Graphics Interface Conference 2015, 3–5 June, Halifax, Nova Scotia, Canada

Hands, Hover, and Nibs: Understanding Stylus Accuracy on Tablets

Michelle Annett* and Walter F. Bischof†
University of Alberta

ABSTRACT

Although tablets and styli have become pervasive, styli have not seen widespread adoption for precise input tasks such as annotation, note-taking, algebra, and so on. While many have identified that stylus accuracy is a problem, there is still much unknown about how the user and the stylus itself influences accuracy. The present work identifies a multitude of factors relating to the user, the stylus, and tablet hardware that impact the inaccuracy experienced today. Further, we report on a two-part user study that evaluated the interplay between the motor and visual systems (i.e., hand posture and visual feedback) and an increasingly important feature of the stylus, the nib diameter. The results determined that the presence of visual feedback and the dimensions of the stylus nib are crucial to the accuracy attained and pressure exerted with the stylus. The ability to rest one's hand on the screen, while providing comfort and support, was found to have surprisingly little influence on accuracy.

Keywords: Pen computing, stylus, pen, accuracy, visual feedback, hand posture, stylus design, nib diameter.

Index Terms: H.5.2. User Interfaces: Input devices and strategies.

1 INTRODUCTION

Stylus-enabled tablets are purported to provide a natural and intuitive method to diagram, sketch, and write, while simultaneously providing functionality not possible with traditional pen and paper (e.g., search, undo, and redo [20]). The use of such a peripheral would be beneficial, as it would allow users to transfer the metaphors and motor skills honed from years of creating precise, legible content with pen and paper to the digital world. The ability to retain such skills and generate content that is near identical to that found with pen and paper is important for users such as architects, graphic designers, artists, or those using tablets for note-taking, mathematics, or annotation, such as students, doctors, or engineers. Such populations require implements that afford the ability to form, join, and terminate strokes with immense precision and accuracy. These users assume that the precision and activities supported by pen and paper will be equivalent to those found with tablets [1]. Although stylus-enabled tablets offer many benefits to such populations, the technology inherent in tablets today does not support precise content creation, resulting in immense user frustration and content that is often illegible, larger than its paper-based counterpart, and visually dissimilar [1, 3, 15, 20].

Unlike pen and paper, tablets utilize sensors to detect the position of a stylus and calibration procedures to map the sensor space to screen space. Tablets such as the Cintiq Companion and Surface Pro support *active styli*, through the use of a multiplexed stylus and touch digitizer or two separate digitizers. Such technology allows for small stylus nibs (usually 1-2 mm in diameter), the ability to disambiguate between touch and stylus input, and harnesses the hover state to provide feedback [5] to reduce inaccuracy. Even with stylus support at the forefront, the accuracy of such devices is still inferior to pen and paper [1]. Alternatively, tablets with only capacitive digitizers, such as the iPad or Galaxy Tab, do not support provide location feedback, or encourage the transfer of traditional pen and paper hand postures [1]. Such devices are not designed for stylus input, although a growing ecosystem of users employ 'finger emulating' *passive styli*. While such styli accommodate low-accuracy selection and are good for less-precise tasks such as sketching, their large deformable nibs (i.e., 5-7 mm) make them less than ideal for high precision-based tasks. Given the technological diversity and the user experiences that result, it is unsurprising that the stylus and tablet have yet to replace traditional pen and paper.

Although stylus inaccuracy has long been identified as problematic [1, 15, 19, 21, 23], it has yet to be solved. As recognized by Vogel and Balakrishnan [25], ongoing work within industry has improved many tablet-based issues such as parallax, latency, and calibration. Due to technical and manufacturing limitations, completely eliminating these hardware issues is difficult. As such, a number of software-based widgets including CrossY [2], Hover-Widgets [8], and Pointing Lenses [19] have been developed to overcome inaccuracy. While such hardware and software advancements have improved the user experience, there has yet to be a holistic understanding or systematic evaluation of all the factors that influence stylus accuracy.

The present work identifies and presents ten factors that influence stylus accuracy on tablet devices today. As it is infeasible to tease apart the relations between such a multitude of factors in a singular piece of work, we report on a two-part experiment that assessed a subset of these factors, namely *visual feedback*, *natural hand postures*, and *nib diameter*. A richer understanding of how these factors influence accuracy was attained using standardized tasks and measures. The quantitative results and participant comments underscored the importance of visual feedback and nib diameter to the inaccuracy experienced, while devaluing the (previously hypothesized) importance of supporting natural hand postures.

2 SOURCES OF INACCURACY

To obtain a better understanding of the factors that could be implicated in the inaccuracy found today, a literature review and summarization was conducted. Keeping both industry, research, and end-user interests in perspective, this exploration focused on not only the hardware limitations that result in inaccuracy, but also factors relating to the user, their expectations, and the stylus. The review resulted in the identification of ten factors (grouped into three categories), that are hypothesized to influence accuracy.

2.1 Underlying Technology

The choices made by manufacturers, in addition to current technological limitations, have resulted in at least five factors that influence accuracy. One of the foremost factors, identified by Lee et al. [12], Nescher and Kunz [16], Ramos et al. [19], and Vogel and Balakrishnan [25], is *display* or *visual parallax,* (i.e., the distance between the surface of the screen and the underlying digitizer). The thicker the glass used in the display and the farther the digitizer is from the surface, the more inaccuracy. While advancements have been made to combine touch and stylus digitizers, or bond digitizers directly to the screen, such disparities make it difficult for the user to determine the precise location of the stylus.

---

* email: mkannett@ualberta.ca
† email: wfb@ualberta.ca

Transducer parallax, also influences accuracy. With the use of electromagnetic or active capacitance with active styli, the circuitry required for sensing is contained within the barrel of the stylus, not the nib itself. This is due to the sheer size of the transducers that are used. As explained by Ward and Phillips, this creates an offset between the tip of the nib and stylus sensor itself [26]. When the stylus is held vertically, inaccuracy is not apparent, but when held at an angle (as is common), this offset manifests on-screen.

For consistent accuracy, Ward and Phillips also suggested that *sensor linearity* is important [26]. To ensure sufficient signal coverage in all areas of the screen, sensors must extend past the edge of the display to match the accuracy in the corners with the centre. When this does not happen, the stylus location is not sensed correctly in all areas of the screen. This can be additionally compounded when using an active stylus with electromagnetic sensing. The introduction of novel magnetic fields from a metal table or the tablet case could then cause drift, making a device appear miscalibrated and inaccurate, seemingly at random.

