Get a Grip: Evaluating Grip Gestures for VR Input using a Lightweight Pen

Nianlong Li¹,² Teng Han¹,³∗ Feng Tian¹,³ Jin Huang¹,³ Minghui Sun⁴
Pourang Irani⁵ Jason Alexander⁶

¹State Key Laboratory of Computer Science and Beijing Key Lab of Human-Computer Interaction, Institute of Software, Chinese Academy of Sciences, Beijing, China
²School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing, China
³School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
⁴Department of Computer Science, Jilin University, Changchun, China
⁵Department of Computer Science, University of Manitoba, Winnipeg, MB, Canada
⁶Department of Computing and Communications, Lancaster University, Lancaster, UK
linianlong16@mails.ucas.ac.cn, {hanteng, tianfeng, huangjin}@iscas.ac.cn, smh@jlu.edu.cn, pourang.Irani@cs.umanitoba.ca, j.alexander@lancaster.ac.uk

Abstract
The use of Virtual Reality (VR) in applications such as data analysis, artistic creation, and clinical settings requires high precision input. However, the current design of handheld controllers, where wrist rotation is the primary input approach, does not exploit the human fingers’ capability for dexterous movements for high precision pointing and selection. To address this issue, we investigated the characteristics and potential of using a pen as a VR input device. We conducted two studies. The first examined which pen grip allowed the largest range of motion—we found a tripod grip at the rear end of the shaft met this criterion. The second study investigated target selection via ‘poking’ and ray-casting, where we found the pen grip outperformed the traditional wrist-based input in both cases. Finally, we demonstrate potential applications enabled by VR pen input and grip postures.

Author Keywords
Virtual Reality; pen input; finger and wrist dexterity; grip postures; handheld controller; spatial target selection.

CCS Concepts
•Human-centered computing → Virtual reality; User studies;

Introduction
Virtual Reality (VR) is emerging as the next generation computing platform for diverse domains including entertainment, education, training, research, clinical practice, and productivity [15]. While the visual quality and immersive capabilities of current consumer VR devices have reached a steady state for presenting impressive graphics, the corresponding input technologies have seen much greater variety in terms of form-factors and user interactions, which remain under active exploration [19, 30, 42, 62]. While researchers have proposed many input modalities, including bare hand gestures [67], gaze [13] and speech [21], application developers heavily rely on handheld controllers as the primary means for 3D spatial [18] and consumer level VR input. This is evident by the large number of multi-button and multi-function controllers with devices such as the HTC Vive [61] and Oculus Rift [54].

For the majority of current VR controllers, users grasp them in a manner to primarily exploit “wrist rotations” [52]. For many applications, including gaming and menu interactions, wrist-based movements are adequately efficient (Figure 7b). However, such grips often require that all fingers maintain a force to hold the device, inevitably restricting movements and input ranges that could normally be achieved if users’ fingers were unconstrained [52]. As VR use proliferates new domains, such as immersive analytics [47], artistic endeavors [4] and clinical simulations [25], refined and precise input operations will be required, needing new grip styles.

We draw attention to the potential of pen-style controllers and grips that can exploit our fingers’ rich dexterity for VR input (Figure 1). We advocate the recent commercial VR/MR pen controllers (e.g., VR Ink [23], Holol Stylus [34], Massless Pen [46]), and explore the use of relaxed grip postures to better involve the fingers’ movement in a skillful and coordinated manner for precise operations. Pen input has a long tradition...
We take inspiration from the ongoing research exploring VR with poking gestures. This paper makes the following contributions: (i) empirical evidence shows the potential of using a pen to afford more precise and dexterous operations in VR; (ii) studies that identified a suitable grip posture for a pen when trying to achieve the largest comfortable range of motion; (iii) studies that evaluated the pen grip posture’s performance in target selections; (iv) demo applications that showed potential applications enabled by VR pen input and grip postures.

**RELATED WORK**

We take inspiration from the ongoing research exploring VR input techniques, and from the recent introduction of pens in VR. In this section, we first look at related work on VR input techniques, and then we examine previous research work exploring pen based interactions. We end with a brief presentation of prior ergonomic studies on pen grip postures.

**VR Input Devices and Techniques**

Input devices play an important role in the creation of successful VR experiences. Traditional desktop input devices like the mouse, joystick, and keyboard are not compatible with fully immersive virtual environments due to the lack of physical surfaces [15]. Researchers have proposed a variety of input methodologies that specifically suit virtual environments. Notable examples include wearable devices [40], vision based tracking devices [49], and handheld controllers [57].

