Master of Science in Informatics at Grenoble
Master Mathématiques Informatique - spécialité Informatique
option GVR

Digital Ink vs. Physical ink
User Perception Limits when Writing with a Digital Pen

Elie Cattan

June 23rd, 2014

Research project performed at LIG

Under the supervision of François Bérard, Grenoble-INP

Defended before a jury composed of:
Prof James Crowley
Dr Olivier Aycard
Prof François Faure
Prof Jean-Claude Fernandez
Dr Caroline Jolly

June 2014

Abstract

Despite the spread of tablet computers (or “tablets”), people still take notes with physical pens on paper. Yet, tablets are a good tradeoff between laptop computers and physical paper. They are lighter and more suitable for complex annotation than computers, and they do not suffer from the limitations of physical paper: tablets are dynamic, they allow easy duplication, distribution and archival.

This work is an attempt to understand why paper is till preferred to tablets when it comes to taking notes. We worked with the hypothesis that the technical performances of digital pens are not good enough to replace physical pens. Our approach was to create an experimental high performances digital pen and selectively degrade various aspect of our pen. We ran two user studies to measure the effect of these degradations on the users of the pen.

In a subjective experiment, we collected the general feelings of participants when writing with a physical pen on paper, a commercial tablet and our device. Contrary to our expectations, we discovered that the smooth contact of a plastic nib on a glass display was not a hindrance. In an objective experiment, we revealed the thresholds of human perception on various technical characteristics of our digital pen. In particular, we showed that current commercial tablets do not satisfy the requirements of human perception in term of offset and latency.

This work provides new knowledge on digital pens requirements, which should contribute to the ultimate goal of providing usable digital services to the task of note taking.

Contents

— 1 —
Introduction

In 2013, almost 200 million tablet computers (or "tablets", in short) were sold[1]. More and more people own tablets and use them in many situations such as taking photos, browsing the Internet, and reading. Tablets can be used with a digital pen to facilitate note-taking. However, when it comes to taking notes during a presentation or a lecture for example, most people who own a tablet don't use them. Either a laptop computer is preferred when the notes are mostly textual, or users revert to physical ink on paper when the notes require the drawing of complex shapes such as boxes and arrows, diagrams, or mathematical curves and formulas. This may be explained by the inability of tablets to reproduce some of the desirable properties of drawing ink on paper. Guimbretiere provides the following analysis of the affordances and limitations of paper [9]: paper is light, flexible, and easy to annotate. Paper is also easy to navigate and it provides large, inexpensive, high-resolution displays. But, paper is static and requires efforts to duplicate, distribute and archive.

A tablet with a stylus appears as a good tradeoff between a laptop and paper to take notes. Tablets are lighter than computers and are more suitable for complex annotation. Tablets do not suffer from the limitation of paper: they are dynamic, allowing easy duplication, distribution and archival. In this context, why is paper still largely preferred than tablets when it comes to taking notes? The work presented here is an attempt to provide elements of answer to this question.

In this report, we use the term "inking" to refer to writing and drawing; we use the expression "physical pen" to refer to normal pens which draw *physical ink on paper*; and we use the expression "digital pen" to refer to systems that include a display surface and a pen and produce digital ink[2]. The main hypothesis of our work is the following: current digital pen technologies do not emulate physical pens well enough that they would allow efficient and comfortable inking. As a result, people prefer to use physical pens for inking, even if doing so they lose all the benefits of the digital world.

Our objective is thus to determine the limitations of digital pens, with the idea that problems can only be solved once they are well identified and understood. The main principles of our approach are to create an experimental system that implements a digital pen which is as close as possible to physical pens, and then selectively degrade the peformances of our system and measure the effect of these controlled degradation on the usability of the system. We are looking in particular to identify which technical characteristics of digital pens are key to usability,

[1]http://www.gartner.com/newsroom/id/2674215

[2]The Anoto pen [11], for example, is not considered a digital pen because it produces physical ink even though it also includes digital services.

and what level of performance, for these key characteristics, are required to provide adequate usability of digital pens. As we are focusing on the inking task itself, and not the environment where the task takes place, we can disregard some of the paper affordances, such as lightweight and mobility, and focus on the fidelity of the physical inking on paper reproduction.