The *calibration techniques* aligning the sensor and display spaces are also often a source of inaccuracy [1, 26]. If inadequate calibration is performed by the user, in the factory, or within firmware, the location of the stylus as displayed on the screen could appear offset from its real world location. While users can adapt to such offsets, switching styli or devices would require a relearning and adaption to the offset.

Friction and *surface texture* have also been cited by Annett et al. [1], Mohr et al. [15], and Sun et al. [23] as important. With most device manufacturers utilizing non-textured glass for the surface of the display, it becomes easy to lose control of the stylus. While these materials allow for vibrant, aesthetically-pleasing displayed content, they do not provide the tactile feedback or friction necessary for inking. Increased effort is thus needed to ensure that the stylus lands, and moves across the screen, as desired.

2.2 Stylus Design

Occlusion is often cited as a major source of inaccuracy, by, among others, Annett et al. [1], Badamet et al. [2], Ramos et al. [19], and Vogel and Balakrishnan [24]. Similar to traditional writing implements, the diameter of the stylus nib (e.g., 0.7 mm with traditional pens and pencils versus 1.6 mm - 7 mm with digital styli) and the stylus barrel naturally occlude on-screen content. Techniques such as Vogel and Balakrishnan's callout widgets [24] or Lee et al.'s PhantomPen [12] propose spatially offsetting the location of content or virtually reconstructing the nib to allow hidden content to become visible.

As users are particular about the writing and sketching implements they prefer, it may well be the case that the *design of a stylus* influences precision and accuracy as well. Although Goonetilleke et al. [7], Park et al. [18], Ren and Mizobuchi [21], and Wu and Luo [27] evaluated user comfort and preferences for various barrel geometries, lengths, and weights, little attention has been devoted to other important facets of stylus design. As identified by Annett et al. [1], nib malleability is important in the perception of accuracy, as the deformation resulting from pressing the stylus to the screen creates occlusion and ambiguity with regards to the location of the stylus (similar to Holz and Baudisch's findings with finger-based interaction [9]). Similarly, the shape of the nib [1], colour and transparency of the nib, and taper of the barrel [26] have been hypothesized to influence accuracy as well.

2.3 User Interaction

Accuracy is rarely hypothesized to be influenced by user factors or behaviours. Prior work by Annett et al., identified *hand posture* adaptations unique to digital devices that were the by-product of poor palm rejection and inadequate surface textures [1]. They hypothesized that unnatural postures were the cause of inaccurate and messy content because such postures did not provide the hand stability needed to write and sketch. While Matulic and Norrie did not find any differences between resting versus holding the palm above a multi-touch table top while tracing shapes, they suggested that stability may influence targeting moreso than other motions [14]. Both of these studies identified the role that hand posture and support seem to have on accuracy, but did not investigate further.

As it has long been identified as beneficial for aiming movements [6], *visual feedback* can also play a role in accuracy. As originally suggested by Buxton [5], tablets that have the ability to detect the position of a stylus before it touches the screen can provide visual feedback in the form of an on-screen cursor [1, 22]. Although such cursors do not exist with traditional pen and paper, exploiting the hover state provides users the opportunity to correct their movements before any actions are taken. It is unknown to what degree users depend on this information or how much the user experience degrades when such feedback is unavailable.

In addition to hand posture, *prior experience* and *expectations* with pen and tablet systems can influence perceived accuracy as well [1]. Many users erroneously assume the stylus is sensed using resistive, pressure sensitive means, similar to early touch screens. Others may have had poor experiences with prior systems and automatically adapt their behaviour to overcome issues that no longer exist. Those only having experience with passive styli are aware they can be imprecise and could have developed biases towards styli in general. Each of these scenarios decreases confidence with stylus-based input, possibly affecting one's unconscious motor commands as well. This lack of confidence and frequent comparisons between pen and paper and poor digital experiences could manifest itself in the content created.

2.4 Summary

Although we do not claim to be exhaustive, our review identified that there many factors can influence accuracy. From a pragmatic perspective, it is difficult to evaluate all aforementioned factors: some factors are limited by technological restrictions that cannot be evaluated without prototype systems (e.g., see recent work on latency by Ng et al. [17]), whereas others are compounded by numerous sub-factors or are intertwined with others. We can, however, explore a subset of these factors. Herein, we describe a two-part experiment that was conducted to understand the influence of three lesser-understood, but increasingly important factors: *hand posture*, *visual feedback*, and stylus design (i.e., *nib diameter)*.

3 EXPERIMENT PART I: HAND POSTURE AND VISUAL FEEDBACK

In the first part of the experiment, we focused on two factors that are closely intertwined in the visuomotor system, i.e., *hand posture* and *visual feedback*. As Annett et al. identified, there are many behavioural adaptations users employ to overcome inadequate surface texture or unintended touch [1]. When the wrist or hand has to be elevated above the surface, users make compensatory movements to stabilize forearm, upper arm, and stylus as they cannot isolate movements from the fingers, wrist, and elbow [4]. It is possible that these additional movements add 'noise' to the motor system, generating the wobbly lines, messy content, and inaccuracy that is often found [1, 14]. Unlike prior work with a pen-based multi-touch tabletop input [14], the present exploration explicitly allowed or disallowed participants to rest their palm (and other areas that generate spurious touch input such as the forearms, wrists, and fingers) on a tablet, thus allowing the 'noise' introduced by such unnatural behavioural adaptations to be measured.

Although the hover cursor is a visual feedback mechanism, there is still much that is unknown about the usefulness of said feedback. When the 'noise' from holding one's hand aloft is coupled with the ability to receive feedback about the location of the nib, how much does it improve one's accuracy? What happens when there is less noise (because the palm is resting) but no visual feedback about the cursor location? Prior work explored different feedback modalities for pen-based input [22], however the complex experimental design (i.e., 6 factors with multiple levels) and lack of post-hoc analysis, left the role of visual feedback unknown. Within the present exploration, our straightforward modification and examination of the presence and absence of the hover cursor allowed for an explicit understanding of the role visual feedback plays in accuracy.

3.1 Methods

To evaluate the effects of the visual feedback and hand posture on accuracy, a 2D Fitts' law task and a simple writing task were used.

3.1.1 Participants

Sixteen participants (6 female) from our institution were recruited to participate (M = 24.9, range 18-34 years). All participants were right handed, were naive to the purpose of the experiment, and the majority had prior experience using touchscreen-based tablets. Three participants used a stylus on a regular basis. Each participant participated in Part 1 and Part II of the experiment, with half starting with the first segment, and half the second. Participants received a $20 honorarium for their participation.