In recent years, wearable input devices have gained popularity both in the research community and commercially. They provide high quality tracking capabilities while enabling natural use of the hand. For instance, data gloves facilitate hand motion tracking, gesture recognition, and haptic feedback [40]. As gloves are often bulky, ongoing research efforts have explored lightweight sensing [45] and haptic rendering mechanisms [31] for data gloves. Similarly, supporting direct hand manipulation in VR via bare hand tracking (e.g., Leap Motion [49]), has drawn significant interest, albeit with a lack of haptic feedback. Alternative input modalities include eye gaze [13] and speech [21]. Though promising, these approaches still require further development for fluid spatial 3D input [15].

Mainstream VR manufacturers commonly use handheld controllers as their primary input device. Early controller prototypes focused on the game-play, not necessarily representing the hands’ essential functions [56]. Current handheld controllers [54, 61] typically include haptic and touch-sensitive controls, allowing users to launch continuous and discrete commands or movements in the virtual environment [15]. These controllers deploy different forms of hand grip in an attempt to retain ergonomic capabilities that mimic humans’ physical and manual dexterity and agility [10]. In essence, controllers exploit humans’ inherent abilities developed for manipulating physical tools [8, 18]. This has motivated a rich set of VR controller designs that capture fine hand and finger motions and provide realistic haptic sensations [42, 57]. Recently released commercial controllers [38] demonstrate the capabilities of capturing both static and dynamic hand postures, as well as grip force, which enrich interactions in VR.

This paper further expands this design space by investigating the input characteristics of a pen-like controller and its grip postures.

**3D Selection Techniques**

Beyond hardware innovations, novel interaction techniques can increase the efficiency of target selection and manipulation in VR. Grasping metaphors emulate natural (real world) object grasping and manipulating actions with the hand [41]. Virtual hand representations are often used, and directly mapped to a user’s physical hand motions [48]. However, this limits reachability, where objects out of a user’s physical reach are not selectable in the virtual world. Non-linear mapping methods can be used to resolve out-of-reach concerns. For instance, Chae et al. [11] recently proposed a method to shrink the virtual space to select distant objects.

Ray-casting is widely used and explored for distant pointing—a virtual ray is ‘casted’ from a point of origin along a given direction [41]. Ray-casting allows the user to select objects beyond their area of reach and requires little physical movement [3]. Research into ray-casting covers various aspects including
how the ray is controlled, the ray shape, and the method to disambiguate among small, crowded, and occluded targets [30]. Other techniques such as using a touch surface [5], and designing physical proxies [12] also facilitate 3D selections.

In our work, we use ‘poking’ gestures for near target selection and ray-casting for distant target selection, and explore the distinctions imposed by different pen grips on these tasks.

**Pen Based Interaction Techniques**

Pen-based input has a long tradition in interactive systems, as it enables precise input on digital device, mimicking our handwriting in physical environments. Cockburn et al. investigated the stylus’ performance in tapping and dragging on touchscreen, and compared it with the finger and mouse [16]. Their results revealed that pen input is accurate even with small targets in pointing tasks, and also facilitates dragging tasks. This was followed by many works seeking to understand pen input characteristics on tablets and phones [1, 14].

Pen input exploits our wrist and fingers’ dexterous motions which can be leveraged for multi degree-of-freedom input [29]. Pressure Widgets [53] and HoverWidget [28] explored the use of continuous pressure sensing and hover sensing to operate onscreen widgets. Closer to our work, Tian et al. [59, 60] explored pen tilting gestures for cursor and menu design. Bi et al. [6] used pen rolling based interaction techniques. Xin et al. [66] carried out a more systematic investigation on how pressure, tilt and azimuth information could be leveraged to enhance pen input. The pen was also found to be useful in designing multi-modal input on touchscreens [7, 32, 33, 51].

Recently, Wu et al. [65] and Wacker et al. [62] demonstrated pen-based object selection and manipulation techniques in Augmented Reality. Our work takes an empirical, study driven approach to design and understand the use of pen and grip postures in VR.

**Ergonomic Study of Grip Postures**

A pen’s input capabilities can be enhanced to a large extent by sensing the user’s grip posture. For instance, Song et al. [55] implemented a multi-touch pen that supported the detection of various grips to enable implicit input mode switching. Grip postures are also widely studied in both the interaction and ergonomic literature [27]. Rahman et al. [52] investigated the dexterity of wrist-based input when a mobile phone was gripped in the hand. This included a framework that measured ergonomic factors like axial range-of-motion and discretization of tilt angles. Eardley et al. [22] examined how mobile devices’ form-factor affected users’ grip postures and hand movements. There is significant literature seeking to better understand how pen grip postures affect children’s learning [44], handwriting performance [35], and stability [64].