We can thus divide our approach into three steps:

1. Create an experimental system that implements a high performance digital pen.
2. Assess that the quality of the system is close to the one of physical pens.
3. Selectively degrade various technical performances of our system, such as latency or sampling frequency, and evaluate in a user experiment the effect of the degradations on the system's usability.

This report is organized as follow. We first provide a review of the state of the art in various domains which are related to our problem: digital augmentation of physical writing, human writing productions, and human limits of perception. We then present our approach in detail. This is followed by a presentation of our experimental system and the evaluation of its performances. The system was used to conduct two user experiments that we present in the following chapter. Finally, we provide a discussion of this work and conclude.

— 2 —
State of the art

The inadequacy of digital pens for inking has been recognized for a long time, we review previous work on pen and paper in the digital work in the first section of this chapter. Our approach requires that we are able to compare the quality of our digital pen at its various levels of performance. One way to do this is to compare the quality of the users' productions. The second section is thus a literature review on ways to quantify the quality of writing and drawing productions. Another approach is to check at which level the performance variation of our digital pen are perceptible by users. We thus review in the final section a set of studies that are focused on human perception thresholds.

2.1 Pen and paper in the digital world

Researchers have recognized for a long time that users were not able to be as efficient in drawing and writing with digital pens compared to physical pens. While our work is a contribution to the improvement of digital pens, another approach is to keep using physical pens and attempt to *augment* them with digital capabilities. More than 20 years ago, Wellner pioneered this idea with the Digital Desk [26]. A standard physical desk, supporting pen and paper, was augmented with digital informations. As a physical desk, the Digital Desk keeps all the good properties of paper, but digital informations are also brought on the desk through projection. Users interact with this digital ink with bare fingers or physical objects such as sheets of paper and erasers.

This idea of keeping physical pen and ink and adding digital information led to a lot of research. Bérard developed the "magic table" [5] which is a multi-user whiteboard mixing both physical and digital ink. It enables people to write on a whiteboard with a normal dry-erase pen and then, after a scanning, physical ink is changed to digital ink and can be easily transformed (moved, rotated, enlarged).

Augmenting the physical ink is also possible by recording the trajectory of the pen. The Anoto pen [11] can track the position of its nib with great accuracy on a special dot-marked paper. It can then send the stroke trajectories to a computer where they will be handled by an application. Based on the Anoto tracking system, Guimbretiere created PADD [9]. PADD enhanced the cohabitation between the paper in the physical and digital world. It facilitated the annotation of paper with a pen in the physical world, switch to the digital world where it is easier to make correction (adding a word, correct a fault...) and switch back to the physical world again.

Research projects such as PaperWindows [10] or XLibris [19] do not use physical ink at all. In PaperWindows [10], a dry pen is tracked with optical markers and digital ink is projected on the paper from above. Digital manipulations such as “copy-paste”, “scroll” or “save” are then possible by manipulating the sheets of paper. XLibris [19] is based on a tablet computer. To be as close as possible to physical pen and paper, it uses the *paper document metaphor*. The display, the interface and the interaction technics are emulating the appearance of a sheet of paper. It also goes beyond paper because it uses the annotations made by the users to search and organize information.

Except XLibris [19], which works with a tablet computer, the other projects really aim to keep the paper and enhance it with digital properties while keeping physical pen.

In our project, we are studying the technical performances of digital pens, and we need to control these performances in realtime. As a consequence, our device cannot produce physical ink. We note, however, that using a projector allows to reproduce the feeling of the pen-nib and paper contact.