3.1.2 Apparatus

The experimental setup for Part I and Part II of the experiment were identical. A Samsung Series 7 Business Slate, running Windows 8.1 with a resolution of 1366 x 768 pixels, a PPI of 135, and hover height of 2 centimetres, was placed on a desk in landscape orientation (Figure 1). Participants could reorient and reposition the tablet as needed to ensure that they were comfortable, but it was required to remain on the desk. Custom software controlled the presentation of each task and recorded the data generated by participants. To ensure the stylus was calibrated as accurately as possible, a 273-point calibration procedure was performed [28].

During the experiment, participants were given a Surface Pro stylus (Figure 1) with a standard 1.6 millimetre blue nib that protruded 1.8 millimetres from the tip of the stylus. The buttons on the barrel and end of the stylus were disabled to prevent any erroneous input from occurring. Throughout the experiment, all touch input was also disabled to ensure participants could touch the screen without fear of accidental inking or navigation.

3.1.3 Experimental Conditions

Two factors were evaluated. For the first factor, *Hand Resting*, participants' ability to rest their wrist, fingers, knuckles, etc. on the screen was manipulated. In the *on-screen* condition, participants were encouraged to employ their natural 'pen and paper' hand posture, i.e., they were permitted to rest their wrist, fingers, knuckles, and so on, on the screen. In the second condition, *in air*, participants were explicitly instructed to avoid touching the screen and bezel with their wrist, fingers, knuckles, forearm, and so on. Participants were thus forced to 'float' their hand and forearm above the tablet [1]. As tablets are unable to sense touch input along the bezel, and tablet instrumentation and augmentation was not desired, we opted to watch each participant via web camera and provided them with verbal feedback whenever their skin touched the surface of the tablet in this condition. Whenever this occurred, the task was restarted.


Figure 1. (Top) Experimental setup with the Samsung Slate and keyboard used in the writing task. (Bottom) The Surface Pro pen.

For the second factor, *Hover Cursor*, the feedback available whenever the stylus was in the 'hover state' [5] was manipulated. In the *cursor present* condition, an on-screen cursor (identical to the diamond-shaped Windows 8.1 hover cursor) was visible whenever the stylus was detected above the screen (i.e., from 0 to 2 cm, as per the tablet specifications), disappeared when the stylus was in contact with the screen, and was invisible when the stylus was outside the hover state (i.e., greater than 2 cm above the screen). The cursor visualization was thus identical in behaviour and visual appearance to the Windows-based pen-enabled systems of today. In the second condition, *cursor absent*, the cursor was set to invisible and remained so for the duration of the task.

Across both conditions, participants were allowed to rest their elbow on the surface of the desk and encouraged to take breaks to mitigate fatigue. While we could have included other conditions where the elbow also had to be elevated or the orientation and angle of the tablet was modified, we opted to constrain our analysis to scenarios and behaviours that were representative of commonplace inking activities and similar to prior work [1].

3.1.4 Tasks and Procedure

A majority of the stylus-based activities can be categorized into one of two types of interaction: selection or stroking (inking). As such, two tasks from the ISO 9241-9 [10] and ISO 9241-411 [11] were used. In the first task, *selection*, participants performed the ISO 9241-411 2D multi-directional tapping task. Participants were presented with a radial array of targets and were required to select the highlighted target as quickly and accurately as possible (Figure 2). Three target diameters (i.e., 12, 32, and 53 pixels) and three target amplitudes (i.e., 140, 463, and 683 pixels) were used, resulting in 9 experimental conditions. As each condition contained 13 targets, 12 selections were made before the next trial began. This task simulated the selection of targets or icons in an interface.

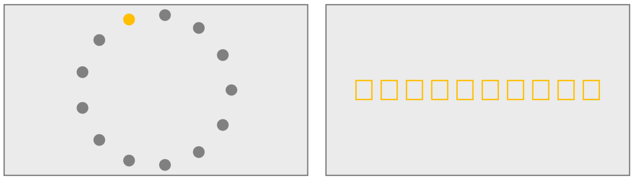
Figure 2. Screenshots of the tasks performed by participants, (Left) selection, (Right) writing.

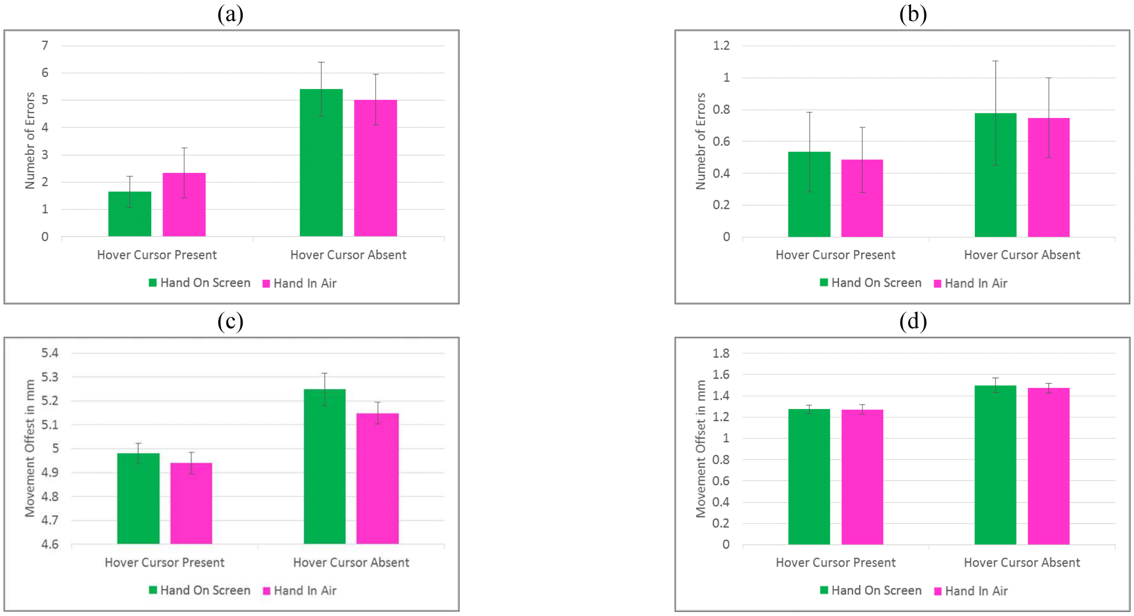

Figure 3. (a) Error for the selection task, (b) Error for the writing task, (c) Movement offset for the selection task, and (d) Movement offset for the writing task. All error bars represent the standard error of the mean.