Inspired by these prior results, we first study how different pen grip postures affect ergonomic capabilities such as the range of motion and comfort level.

**STUDY 1: EFFECT OF PEN GRIP POSTURES ON WRIST AND FINGER MOTION**

The overarching aim of this study was to examine the effect of pen grip style on the possible range of wrist and finger motion. Participants performed a tilt motion in eight cardinal directions while holding a pen in five common grips.

**Candidate Grip Postures**

We assessed a total of five grip postures based on prior work [20, 33, 55, 64], as well as those commonly used in art [63]. We excluded postures that add firm constraints to the finger/wrist movements. Figure 2 illustrates the five grip postures that were investigated in this study, their descriptions follow:

- **Tripod at Front End (TFE):** This is the most common grip posture for precise writing and drawing on a surface (Figure 2a). It is considered the most popular and useful grip to support precise mid-air drawing and sketching [64] and often adopted in commercial VR pens. Tripod grip, with fingers positioned at the front of the shaft (TFE), facilitates writing. It is however unclear how such a grip impacts operations such as menu selection and target acquisition in VR, that often come with the requirement for agile motions of larger range. In mid-air, there is limited support for users to rest their hands.

- **Tripod at Rear End (TRE):** This represents an alternate tripod grip where users hold the pen shaft at the rear end (Figure 2b). Wu et al. [64] observed many users held pens in this way during surface pointing and clicking tasks. Compared to TFE, gripping a pen at the rear encourages its use as a lever (with the tip feeling ‘weightier’).

- **Quadropod at Rear End (QRE):** A quadropod grip is similar to the tripod grip, except that two fingers (i.e., index finger and middle finger) are used to control the pen together with the thumb (Figure 2c). The ring finger is used to rest the pen. QRE is often seen from holding a writing brush in Chinese calligraphy, where it is easy to twist the pen shaft while keeping the pen stable.

- **Pinch:** In this grip, users perform a pinch action, holding the pen between their thumb and fingers, as if they have picked it up from a table (Figure 2d). With Pinch, the pen can be positioned parallel to the writing surface. In the context of paper writing or drawing, this helps to loosen the user’s wrist and to move the pen by hovering it over the paper.

- **Overhand:** Overhand is the posture used to hold a pen to allow the tip’s use for line drawing and shading on paper in a more versatile way (Figure 2e). The hand is relaxed with fingers...
and thumb lightly holding the pen. It creates several ways to control the pen, such as rolling forward or backward, sliding and even flipping.

**Hardware Configuration**

In this study we do not consider factors such as the pen’s weight, length, or radius as there exist sufficient evidence showing how these factors affect one’s grip capabilities[26]. We do not attempt to design the best pen form-factor as we instead focus on studying how pen grip affects motion. Considering that current commercially available VR pens are commonly bulky in size, we examined the design of over 10 commercial digital pens for mobile phones and pads, and decided to use the Apple Pencil [50] as the design reference. The Apple Pencil is of a compact size, lightweight and reasonable length. This offers many grip possibilities ergonomically, thus fits our study purpose.

A pen proxy, similar in size to the Apple Pencil [50] was 3D printed, as shown in Figure 3a. It measures 178 mm in length, and has a radius of 9 mm. Four additional stickers with reflective markers were attached to the pen for tracking purposes. Altogether it weighs 11 g. The pen was tracked with an OptiTrack V120: Trio¹, which was driven by a Windows 10 laptop. This allows us to retrieve the spatial position and orientation of the pen at 120 fps.

**Participants**

Twelve (12) participants (5 female, avg. age = 24.6) were recruited from a local university for the study. All were right-handed. Each participant was rewarded with $USD20 for their participation. We first described the purpose of the study and instructed each participant to familiarise themselves with the pen in a natural position and pressed the Ctrl button to initialize the coordinates of the reference center. Once ready, a trial started when the participants pressed the Spacebar key and the target direction was prompted on screen. After performing the tilt motion, participants returned their posture to the home position, and pressed the Spacebar key. This ended the current trial and started the next one. Each direction was repeated three times and the same procedure was repeated for each grip posture. The order of the grip gestures used in each block was randomized. During the study, the participants were free to use their fingers and wrist to tilt the pen (i.e., no restrictions were imposed on how they should use their fingers and wrist), and were allowed to rest between trials. The tracker captured the pen tip’s spatial position and the pen’s orientation information at every frame during a trial (120 fps). In total, the system captured 3 blocks × 5 postures × 8 directions × 3 repetitions = 360 data logs for each participant. The participants were asked to complete a NASA-TLX form to rate workload factors from 1 to 7. The study lasted on average 40 mins per participant.