2.2 Evaluation of handwriting and drawing

Our approach involves comparison of the quality of various devices (or a single device with various parameters) for inking. These comparisons can be done on the *productions* achieved with the various devices. We are focusing on drawing and writing as representative tasks of note-taking. We are not aware of any work that specifically compare physical and digital ink in a quantitative manner. This illustrates the novelty of our approach. There has been research efforts, however, which attempted to quantitatively compare physical handwriting or drawing. We first review the litterature focused on the drawing task. The simplest analytically expressed shapes are polygons. Arkin et al. defined a metric that measures the similarity of polygons [2] and is invariant under translation, rotation and change of scale. The polygons are described with a function as a serie of angles, the distance between two polygons is then computed by integrating the difference between the two polygons function. Alt et al. [1] offered a mathematical solution to the problem of polygon matching by finding the isometry which minimize the Hausdorff distance between the two shapes. Other research efforts used the Hough transform to match simple line drawings [3, 8], or the angular signature of the strokes [24]. However, these drawings comparison algorithms are used to differentiate different models, i.e. drawings that are quite different. We did not find any research efforts on the comparison of drawings based on similar drawings that reproduce the same model.

Focusing now on the task of writing, Rosenblum et al. [18] offered a review of the methods to compare handwriting to measure the handwriting capabilities of children. They found two ways of handwriting evaluation. The first is a global evaluation. An evaluator has to compare a production with a given scale of writing samples and find the closest one. The second is an analytical evaluation: based on some criteria as size, slant, spacing, shape or general merit, the evaluator has to put grades on each criteria. Rosenblum et al. pointed out a set of methodological issues in handwriting evaluation. There are a lot of criteria such as the evaluator, the instructions, the writing speed measurement or the writing accessories which are not well defined and that limits comparison between studies.

To summarize, drawing comparison metrics were only studied in the case of different models, and may not be accurate enough to reveal a device effect. In addition, research on handwriting quality shows that despite agreement on assessment criteria, methodological issues are too important to define an irrefutable protocol. This review raises a question: are drawing or handwriting comparison approaches accurate enough to make the difference between a production done on two different devices? We will present in Section 3.3 our efforts and results on this matter.

Moreover, production comparisons only focus on the results. But feelings when inking are also very important in order to assess the comfort of the tool.

2.3 Perception, Just Noticeable Difference (JND)

When comparing inking devices, instead of comparing the productions of the devices, we can aim at a more direct comparison of the inking action itself. In other words, we can ask users which device they prefer. However, the differences between devices may be subtle. Asking users directly what aspect of a digital pen is inferior to a physical pen may not provide a detailed nor accurate picture of the digital pen limits. More robust protocols are required to estimate what differences in performance are perceived by users.

Ng et al. focused on one particular parameter of input devices: their *latency*. The latency of an input device can be defined as the delay between the time where the device is positionned at a particular location, and the time when the graphical feedback is displayed at this location. Ng et al. worked on finding the limit of perception of latency in touch devices [17]. They used a protocol, first introduced by Kaernbach [13], which is called the Just Noticeable Difference (JND). The JND reveals the smallest difference in the tested parameter that users can perceive with at least 75% accuracy. Using the JND protocol, Ng et al. revealed that users are able to perceive latencies at very low levels: 6ms for a dragging task with the index finger.

In a very recent paper, Ng et al. [16] focused again on user perceivable latency, but this time for a digital pen used in various tasks, including a "scribbling" task that shares similarity with the "inking" tasks that we focus on. They implemented a very low latency digital pen: its latency was estimated at 1.4ms for dragging tasks and 7ms for scribbling. Using the JND protocol again, Ng et al. found that people were able to perceive latency down to 2ms (respectively 6ms) when dragging a small (respectively large) box and down to 40ms when scribbling. One important result of this study is that the level of perceiveable latency is strongly dependent on the task. When scribbling, participants had to maintain the shape and the speed of stroke at each trial. This was more cognitively demanding than dragging a box and may have distracted participants from the perception of latency. The study also revealed that currently available digital pens, which latencies are estimated in the range 50-200ms [17], do not satisfy the requirement of immediacy in terms of Human perception. Perceiving latency does not mean that the performances will be necessarily degraded, but the negative effect of latency was already noted [12, 15]. Moreover, even though it does not affect performances, it could also affect the comfort when executing the task.