For the *writing* task, participants performed the ISO 9241-9 free-hand input test, wherein participants wrote the digits from 0 to 9 in nine target rectangles on the screen, with one digit per rectangle (Figure 2). Participants wrote each digit such that it was centred and completely contained within each rectangle. After participants wrote the last digit, they pressed the space bar on a keyboard to advance to the next trial. Three rectangle widths (i.e., 32, 52, and 78 pixels) and three inter-rectangle distances (i.e., 1, 36, and 53 pixels) were used, resulting in 9 experimental conditions. As specified by the ISO 9241-9 standard, this task simulates the natural free-form input, i.e., inking, that is common with a stylus.

Throughout the experiment, participants were seated in a computer chair at a standard office desk and given a short demonstration of the task to be performed. Three training trials (that employed parameters not evaluated during the experiment) were then performed to familiarize participants with the stylus, tablet, and tasks. As multiple tasks and conditions were evaluated, task presentation order and the order of trials was counterbalanced using a Latin square design. At the conclusion of each segment of the experiment, participants answered a short Likert-scale based questionnaire on noticeability, accuracy, and preferences (e.g., "The hover cursor was noticeable", "I was more accurate with the hover cursor than without", "I would prefer to have a hover cursor while writing, sketching, or annotating a document", and so on).

Although many other tasks could have been chosen (e.g., steering, drawing lines, tracing, copying sentences, etc.), we wanted to utilize tasks that had low cognitive load, could be performed quickly (given the number of trials required), were representative of the tasks that users would perform with a stylus, and allowed for quantifiable changes to be measured and replicated by others.

3.1.5 Measures

Although many methods can be used to measure the accuracy of target selections, few are also applicable to written content [20]. We thus used a fine and a coarse measure adapted from Mackenzie et al.'s accuracy measures [13] to measure accuracy. For *error*, the number of attempts necessary to tap each target during selection was computed and for the writing task, the number of strokes that fell fully or partially outside each target rectangle was computed. We also computed the *movement offset* for each task (i.e., the deviation in millimetres from the centre of the selection target to the stylus location and the distance in millimetres from the centre of each written digit to the centre of the target rectangle). We also computed *movement time* (i.e., the average time to complete each of the 12 selections during the selection task and the difference in time between the start of the first stroke and end of the last stroke for writing). Lastly, in pilot studies, many users modified the pressure exerted with the stylus in the *in air* condition. We thus computed the *pressure exerted* (from 0 - 255 units) as well.

3.2 Results

A within-subjects, repeated measures ANOVA with *Hover Cursor* (levels: Cursor Absent, Cursor Present) and *Hand Resting* (levels: On Screen, In Air) was performed. Each task was analysed separately, as the computations for each measure varied between tasks (e.g., movement time averaged across the 12 selections for selection compared to total movement time for writing). For the questionnaire data, responses were encoded on a 7-point Likert scale, with 1 corresponding to "Strongly Disagree", 4 to "Neutral", and 7 to "Strongly Agree". All responses were compared to the neutral response using one-sample t-tests.

3.2.1 Errors

With Hover Cursor, participants made fewer errors when the cursor was present compared to absent (Figure 3a; $F_{1,15} = 23.7$, $p < .001$, $\eta^2 = .38$). Hand Resting was not found to influence error ($F_{1,15} = .3$, $p = .62$, $\eta^2 = .00$) and no interaction was found between Hover Cursor and Hand Resting ($F_{1,15} = 1.9$, $p = .19$). The writing task demonstrated similar results: fewer errors were found for the Hover Cursor factor when it was present (Figure 3b; $F_{1,15} = 9.8$, $p < .001$, $\eta^2 = .18$). Hand Resting was not found to influence error ($F_{1,15} = 0.7$, $p = .71$, $\eta^2 = .00$) and no interaction was found between Hover Cursor and Hand Resting ($F_{1,15} = .4$, $p = .84$). Thus, visual feedback decreases error and helps targeting movements and inking.

3.2.2 Movement Offset

With the Hover Cursor conditions, participants' selections were more accurate when the cursor was present than absent (Figure 3c;

$F_{1,15} = 15.7$, $p < .001$, $\eta^2 = .49$). With Hand Resting, participants exhibited less variability when their hand was in the air (Figure 3c; $F_{1,15} = 5.2$, $p < .05$, $\eta^2 = .03$). No interaction was found between the Hover Cursor and Hand Resting factors ($F_{1,15} = 0.9$, $p = .36$). While visual feedback and hand posture appear important, the difference in effect sizes suggest visual feedback plays more of a role in decreasing error. While writing, the Hover Cursor factor resulted in digits that were significantly more centred when feedback was present (Figure 3d; $F_{1,15} = 44.2$, $p < .001$, $\eta^2 = .50$). Hand Resting had no influence on error, ($F_{1,15} = 0.15$, $p = 0.70$, $\eta^2 = .00$) and no Hover Cursor and Hand Resting interaction was found ($F_{1,15} = 0.25$, $p = .62$). As with the selection results, the stability afforded by the screen was less important than the cursor presence.

Movement Time in Milliseconds

		Selection		Writing	
		M	SEM	M	SEM
Hand On Screen	Hover Cursor Present	900	36	6660	312
	Hover Cursor Absent	1020	53	6481	332
Hand In Air	Hover Cursor Present	1006	47	7649	371
	Hover Cursor Absent	994	40	6945	299

Pressure Exerted

		Selection		Writing	
		M	SEM	M	SEM
Hand On Screen	Hover Cursor Present	27.73	4.32	126.23	8.39
	Hover Cursor Absent	29.48	.34	126.54	9.25
Hand In Air	Hover Cursor Present	19.77	3.42	99.13	7.93
	Hover Cursor Absent	23.75	3.15	107.42	7.06

Table 1. Results for Movement Time and Pressure Exerted.

3.2.3 Pressure Exerted

For selection, Hand Resting influenced the pressure exerted, with more pressure exerted when their hand was on the screen than in the air (Table 1; $F_{1,15} = 8.9$, $p < .01$, $\eta^2 = .23$). The Hover Cursor factor did not influence the pressure exerted ($F_{1,15} = 2.0$, $p = .18$, $\eta^2 = .03$), nor was there a Hover Cursor by Hand Resting interaction ($F_{1,15} = 1.0$, $p = .33$). Hand Resting influenced the pressure exerted while writing, with significantly more pressure being exerted while the hand rested on the screen (Table 1; $F_{1,15} = 8.6$, $p < .01$, $\eta^2 = .29$). The Hover Cursor factor did not influence the pressure exerted ($F_{1,15} = 3.06$, $p = .10$) and a Hover Cursor by Hand Resting interaction was not found ($F_{1,15} = 1.02$, $p = .33$). These results complement the other accuracy measures and suggest that the use of pressure to assess perceived accuracy may be useful.