**Results**

As the tilt actions are centred around the wrist and not on a gripping point on the pen, the resulting tilt motion is the joint effects of pen tip translation and pen shaft rotation for each grip posture and for each direction. These two values were calculated by averaging the euclidean distance travelled by the pen tip, and the angle tilted by the pen shaft (i.e., the angle between two vectors standing for the initial and maximally tilted orientations of the shaft). Data were normally distributed according to a Shapiro-Wilks test at the 5% level. Results were analyzed using a repeated measures ANOVA and are illustrated in Figure 4.

**Distance Pen-Trip Traveled**

Overall there was a significant main effect of Grip Posture on the pen tip’s moving distance ($F_{4,44} = 23.557, p < 0.001$). Pairwise comparisons using the Bonferroni adjustment yielded the moving distance with TFE grip ($\mu = 216.79mm, \sigma = 8.55mm$) was significantly larger than all others (all $p < 0.05$), and the moving distance of the TRE grip ($\mu = 151.06mm, \sigma = 5.00mm$) was significantly shorter than all others (all $p < 0.05$). There were no significant differences among Pinch

---

¹https://optitrack.com/products/v120-trio/
Across postures, the moving distance in the Southwest direction was the largest ($\mu = 159.80\text{mm}, \sigma = 9.44\text{mm}$), and the shortest distance was in the East direction ($\mu = 159.80\text{mm}, \sigma = 9.44\text{mm}$) (both $p < 0.05$). This is in line with expectations, as all of our participants were right-handed, with their actions appearing to be more flexible when tilting the wrist and fingers inwards rather than outwards. However, significant effects were not found on every pair of the directions, and there was an interactive effect between Grip Postures and Directions ($F_{28,308} = 15.21, p < 0.001$). The moving distance of TRE and QRE postures were significantly larger than others in the East, Southeast and South direction (all $p < 0.05$, except for the pair of QRE and Pinch in the East direction where $p = 0.078$). In both North and Northwest directions, Overhand had the largest moving distance, but was not significantly different to TRE in North ($p = 0.056$), and to TRE, and Pinch in Northwest (all $p > 0.05$). Besides, in the two directions, TFE and QRE postures led to significantly shorter distance than others (all $p < 0.05$, except for the pair of QRE and Pinch in Northwest where $p = 0.064$).

Tilt Angle
There was a significant main effect of Grip Posture on the tilt angle ($F_{3,44} = 26.19, p < 0.001$). Pairwise comparisons using the Bonferroni adjustment yielded a tilt angle with TRE ($\mu = 57.33^\circ, \sigma = 3.42^\circ$) was significantly larger than other postures (all $p < 0.05$), except for QRE ($\mu = 52.89^\circ, \sigma = 3.93^\circ$) ($p = 0.65$). The Pinch posture ($\mu = 38.87^\circ, \sigma = 2.70^\circ$) led to a significantly smaller tilt angle than other postures (all $p < 0.05$), except for Overhand ($\mu = 42.59^\circ, \sigma = 2.95^\circ$) ($p = 0.237$).

Across postures, Direction also had a significant main effect on the tilt angle ($F_{7,77} = 21.10, p < 0.001$). The tilt angle in the West direction ($\mu = 59.67^\circ, \sigma = 5.51^\circ$) was significantly larger than that of East, Southeast, and Southwest directions (all $p < 0.05$). The East direction let to the smallest tilt angle ($\mu = 34.72^\circ, \sigma = 2.67^\circ$), which was significantly smaller than other directions (all $p < 0.05$) except for the Southeast direction ($p = 0.73$).

Similar to the moving distance, there was also an interactive effect between Postures and Directions ($F_{28,308} = 13.03, p < 0.001$). In the East direction, the TRE and QRE postures led to significantly larger tilt angle (all $p < 0.05$). In the Southeast, South, and Southwest direction, the tilt angle of Pinch and Overhand postures were significantly smaller than all other postures (all $p < 0.05$). No significant effect was found among the five postures in the North and Northwest direction except for the pair of TRE and QRE ($p = 0.01$; $p = 0.015$).