In our study, we will use the JND protocol, as detailed in Section 3.4, to study the limit of user perception of latency while inking, but also the percetion limits of other parameters detailed in Section 3.2.

— 3 —
Approach

In this chapter, we detail the approach to understanding the limits of digital pens that we outlined in the introduction.

3.1 Main principles

We make the hypothesis that physical pens are favored to digital pens for note-taking because digital pens do not reproduce some of the key characteristics of physical pens. Our main objectives are thus to identify these deficient characteristics, and to quantify them. To achieve these objectives, our strategy consist in selectively degrading some *characteristics* of an *ideal digital pen*, and quantify the resulting degradation, if any, of the usability of the pen. This strategy present several difficulties, as detailed below.

We need to have an ideal pen on which we can selectively degrade some characteristics in a controlled manner. Any physical pen is considered an ideal pen for our purpose, but there is not practical way to degrade, for example, the latency or the spatial resolution of a physical pen. We thus chose to create a digital pen which is identical, or as close as possible, to a physical pen. While commercial digital pens have failed to achieve this goal up to now, we don't have the same constraints as commercial products. In particular, we can implement our experimental digital pen with the use of heavy, expensive, but high performance equipment. In addition, our pen can work in a very controlled environment tailored to our experiments. The design and implementation of our experimental pen is presented in the next chapter "A high performance digital pen".

We need to select the set of characteristics of our digital pen that we will control and degrade in order to evaluate their effect of the pen's usability. This choice is presented in the following section "Technical characteristics of a digital pen".

We need to quantify the usability of a pen. The usability of a pen can be assessed from the quality of users' productions using the pen. In the section "Evaluating inking productions", we discuss our efforts in this domain, and the difficulties that we encountered. These efforts led us to a re-orientation of our usability evaluations.

Eventually, we decided to assess the quality of our digital pen based on users ability to perceive its technical performances. This is presented in the section "Perception limits of technical performances".

3.2 Technical characteristics of a digital pen

Our approach requires that we degrade some of the parameters of an ideal pen to measure the effect on usability. But what are the key parameters of a pen that influence its usability? We identified the four following parameters based on discussion with several researchers. This procedure, however, does no guarantee that we did not miss an important parameter. Hence, this work must be considered as a "best effort". In any cases, our work provides a quantification of the required performance for the parameters that we considered. In the Section 6.1 we discuss other parameters that could have influenced the quality of a digital pen.

- The contact of the nib of the pen on the surface: on paper, with a pen or a pencil, the paper grain hooks a bit of ink or graphite from the nib. This provides resitance to the motion of the nib and gives a very particular feeling when writing. On a tablet, the pen nib is most of the time made of resin or plastic which slips with almost no resistance on the smooth glass surface. This may require more effort to precisely control the motion of the nib and is sometimes perceived as unpleasant.

- The offset: on paper, ink always appears under the pen nib. But on a tablet, a bad calibration or the parallax generated by the thickness of the glass may create an offset between the position of the nib and the position of the digital ink on the screen.

- The latency: various processings are required to generate digital ink, such as pen sensing, data transmission, and graphical rendering. All these processings take time to perform. As a result, ink does not appear immediately where the nib is moved, but after a small delay: *the latency* of the system.

- The spatial resolution: the resolution of a tablet sensor is the minimal nib motion that can be detected. If the resolution is not high enough, users may have trouble aiming at precise location, and smooth motions of the pen generate curves that are not smooth, even at low motion speed.

- The temporal resolution: the digital world is not continuous but discretized. That implies that the tablet receive events at a given frequency. If this frequency is too low and the user is fast, for example, smooth curves of the digital pen generate broken lines.

The first parameter, the nature of the nib-surface contact, has a different nature compare to the four others: it is not continuous. To find out if people prefer writing on a very smooth surface rather than on a grainy one, we designed and ran a specific experiment where we tested both conditions.