3.2.4 Movement Time

During selection, the participants were faster when the hover cursor was present than not (Table 1; $F_{1,15} = 4.3$, $p < .05$, $\eta^2 = .07$). Hand Resting did not influence the time taken to complete the task ($F_{1,15} = 1.6$, $p = .24$, $\eta^2 = .03$) and no interaction was found between the Hover Cursor and Hand Resting factors ($F_{1,15} = 7.0$, $p = .06$).

While writing, the Hover Cursor and Hand Resting factors did influence movement time. Participants were faster when allowed to rest their hand than when not allowed to do so (Table 1; $F_{1,15} = 11.7$, $p < .01$, $\eta^2 = .27$). Participants were slower when the hover cursor was present (Table 1; $F_{1,15} = 13.5$, $p < .01$, $\eta^2 = .10$). No interaction was found between the Hand Resting and Hover Cursor factors ($F_{1,15} = 4.2$, $p = .06$). The feedback from the hover cursor appears to slow users down, possibly because they wait for feedback before interacting. When it comes to the speed of inking, the effect sizes suggest that hand posture may be more important than feedback.

3.3 Discussion

The results demonstrated that hand posture plays little role in accuracy on tablets. When the hand is in the air, the increased degrees of freedom from the fingers, wrist, and forearms do not introduce significant instability or inaccuracy, contrary to hypotheses from prior work. The resting of the elbow (and forearm) on the table provided enough stability for participants to perform the necessary movements. Resting the hand on the screen, while useful for writing, was not helpful while selecting targets. This is likely because the larger inter-target distances placed targets outside the natural range of motion of the wrist, hence participants were lifting their hand to complete the task regardless.

Hand resting led to participants to exert 22% more pressure while targeting and 11% more pressure while writing. Although resting the hand should have decreased pressure, these differences may be due to the slipperiness of the nib on the surface. When the hand is not available to stabilize the stylus, the loss of control leads to a loss of exerted force. Participants also demonstrated an 11% increase in speed while writing with their hand on the screen. As holding the hand in the air is uncomfortable and unnatural, participants were much more careful, so as to not have to repeat movements.

Many participants felt it was distracting, fatiguing, or uncomfortable (M = 5.07, SD = 1.83; t(15) = 2.26, $p < .05$) to hold their arm in the air. One participant stated “I never, ever want to buy a tablet that doesn't allow me to place my hand on the screen.” Although a decrease in accuracy was not found, participants felt that they were less accurate when holding their hand in the air (M = 5.93, SD = 1.28; t(15) = 5.85, $p < .001$). We also observed many participants grasping higher on the barrel when their hand was in the air. While many participants were unaware of this, one participant made “a conscious decision as to where [he] held along the pen, (where [his] fingers actually touched the pen, near the bottom, or the middle of the pen).” Such compensatory behaviours could have implications while interacting for extended periods.

The results demonstrated that visual feedback, more so than resting the hand, influences accuracy. The hover cursor enabled participants to be 161% more accurate and 5% faster while targeting. The increased accuracy is likely due to participants using the feedback to assist their targeting movements, and the slight increase in speed is likely a by-product of the latency inherent in the cursor. When feedback was unavailable, users made slower, more thoughtful movements, knowing there was no feedback to support corrective movements. While writing, visual feedback led to a 49% improvement in accuracy and a 6% increase in movement time. When the cursor was present, participants were more likely to capitalize on its feedback, moving cautiously before tapping the screen. When the cursor was absent, users likely realized that realignment would not increase accuracy, hence they moved faster.

Participants reported that they noticed the hover cursor (M = 5.88, SD = 1.61; t(15) = 4.4, $p < .001$) and believed it made them more accurate (M = 4.94, SD = 1.70; t(15) = 2.2, $p < .05$), but were mixed on its usefulness for inking (M = 4.3, SD = 2.12; t(15) = .6, $p = 0.28$) because “writing is allowed to be a messy task to begin with” and “I can make corrections on the fly while writing so some inaccuracy is ok”. While the hover cursor has been viewed as a by-product of inaccuracy, 43% of participants believed it would be helpful, even if the stylus was perfectly accurate. With a traditional pen, the user knows where ink will be deposited, as the ink or lead leaves marking at those points that touch paper. On a digital device, the cursor appears to act as a backup feedback system.

4 EXPERIMENT PART II: INFLUENCE OF NIB DIAMETER

The second part of the experiment explored nib diameter. Although other stylus-based factors, such as nib malleability, barrel contour, and nib transparency could be explored, nib diameter was chosen as it is one of the most prevalent differences between active and passive styli today. Given the diversity of nibs available, from ~1 mm to 7 mm, it is important to understand how they affect user experience. Are they intolerable? Can users adapt to them? Although one can assume the larger the nib, the less accuracy (due to the increased occlusion and resultant uncertainty), is the difference substantial enough to warrant increased engineering to reduce nib diameters? By using identical sensing mechanisms and only modulating the nib diameter, we evaluated the role that diameter has on the perception of accuracy.

4.1 Methods

As in Part 1, ISO 9241 tasks were also used in Part II.

4.1.1 Experimental Conditions

Only one factor, *nib diameter,* was manipulated in the second part of the experiment. As we were unable to create nibs thin, yet strong enough to fit inside the nib chamber, ‘caps’ were 3D printed using non-flexible Polylactic Acid and glued to the end of a standard nib (Figure 4). The printed caps were designed such that they had the same shape and rigidity, varying only in diameter. The diameters were chosen such that the largest cap replicated the diameter of passive styli today (i.e., 5.6 mm); the remaining diameters were distributed within this range (i.e., 4.4 mm, and 3.5 mm). With the caps, the same activation force was necessary to trigger the stylus.

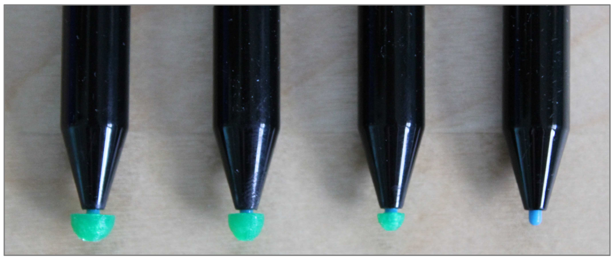

Figure 4. The four nibs used in the Part II of the experiment. From left to right: 5.6 mm, 4.4 mm, 3.5 mm, and 1.6 mm.