Subjective Workload
Study results were analyzed using a Friedman test with Wilcoxon signed rank tests for pair-wise comparisons. Overall, average NASA-TLX ratings for all grip postures were less than 3.5 (with 1 indicating the posture was very easy and 7 being very hard to learn) except for QRE ($\mu = 4.127$). This suggests that most of the participants could use these postures well except for QRE (Figure 5). Some participants said that they had no experience in writing with Chinese brush. The Friedman test yielded a significant difference in Grip Posture ($\chi^2(4, N = 12) = 23.54, p < 0.001$). The TRE posture was rated as significantly easier ($\mu = 1.87$) to use than others (all $p < 0.05$) except for TFE ($\mu = 2.08$) ($p = 0.747$).

Discussion
This study focused on identifying the effect of different pen grips on achieving the largest comfortable range of motion. It allowed us to examine the extent to which, the joint motions of wrist and finger movements contribute to the tilting actions. Different grip postures added various constraints to not only to the fingers’ extension, but also to the wrists. Amongst the five grip postures, Pinch and Overhand were found to produce smaller range of motions compared to the others.

Overall, TRE performed better than the other grip postures. It led to larger moving distances and tilt angles in most directions. This is understandable, as compared to TFE, gripping the pen at the rear end places the fulcrum further from the pen tip, allowing a larger radius of movement by the pen tip. Furthermore, a relatively more loaded force can be sensed by hands. Moreover, the TFE grip often rests the pen on the thumb cleft, that restricts the tilt range of the pen. Participants also confirmed that TRE felt more comfortable and relaxed.

QRE also grips the pen at the rear end, and similar to TRE, it shows advantages in moving distances and tilt angles in East, Southeast, and South. However, TRE significantly outperformed QRE in the North and Northwest directions. Compared to TRE, QRE grips the pen more strenuously and the two fingers (i.e., index and middle) used to control the pen restrict its upward tilt movement (i.e., North).

The effects of Direction matched our expectations, where the fingers became more engaged while tilting the pen upwards (North) and downwards (South). As confirmed by the par-
In contrast to VR interactions, where menus are floated in space beyond a user’s reach, pens are physically reachable. The action is similar to touching a screen, however, users do not perceive any haptic feedback. It is intuitive and normally requires no trigger mechanism.

**Tilt:** In contrast to *Poke*, *Tilt* gestures are often used alongside ray-casting for distant pointing. This is especially common in VR interactions, where menus are floated in space beyond a user’s zone of physical reach. A trigger mechanism is needed to confirm the user’s selection, e.g., button or dwell.

**Poke:** *Poke* gestures are used to select items on menus that are physically reachable. The action is similar to touching a screen, however, users do not perceive any haptic feedback. It is intuitive and normally requires no trigger mechanism.

**Pen Grip**

Beyond the *Grip Posture* (*Pen Grip*, *Palm Grip*), the *Target Width and Target Distance* were set following Teather et al. [58], where the *Poke* condition, the target widths (i.e., diameter of the target ball) were set to (0.018, 0.03), and the target distance were set to (0.14, 0.30, 0.38). In the *Tilt* condition, the target widths were set to (0.054, 0.09), and the target distance were set to (0.42, 0.90, 1.14), all with the units of meters. This resulted in Fitts’ IDs of (2.50, 3.13, 3.46, 3.77, 4.14, 4.47) in both conditions.

The order of the *Pen Grip* and the *Palm Grip* was counterbalanced, while each pair of *Target Size* and *Circle Radius* was repeated 5 times, and appeared in random order. In a similar manner to Teather et al. [58], a trial was counted when the participant completed the selection of all 13 targets around the circle, thus in each session, there were 2 (grasp postures) × 2 (target size) × 3 (circle radius) × 5 (repetitions) = 60 trials, that included 60 (trials) × 13 (targets) = 780 poke or tilt selections.

**Apparatus and Procedure**

This study used the same tracking device and pen prototype as used in Study 1. A different group of 16 right-handed participants (7 female, avg. age = 23.63) were recruited from the same university. In the same manner as Study 1, each of the participants were rewarded with $20 for their participation. They were introduced to the idea and the study setup, and were asked to stand in front of a table while wearing an HTC Vive Headset (Figure 6b and 7b) that ran the experimental (Unity) application. They were provided sufficient time to get familiar with the tasks and techniques before the formal study started.