For the four other parameters, we consider a physical pen to have optimal performances: there is no offset and no latency, temporal and spatial resolution are infinite. With a digital pen, these parameters are never optimal due the discrete nature of digital computation. However, Human perception is not infinite. We can expect that at some very high level of temporal and spatial resolution, for example, Humans are not able to perceive the discontinuities. Therefore, for each of these parameters, we want to measure what is the level of performance at which users are not able to perceive a difference with the ideal pen.

3.3 Evaluating inking productions

One way to evaluate the quality of a pen is to evaluate the inking productions (writings or drawings) made by users of the pen. If we are able to define a quality metric for the productions, then we can compare the productions of our digital pen with an ideal physical pen. We can also compare the productions of our digital pen with its nominal performances and with some degraded performances.

We have seen in Section 2.2 of the state of the art that several protocols have been proposed to measure the quality of a handwriting productions. This illustrates the difficulty of handwriting evaluating. Testing and comparing the various approaches would have required a large effort that would go far beyond the scope of this master project.

We thus chose to focus on a quality metric for drawings, which seemed more tractable. The quality of a drawing can be judged in many ways. We are interested in how users would judge de quality of the drawings. Therefore, we decide that our metric should be validated by comparing its results with the results of a Human evaluation.

We first created a data set. We worked with the hypothesis that the quality of a drawing depends not only on the device used for the drawing, but also on the time used to perform the drawing. We could thus easily create a data set of the same model reproduced at various quality levels simply by controlling the time allowed to do the reproduction. The same ideal device, a normal pencil on paper, would be used for all drawing durations. This data set would allow a first validation of our metric: drawing performed with more time should have a higher quality.

We ran an experiment where a participant had to copy, on a sheet of paper, a line drawing model of a mouse using and a pencil with different time limits: from 3s to 30s with a 3s time step. The maximum of 30s was chosen because it was approximately the time needed to copy the whole model at a normal speed: the participant did not have to draw anymore at the end of the 30s. Below 30s, the participant was still drawing at the end of the timer. The Table 3.1 illustrate the results.

We then asked ten evaluator participants to order these drawings from what they judged to be the closest to the model, to the farthest from the model. We computed an agreement score between the evaluators: 100% corresponds to a total agreement accross evaluators, 0% corresponds to total disagreement, and 50% corresponds to randomness. Considering the ten drawings, from 3s to 30s, the evaluator participants agreed at 84%. However, for the four drawings done with the largest durations (21s to 30s), the ten evaluators were unable to agree on an order: the agreement score was 41% which shows randomness.

When the drawing duration is very constrained (i.e. small durations), it is possible to consistently rank the drawings. But we are expecting small variations in the production quality as we want to control small variation in the digital pen performances. Small variations are difficult to rank: our experiment demonstrated that human evaluators were not able to consistently rank the productions when the increase of drawing duration was only used for subtle improvements. As the ranking algorithm we are looking for is one that corresponds to the Human perception of drawing quality, we cannot expect to "beat" Human evaluators. Indeed, we implemented an algorithm that had similar agreement score compared to Human evaluators. Its agreement score with Human evaluators was in the range of 80% overall, but its agreement was in the range of 40% (randomness) for the most subtle drawings (i.e. the four longest durations).

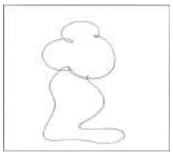
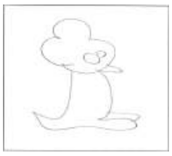
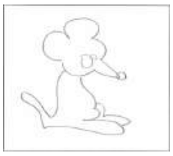
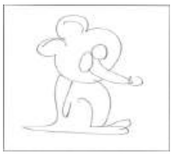
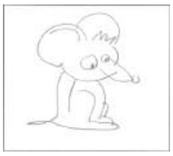
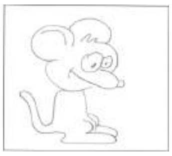
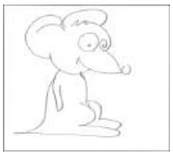
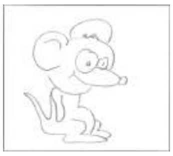
Model					
Time	3s	6s	9s	12s	15s
3s - 15s					
Time	18s	21s	24s	27s	30s
18s - 30s					

Table 3.1: Drawings of a participant using a physical pencil on paper, under various drawing durations.