4.1.2 Tasks, Procedure, and Measures

Participants completed the same selection and writing tasks as in Part 1. Participants were instructed to rest their hand on the tablet during each trial and the hover cursor was visible whenever the stylus was in the hover state. With the 1.6 and 3.5 mm nibs, the hover cursor and targets were not occluded. Depending on how the 4.4 mm and 5.6 mm styli were held, occlusion was possible. The order of the four nib diameters was counterbalanced, as was the presentation of the task manipulations (i.e., target distances, rectangle width, etc.), and the tasks themselves. Similar to Part 1, the number of errors, movement offset, movement time, and pressure were recorded. Participants also completed 7-point Likert scale questions and ranked the styli according to which were the most preferred and most accurate.

4.2 Results

A within-subjects repeated measures ANOVA, with *Nib Diameter* (levels: 1.6, 3.5, 4.4, and 5.6) as the main factor, was performed for the selection and writing tasks. For the questionnaire data, one-sample t-tests, comparing all responses to the neutral response (i.e., ‘4’), and, where appropriate, comparisons were made against the smallest or largest nib size (i.e., ‘1.6 mm’ and ‘5.6 mm’).

4.2.1 Errors

While selecting, Nib Diameter influenced the number of errors made (Figure 5a; $F_{3,45} = 2.9$, $p < .05$, $\eta^2 = .16$). Bonferroni-corrected paired t-tests determined that participants made significantly fewer errors with the 1.6 mm versus 4.4 mm ($p < .05$) and 1.6 mm versus 5.6 mm nibs ($p < .01$). Nib diameter thus influences accuracy, especially when diameters are atypically large (e.g., 4.4 mm or 5.6 mm). Nib Diameter did not influence the number of digits outside the target rectangles (Figure 5b; $F_{1.3,20.75} = .90$, $p = .39$, $\eta^2 = .06$).

4.2.2 Movement Offset

The larger the nib, the less accurate one’s targeting movements (Figure 5c; $F_{3,45} = 3.9$, $p < .05$, $\eta^2 = .21$). Bonferroni-corrected paired t-tests revealed that participants were significantly more accurate with the 1.6 mm compared to 3.5 mm nib ($p < .05$), the 1.6 mm versus 4.4 mm nib ($p < .01$), and the 1.6 mm versus 5.6 mm nib ($p < .05$). The diameter of the nib is thus important when targeting, with a possible ‘inaccuracy threshold’ emerging at 3.5 mm. Nib Diameter was not found to influence the centeredness of the written digits (Figure 5d; $F_{1.9,29.45} = 1.6$, $p = .24$, $\eta^2 = .09$).

4.2.3 Pressure Exerted

In the selection task, Nib Diameter influenced the pressure exerted (Table 2; $F_{3,45} = 8.9$, $p < .001$, $\eta^2 = .37$). Bonferroni-corrected paired t-tests demonstrated that participants exerted significantly more force with the 3.5 mm versus 1.6 mm nib ($p < .01$), the 4.4 mm versus 1.6 mm nib ($p < .01$), and the 5.6 mm versus 1.6 mm nib ($p < .05$). Although the activation force was identical, participants may have perceived the larger nibs as less accurate, exerting more force.

	Movement Time in Milliseconds				**Pressure Exerted**			
	Selection		Writing		Selection		Writing	
	M	SEM	M	SEM	M	SEM	M	SEM
1.6 mm	883	35	7408	301	22.47	2.21	109.49	6.07
3.5 mm	937	42	7252	340	25.98	2.06	116.98	6.18
4.4 mm	954	36	7581	409	31.88	2.66	116.18	7.64
5.6 mm	990	39	7788	532	33.31	3.16	134.03	7.14

Table 2. Results for Movement Time and Pressure Exerted.

Nib Diameter also influenced the pressure exerted while writing (Table 2; $F_{3,45} = 12.6$, $p < .001$, $\eta^2 = .46$). Participants exerted more force with the 5.6 mm than 4.4 mm nib ($p < .001$), the 5.6 mm versus 3.5 mm nib ($p < .001$), the 5.6 mm than 1.6 mm nib ($p < .01$), and the 3.5 mm versus 1.6 mm nib ($p < .01$). Like the selection task, it thus appears that pressure is used as a compensation mechanism to overcome poor perceived accuracy.

4.2.4 Movement Time

With selection, Nib Diameter influenced the time to complete the task (Table 2; $F_{3,45} = 5.4$, $p < .01$, $\eta^2 = .27$). Bonferroni-corrected paired t-tests found that participants were significantly faster with the 1.6 mm versus 4.4 mm ($p < .01$) and 1.6 mm versus 5.6 mm nibs ($p < .01$). Such results align with the error measures and illustrate the speed and accuracy disadvantages of using larger nibs. Nib Diameter did not influence writing (Table 2; $F_{1.9,28.49} = 0.9$, $p = 0.41$, $\eta^2 = .06$). Combined with the accuracy results, it appears that feedback influences accuracy more so than nib diameter.

4.3 Discussion

The second part of the experiment demonstrated how large the disparity between active and passive styli is. When using a nib similar to a passive stylus (i.e., 5.6 mm) compared to a nib found with an active stylus (i.e., 1.6 mm), participants exerted 48% more pressure, exhibited a 12% increase in movement time, and accuracy decreased by 60%. While writing, nib diameter had little influence on the accuracy measured. It is possible that the hover cursor feedback neutralized the impact of the different diameters. As the 4.4 mm and 5.6 mm nibs occluded the cursor and targets, this is, however, unlikely. While writing, if an initial targeting movement