A trial is initialized by displaying 13 target balls at the current condition, and with a target ball highlighted in red. Once the ball is successfully selected, the ball turns blue for 1 second and the next target ball is highlighted in red. The trial starts...
when the participants make the first selection, and ends when
the last target gets selected. The appearances of the target was
randomized. Note the time for selecting the first target did not
count for the trial time.

After each session, the participants were asked to complete
a NASA-TLX form for each of the grip postures. The study
lasted an average of 1 hour per participant.

**Poke Gestures**
The virtual targets were displayed in the VR headset, at a
fixed depth of 0.2m, with their heights adjusted based on
the participants’ hand positions, making it comfortable for
users to perform poke gestures. Specifically, the height value
was determined by averaging the current participant’s hand
position when naturally holding the pen with Pen Grip and
Palm Grip when not moving towards the target. The poking
action is identified by the pen entering the 3D target volume
and then exiting it. To avoid the participant slide over the target
(instead of poking it), they were asked to lift the pen at least
5 cm after a selection. The pen tip entering and exiting a pre-
defined invisible plane (5 cm from the target center) without
touching the target was recorded as an error. Meanwhile, the
targets must be ‘poked’ from the side that faced the participants
in order to be selected. If the participants missed to select a
target, she/he had to continue to select the next one. An error
was counted and the missed target appeared in the last. Thus
for each trial, we recorded the conditions, trial time and errors.
The participants were allowed to take rest between trials, but
not during a trial.

**Tilt Gestures**
In the second session, the Tilt gesture was used. The experi-
ment setup was the same as the first session, except that the
targets were displayed 2m away, and the participants made
the selection with ray-casting. Once the ray is moved within
the target, the target turned to green, and the participants held
it (i.e., dwell) for 500 milliseconds to confirm the selection.
Note that dwell was simply a design artifact of our study, not
necessarily an optimal design choice. Selection can be made
with other methods, e.g., button, force press, finger tap, de-
pending on the tasks and applications. Also in this session, the
participants had to make successful selections before moving
to the next targets, thus there was no error data.

**Results**
Data were normally distributed according to a Shapiro-Wilks
test at the 5% level. We analyzed the results using repeated
measures ANOVA and post-hoc comparisons with Bonferroni
adjustment.

**Selection Time**
For the selection tasks based on Poke gesture, a within-subject
ANOVA analysis on average selection time showed a sig-
ificant effect for Postures ($F_{1,15} = 58.06, p < 0.001$), Tar-
get Width ($F_{1,15} = 138.93, p < 0.001$), and Target Distance
($F_{2,30} = 64.79, p < 0.001$). Post-hoc tests revealed that the
Pen Grip ($\mu = 1023ms, \sigma = 31.22ms$) was significantly faster
than the Palm Grip ($\mu = 1257ms, \sigma = 46.82ms$) ($p < 0.001$).
Based on our observations, with Pen Grip, the participants
were more likely to move their forearm, and co-ordinated
their wrist and finger tilting actions when approaching the
targets. Participants’ actions appeared agile and dexterous. In
contrast, poking with the Palm Grip involved more arm and
forearm extensions, and less wrist motions. Additionally, the
post-hoc tests on other factors can be anticipated: selection
time significantly increased with increasing target width and
target distance (all $p < 0.001$). No other significant interaction
effects were found (all $p > 0.05$).

For the selection task based on Tilt gesture, the 500ms dwell
time was removed, as in this study, our primary goal was to
measure target acquisition times, not target plus selec-
tion. There was a significant effect for Postures ($F_{1,15} = 31.30, p < 0.001$), Target Width ($F_{1,15} = 170.90, p < 0.001$),
and Target Distance ($F_{2,30} = 313.39, p < 0.001$) on average
selection time. Post-hoc tests revealed that the Pen Grip
($\mu = 926.014ms, \sigma = 28.00ms$) was significantly faster than
the Palm Grip ($\mu = 1039.20ms, \sigma = 31.86ms$) ($p < 0.001$).
The participants were able to leverage both finger and wrist
tilting motions to point the ray to the targets when using Pen
Grip, whereas with the Palm Grip they controlled the ray with
wrist rotations. Similar to the previous session, significant interactive effects existed for Postures × Width ($p < 0.001$).
No other significant interaction effects were found.