This experiment demonstrated that ranking the quality of drawing production was probably not the right path to follow for our problem. We thus re-oriented our approach towards the Human perception limits of the digital pen performances.

3.4 Perception limits of technical performances

Estimating the quality of the productions does not seem to provide a manageable way to estimate the effect of various pen characteristics. We thus chose a more direct way which is founded on users' perception of the pen quality. In other words, we will selectively degrade some technical performance of our digital pen, and ask participants if they perceive the degradation. These degradation can be subtle, for example will participant perceive a temporal resolution degradation from 200Hz to 180Hz? Participant answers always have a subjective component. However, we introduced the Just Noticeable Difference (JND) protocol in Section 2.3 of the state of the art. The JND allows to measure an objective threshold of the user ability to perceive a difference in two stimuli. Replaced in our main approach, the JND works as follow: if users are not able to perceive the performance difference between our digital pen and an ideal pen, then this difference should have no effect on users' performance in inking.

Just Noticeable Difference

The JND protocol is based on a weighted up-down method first introduced by Kaernbach [13]. It is designed to measure the minimal difference between two stimuli that participants can perceive. The two stimuli are called the "baseline" and the "probe". When the difference between the baseline and the probe gets close to a threshold between a perceivable and an imperceptible difference, participants ability to distinguish the two may vary accross trials.

The JND handles this participant variability in a statistical way: it computes a threshold where the participant was able to correctly identify the baseline from the probe 75% of the time.

The baseline remains constant during the experiment, the probe depends on the results at each trial. At each trial, the baseline and the probe are randomly presented to the participant, so that the participant has no way to *guess* which is which. Participant are then asked to vote for the stimuli that they think is the baseline. If the answer is correct, the baseline-probe difference is supposed to be above a perceivable threshold, and the probe is moved closer to the baseline by one step. If the participant incorrectly choses the probe as the baseline, then the difference is considered imperceptible, and the probe is moved away from the baseline by three steps. Participant are never told if they were correct or not in their choices. This enables an objective measure of the perceivable threshold. A session ends when a fixed number of *reversals* have occured: a reversal is a switch of probe direction. The threshold is computed as the average of the probe's values at each reversal. This one-step vs. three-step update of the probe converges to a threshold where the participant is correct at 75%. This process is illustrated on Figure 3.1.

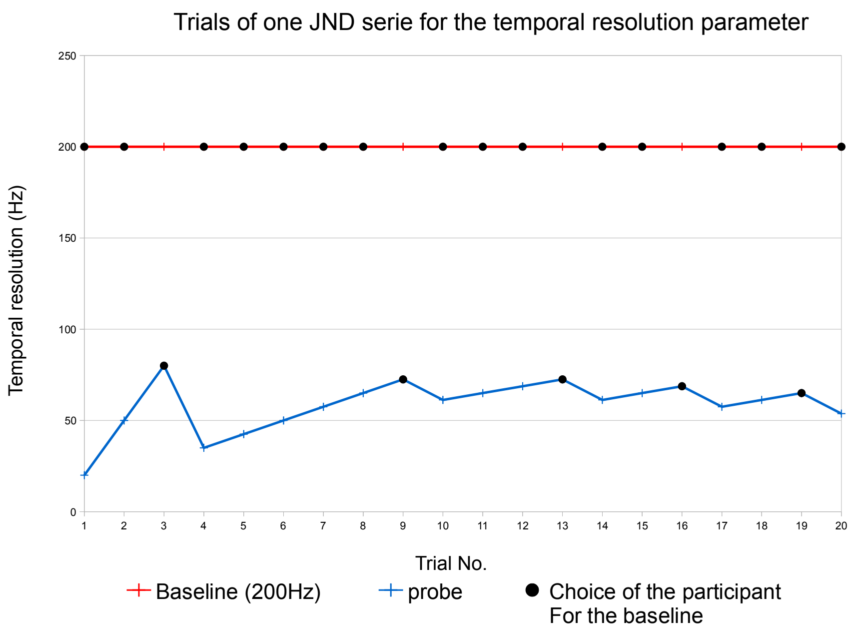


Figure 3.1: JND trial series of one participant for the temporal resolution parameter. On correct identifications of the baseline, the probe is moved closer by one step, otherwise it is moved away by three steps.