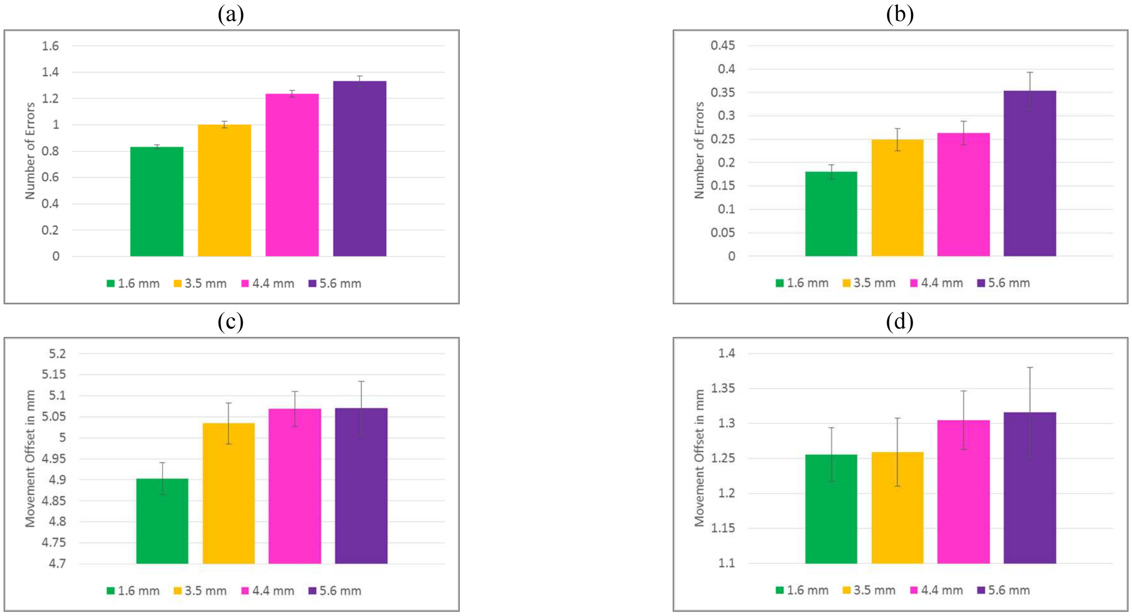
(a)
(b)
(c)
(d)

Figure 5. (a) Error for the selection task, (b) Error for the writing task, (c) Movement offset for the selection task, and (d) Movement offset for the writing task. All error bars represent the standard error of the mean.

was incorrect, there were many opportunities for on-the-fly corrections; with selection, this was not possible.

While the movement time for the 1.6 mm nib (M = 7408 ms, SEM = 301 ms) was longer than the movement time for the Hand On Screen and Hover Cursor Present condition from Part 1 (M = 6660 ms, SEM = 312 ms), we believe this is due in part to natural variability in users' performance, as well as possible carry-over effects from the usage of the other larger nibs in Part II.

Although no accuracy or time differences were found for writing, a significant number of participants believed that the size of the stylus nib influenced their accuracy while inking (M = 5.88, SD = 0.89; t(15) = 8.5, $p < .001$). Participants felt more accurate with the smaller nibs (M = 4.9, SD = 1.4; t(15) = 2.7, $p < .01$). A significant number of participants (i.e., 63%) ranked the 1.6 mm nib as the most preferred (t(15) = 2.8, $p < .01$) and accurate (t(15) = 2.7, $p < .01$) whereas 75% felt the 5.6 mm nib was the least preferred (t(15) = -2.0, $p < .05$) and 63%, the least accurate (t(15) = -2.6, $p < .05$). Even though they were no more accurate, the perception of accuracy plays an important role in the movements made. When asked if a nib thinner than 1.6 mm was desired, none of the participants were receptive to this idea. The lack of desire for smaller nibs contradicts the recommendations of Ren and Mizobuchi's handheld PDA research, but given the limited details provided, we believe these recommendations were unjustified [21]. Although many users indicated they used precise implements daily (i.e., 0.7 mm), different expectations were held for digital devices.

Those who preferred the 3.5 mm and 4.4 mm nibs found the 1.6 mm nib too small (e.g., "the 3.5 and 4.4 mm pens were similar to the markers I use and like") and professed that the 3.5 mm and 4.4 mm nibs appeared identical in size. This visual similarity may explain the plateauing found with the pressure that was exerted. As only perceived accuracy was manipulated, participants unable to discern between the 3.5 and 4.4 mm nibs assumed that they were equivalent, thus they applied the same compensatory pressure.

5 OVERALL DISCUSSION AND IMPLICATIONS FOR HCI

The experiment demonstrated that nib size, visual feedback, and the ability to rest the hand play a role in the speed, accuracy, and pressure with which targets are selected and writing occurs. A comparison between the effect sizes for the three conditions aids in this understanding (Table 3). The feedback provided to the user from the hover cursor appears of highest importance to accuracy. When the cursor was not present or difficult to see (due to nib occlusion), participants were much less accurate and moved slower.

		Hover Cursor η^2	Hand Resting η^2	Nib Diameter η^2
Selection	Error	0.38	-	0.16
	Movement Offset	0.49	0.03	0.21
	Pressure Exerted	-	0.23	0.37
	Movement Time	0.07	-	0.27
Writing	Error	0.18	-	-
	Movement Offset	0.50	-	-
	Pressure Exerted	-	0.29	0.46
	Movement Time	0.10	0.27	-

Table 3. Effect sizes for the Hover Cursor, Hand Resting, and Nib Diameter conditions. (-) denotes the measure was not significant.

Although the hover cursor is a purely digital construct, it is useful. Given the role that the hover cursor plays in accuracy, efforts should be made to increase the height of the hover state and advance hardware such that the hover state exists with capacitive devices. The sooner users receive feedback about the location of the stylus, the better their movements. Given the inaccuracy and legibility complaints found while writing and selecting, it would be fruitful to explore extending precision selection widgets such as CrossY [2] or Pointing Lenses [19] for use in sketching and writing.

Nib diameter was less important than visual feedback, but more important than hand posture. Larger nibs lead to less accuracy during selection and more pressure while selecting and writing, but nib diameter does not influence accuracy linearly. Thus, there could be 'accuracy plateaus', wherein decreases in accuracy and increases in pressure only accompany visually indistinguishable nib diameters. As only four nib sizes were tested, this cannot be confirmed definitively. Such results, however, have implications for multi-pen systems or styli with interchangeable, multi-surface, malleable, or chiselled nibs. Will users unintentionally exert different pressure with different sides of the stylus or nib geometries? Should we compensate or harness this behaviour to obtain contextual information? Although our exploration was confined to nib diameter, these other facets also warrant attention.

The results favoured visual feedback and nib design, but concerns over fatigue and comfort should not be discounted. When asked if they would prefer i) an accurate pen but be prevented from resting their hand or ii) an inaccurate pen but be allowed to rest their hand, 63% of participants favoured hand resting. Interestingly, our population of users were willing to accept inaccuracy if it meant they could interact more naturally. While using a tablet in one's lap, holding it in the air, or interacting for longer durations, fatigue will eventually set in and cause some degree of inaccuracy. Understanding the fatigue/inaccuracy relationship is thus important and provides continued motivation to solve unintended touch and develop surface textures that support smooth hand motions.

Pressure is often harnessed for pressure-sensitive user interface widgets or stroke beautification. Determining the pressure applied when the nib size changes or hand rests, leads to interesting questions for future research. Is consistent ink rendering desired when different styli are used? Should compensatory adjustments be made when the hand is in the air and increased pressure is likely to be applied? As pressure appears to be a compensatory adaptation, it is thus important to consider the implications of pressure beyond accuracy (e.g., for unintended touch, biometrics, and so on).