**Error Rates**
The error rates of the poke gesture were calculated and all three
variables Postures, Width, and Distance had significant main
effects of error rates (all $p < 0.001$). However, no significant
interactive effect was found between them. The error rates of
Pen Grip ($\mu = 0.096, \sigma = 0.012$) was significantly lower than
Palm Grip ($\mu = 0.133, \sigma = 0.012$) ($p < 0.001$). In the Tilt
condition, the participants had to successfully select each tar-
get before proceeding to the next. As such, there was no error
data. The same protocol was used in prior evaluations [17].

**Fitts’ Law Test**
Trial times across the grip postures and selection techniques
were further analyzed to see whether they can be modelled
with Fitts’ law [24]. Pairing the target widths and distances
yielded the same group of five IDs for both sessions. Figure 9
revealed a strong correlation between the trial time and ID for
Tilt gestures, with both $R^2$ values above 0.88. We noticed that
the two lines intersects at very low IDs, and with increasing
IDs, Pen Grip selection is more efficient than Palm Grip.
I can’t feel the difference between Pen Grip and Palm Grip in poking, especially for a long time, but it allows you to select target faster. “ [S14]  

“I can’t feel the difference between Pen Grip and Palm Grip in tilting, but shoulder soreness occurs when using Pen Grip for a period of time.” [S14]

Discussion
The results showed that using the Pen Grip improved target selection time with both poke and tilt gestures, and was less error-prone with poke gestures. This indicates that the Pen Grip is an efficient candidate for selection in VR interfaces. Exploiting the fingers’ dexterous motions helps to accelerate the process of adjusting the pen tip towards the target, while the user’s wrist flexion motion can then quickly ‘jab’ it. Holding the pen with the Pen Grip benefits from less arm movement, where in the Tilt session, users were able to tilt the pen only by moving their fingers, especially when the target distance was small. While in the Poke session, users did not necessarily engage their upper arm motions to make the selection. 

Another factor that may have contributed to this performance improvement, is that with Pen Grip, the pen is held higher than that with Palm Grip, meaning the pen is closer to the eye-line. This makes it easier for users to aim the pen. On the other hand, this might also introduce a shortcoming: when the pen is held higher, the appearance of the virtual pen in the dominant view angle might add obstructions, while this is not the case for holding the pen at a lower position, where the virtual pen only appears in the user’s peripheral vision. However, this needs further investigation to fully understand this effect.

The two grip postures also resulted in different comfort zones of operation. We did not include this factor in the study, but it is clear that with Pen Grip, a user can easily point a pen to the front and downwards, while with Palm Grip, pointing a pen to the front and upwards will be easier. Additionally, the fatigue caused by different grips is also worth deeper exploration. The participants reported that the Pen Grip felt relaxed at the beginning as they could just leverage finger motions to control the pen. However, fatigue was gradually perceived when the forearm were raised and held for a certain time, and participants took more breaks during the test. In comparison, the participants did not report similar issues with Palm Grip. This indicated that the Pen Grip was not suitable for performing long-duration tasks.

INTERACTION TECHNIQUES AND APPLICATIONS
This section presents three scenarios enabled by VR pen input and grip postures.

Switching Grips
Although the Pen Grip is shown to be more capable for quick and precise selection, the Palm Grip has the advantage of familiarization and is less tiring. Inspired by this, we looked at how to combine the two to bring a better user experience when using the pen. Key to this, is switching between gripping postures. It can be easily done, for example, by flipping the pen. This enables new interaction opportunities like input mode switching. For instance, in an interior design application, one can select furniture with the Palm Grip, and switch to the Pen Grip to invoke a menu and perform operations like picking a color from a palette, or adjusting values on a scale by dragging a slider (Figure 11). There are many approaches to
support grip switching actions, such as using capacitive touch sensors on the pen’s surface or an internal motion sensor.

**Pen-based Gestural Widgets**

The user’s ability to exert dexterous and large-range finger motions while gripping the pen at its rear end, allows us to design a set of unique pen based gestures in mid-air that are otherwise hard to do with other postures. With the grip, one can easily perform short and rapid input bursts, such as tilting and poking. These gestures can be captured and recognized either via the external motion tracker, or in a self-contained way (e.g., accelerometer). In this demo scenario (Figure 12), we show that the gestures can be utilized as interface widgets, which trigger scrolling or zooming operations on a web view with tilting and poking the pen respectively. A double tilt makes the page scroll automatically, while a double poke calls a stop. In 3D painting and visual data analysis scenarios, it is more convenient to manipulate 3D models. Gestural widgets as such can help improve the interface efficiency without relying on menus.

**Avoiding the Eye-Hand Visibility Mismatch**

When using virtual pointing techniques to select targets in VR, the user may suffer from eye-hand visibility mismatch [3]. For example, it is difficult for users to manipulate complex molecular substructures or select the human skeleton in medical treatment [2]. More specifically, in the scenario shown in Figure 13a, the user can select an object which is hidden by another object he can’t see with the Palm Grip, leading to misinterpretation. Moreover, in Figure 13b, the larger object is stacked below the target object. While a user can see the target object, the larger object can easily block the pointing ray if the user holds the controller with the Palm Grip. This may cause selection errors and cause user frustration. In contrast, the Pen Grip avoids this problem and keeps a high consistency between the visibility of the eyes and that of the hands. As shown in Study 2, the posture also helps users when aiming the pen at targets where precise selection is needed.

**DISCUSSION AND FUTURE WORK**

The concept of using a pen shaped controller in VR is not new. However, little was previously known on its performance in the fundamental interaction tasks of pointing and selection. Our first study identified the grip posture best suited for a pen when trying to achieve the largest comfortable range of motion. Different from the typical tripod grip that holds the pen at its front end, we found that using the tripod gesture at the rear end resulted in a larger range of motion while maintaining a good level of comfort. With the second study, we confirmed its efficiency in pointing tasks when used for performing tilt and poke gestures. The results indicated there is a good potential to use the pen for VR operations in the future, provided we instruct users on the proper grip. One way of doing so is in the pen design itself, suggest where and how the user should grip.

Nonetheless, our paper has several limitations, some of which are worth exploration in future research. Besides tilt and poke gestures, the pen’s use has been explored in many other ways. Bi et al. [6] investigated the properties of rolling a pen and its design space on a touch surface. This rolling gesture is also a viable input method in VR. One challenge of performing pen rolling gestures in air, is to identify suitable grip gestures that keep a good balance of range of motion (i.e., angle of rotation) and stabilization. The Quadropod grip posture could be a good candidate but requires further evaluation.

In the second study, dwell was used to confirm a selection with tilt gestures and ray-casting. Dwell is considered an experimental artifact, not necessarily an optimal choice to trigger a selection. Designers could also add a physical button to the pen, close to the user’s thumb or index finger when they hold the pen. This could be a more practical solution, but raises a new challenge that as the fingers are primarily used in gripping and wielding the pen in the Pen Grip posture, it is unclear how switching the fingers' roles frequently would affect performance in pointing tasks. The Palm Grip could be better in this case, as it relies on wrist rotations, leaving
thumb and index finger for other operations. This indicates that a joint use of both grip postures could be more practical, as shown in the demo application. Another challenge is the location of physical button, one cannot easily look directly at the physical pen to see where the button is. Some existing methods may be able to solve this problem by using motion sensors for detecting taps without buttons [32, 43].

Similar to most controllers, haptic feedback is critical for creating successful use experiences of the pen. This brings new opportunities as well as challenges for future work. Adding vibrotactile feedback to the pen device is feasible with a high frequency linear actuator, which can be leveraged to render a rich set of haptic profiles with proper mapping to interface interactions.

In our studies, we used a compact-sized 3D printed pen, that was similar to a regular pen, which users are accustomed to. The results may vary if a bulkier or heavier pen was used. However, despite the technical challenges, developers should aim for VR pen’s of this smaller form factor; larger, heavier form factors may play a critical (potentially negative) role in the user’s experience. Meanwhile, other form-factor design options should be considered and evaluated in future work.

CONCLUSION
The current design of handheld controllers, such as those used with the HTC Vive and Oculus Rift, primarily exploit wrist rotation gestures for input. These miss the opportunity to fully utilize the fingers’ capability for dexterous movements for high precision pointing and selection in VR. In this paper, we look into the potential of a pen-style controller and grip that can exploit our fingers’ rich dexterity for VR input. With two studies, we found that the tripod grip at the rear end is an optimal grip posture for a pen when trying to achieve the largest comfortable range of motion. This improved the target selection time with either poke or tilt gestures, and was less error-prone with poke gestures. Finally, we discussed interaction design and application opportunities with pen input in VR.

ACKNOWLEDGMENTS
We thank Luyao Shen for help with figure drawing and video editing, Zongqi Zhang for help with conducting the experiments, Sinong Zhan for help with the software implementation, the members of IEL ISCAS, and the reviewers for their constructive feedback. This work was supported by the National Key R&D Program of China (Grant No. 2016YFB1001402), the National Natural Science Foundation of China (Grant No. 61802379, 61872164) and Youth Innovation Promotion Association CAS.

REFERENCES