Within the present work, visual feedback, nib diameter, and hand posture were evaluated using tasks from the ISO standards. While these tasks are limiting in that they do not examine spontaneous content creation, the results are representative of the 'best case' situation. By the very design of the tasks, they can be generalized to more complex selection and free-form input scenarios. It should be noted, however, that the presence of additional cognitive load or alternative task characteristics such as motivation, loci of attention, and hand movement patterns, would likely influence the degree to which feedback, diameter, and hand posture influence one's accuracy. The continued exploration of such factors and use of other stylus-based activities are thus fruitful areas for future research.

As we have demonstrated, many factors influence accuracy on tablets today. While advancements to underlying stylus technologies should continue, the present exploration identified that much more attention needs to be devoted to the design of the stylus and the user. Given the diverse and complex nature of accuracy, there is much room for future innovation and exploration, especially when performance, speed, and comfort are considered. Inaccuracy can no longer be thought of as simply a by-product of poor hardware decisions or limitations. Manufacturers must consider how technology supports natural behaviour, and developers must understand how users compensate for hardware limitations.

6 CONCLUSION

Inaccuracy has long been a complaint of users while interacting with a stylus. It is often assumed that inaccuracy is only caused by hardware-based factors such as visual parallax. The present work, however, determined that many factors contribute to inaccuracy and require further exploration. Using ISO 9241 selection and free-form input tasks, a subset of these factors (i.e., nib diameter, the feedback provided by the hover cursor, and the ability to rest one's hand) were evaluated. The results found that the visual feedback from the hover cursor is most helpful to decrease accuracy errors, and that the diameter of the nib (and users' perception of a stylus) influenced the pressure exerted and accuracy. Although increasing comfort, the ability to rest one's hand on the screen did not influence accuracy. The present exploration thus provided rich information regarding the factors influencing accuracy, and helps to explain the suboptimal experiences found with styli and tablets today.

REFERENCES

[1] M. Annett et al. The Pen is Mightier: Understanding Stylus Behavior While Inking on Tablets. *Proc. of GI*, 2014, 193-200.

[2] G. Apitz & F. Guimbretiere. CrossY: A Crossing-Based Drawing Application. *Proc. of UIST*, 2004, 3-12.

[3] S.K. Badamet et al. Tracing and Sketching Performance using Blunt-tipped Styli on Direct-Touch Tablets. *Proc. of AVI,* 2014, 193-200.

[4] N. Bernstein. *The Coordination and Regulation of Movements*, 1967. Oxford: Pergamon Press.

[5] W.S. Buxton. A Three-State Model of Graphical Input. *Proc. of CHI*, 1990, 449-456.

[6] L.G. Carlton. Visual Information: The Control of Aiming Movements. *Quarterly Journal of Experimental Psychology*, 1981, 33(1), 87-93.

[7] R.S. Goonetilleke, E.R. Hoffman, & A. Luximon. Effects of Pen Design on Drawing and Writing Performance. *Applied Ergonomics*, 2009, 40, 292-301.

[8] T. Grossman et al. Hover-Widgets: Using the Tracking State to Extend the Capabilities of Pen-Operated Devices. *Proc. of CHI*, 2006, 861-870.

[9] C. Holz & P. Baudisch. Understanding Touch. *Proc. of CHI*, 2011, 2501-2510.

[10] ISO/TS 9241-9:2000E: *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices*. International Organization for Standardization, 2000.

[11] ISO/TS 9241-411:2012(E): *Ergonomics of human-system interaction-Part 411: Evaluation methods for the design of physical input devices*. International Organization for Standardization, 2012.

[12] D. Lee et al. PhantomPen: Virtualization of Pen Head for Digital Drawing Free from Pen Occlusion & Visual Parallax. *Proc. of UIST*, 2012, 331-340.

[13] I.S. MacKenzie, T. Kauppinen, & M. Silfverberg. Accuracy Measures for Evaluating Computer Pointing Devices. *Proc. of CHI*, 2001, 9-16.

[14] F. Matulic & M.C. Norrie. Empirical Evaluation of Uni- and Bimodal Pen and Touch Interaction Properties on Digital Tabletops. *Proc. of ITS*, 2012, 143-152.

[15] A. Mohr, D.Y. Xu, & J. Read. Evaluation of Digital Drawing Devices with Primary School Children - A Pilot Study. *Proc. of ICL*, 2010, 830-833.

[16] T. Nescher & A. Kunz. An Interactive Whiteboard for Immersive Telecollaboration. *The Visual Computer*, 2001, 27(4), 311-320.

[17] A. Ng et al. In the Blink of an Eye: Investigating Latency Perception during Stylus Interaction. *Proc. of CHI*, 2014, 1103-1112.

[18] E. Park, K.J. Kim, & A.P. del Pobil. Does Length Matter? A Study Examining How Length of Stylus Pen Helps Effective Electronic Documentation. *Proc. of ICIS*, 2011, 185-188.

[19] G. Ramos et al. Pointing Lenses: Facilitating Stylus Input through Visual- and Motor-Space Magnification. *Proc. of CHI*, 2007, 757-766.

[20] J.C. Read. The Usability of Digital Ink Technologies for Children and Teenagers. *People & Computers XIX-The Bigger Picture*, 2006, 19-35.

[21] X. Ren & S. Mizobuchi. Investigating the Usability of the Stylus Pen on Handheld Devices. *Proc. of SIGHCI*, 2005, 12.

[22] M. Sun & X. Ren. An Evaluation of Multimodal Feedback in Tracking State for Pen-Based Interfaces. *Proc. of IEEE Mechatronics and Automation*, 2009, 72-77.

[23] M. Sun et al. An Investigation of the Relationship between Texture and Human Performance in Steering Tasks. *Proc. of APCHI*, 2012, 1-6.

[24] D. Vogel & R. Balakrishnan. Occlusion-Aware Interfaces. *Proc. of CHI*, 2010, 263-272.

[25] D. Vogel & R. Balakrishnan. Direct Pen Interaction with a Conventional Graphical User Interface. *Human-Computer Interaction*, 2010, 25(4), 324-388.

[26] J.R. Ward & M.J. Phillips. Digitizer Technology: Performance Characteristics and their Effects on the User Interface. *Computer Graphics and Applications*, 1987, 7(4), 31-44.

[27] F. Wu & S. Luo. Performance Study on Touch-Pens Size in Three Screen Tasks. *Applied Ergonomics*, 2006, 37, 149-158.

[28] Xda Developers Forum. forum.xda-developers.com/showthread.php?t=2171198. September 2014.