Jota et al. proposed an improvement of the JND in order to accelerate its convergence [12]: the step used for the probe update is large at the beginning of a session, but it is divided by 2 at every probe reversal until it reaches a minimum step size defined by the experimenter.

JND for absolute perception thresholds

The JND protocol was designed for the measurement of the *minimal perceivable difference between two stimuli*. In particular, its was not specifically designed to measure an *absolute threshold of Human perception*. However, if the baseline is well below the threshold of Human perception, then the threshold measured by the JND is actually the absolute threshold of Human perception that we are looking for. This is illustrated on Figure 3.1: the end of the probe's blue curve shows that the participant is able to distinguish temporal resolutions which differ by around 25Hz (i.e. between 50Hz and 75Hz). But the baseline-probe difference remains at a much larger level (more than 100Hz). This shows that the perception bottleneck is not the proximity of the baseline and the probe, but rather an absolute perception limit.

As a consequence, using the JND protocol to measure an absolute threshold of Human perception requires that the baseline is *below* this threshold. In the physical world, some baselines can be null, such as the latency when drawing physical ink on paper. In such cases, the baseline is clearly below the threshold of Human perception. Hence we could consider to use a physical pen as our baseline, however this is not compatible with the JND protocol. The protocol requires that participants have no means to guess which is the baseline among the two stimuli; and asking participants to switch from physical to digital would clearly tell them which one is the baseline. We thus used our digital pen for both the baseline and the probe in our JND experiments. As a consequence, we need to check that the JND observed in the experiments represents an actual absolute threshold of Human perception, and not simply the JND with an *above the threshold* baseline. A clear indication that the baseline is below the threshold of Human perception is a JND that is far greater than the baseline (such as in Figure 3.1). On the contrary, if the observed JND is close to the baseline and if the probe varying parameter sometimes reach the baseline, then it can be explained simply because the baseline was too close or even above the threshold of Human perception.

To summarize our approach, we will run a set of user experiments following the JND protocol in order to reveal the thresholds of Human perception concerning four important technical parameters of digital pens: the nib-ink offset, the latency, and the spatial and temporal resolutions. We will also run a subjective user experiment concerning the nib-surface contact. This experiment requires the availability of a high performance digital pen, this is the topic of the next chapter.

— 4 —

A high performance digital pen

In the previous chapter, we outlined our need for a digital pen which performances are as close as possible as those of a physical pen, that we consider ideal. In this chapter, we explain the design choices for our high performance digital pen. We detail our approach to optimize its technical performances, then we explain how we proceeded to selectively degrade its performances in a controlled manner.

4.1 Apparatus

Display

Steimle provides a survey of the technology used in "Pen-and-paper user interfaces" [23]. He lists different ways to display a digital output on paper. Using any form of screen (such as a CRT or LCD) introduces a glass or plexiglass layer between the digital ink and the pen's nib. This layer creates a spatial offset between the nib and the ink, and the nib-layer contact as very little friction, unlike a physical pen on paper. This approach is not suitable for our problem. Electronic paper looks and behave like physical paper, but it has the digital ink embedded in the paper [6]. The update rate, in the order of a second, is too low for our purpose.

We thus chose overhead projection as our display solution. Projection allows to use a normal sheet of paper as the inking surface, and thus to had a nib-paper contact similar to the one of physical pens. Projectors tend to be heavy and are thus usually static. We lose some of the affordances of paper such as light weight, and flexibility. Researchers have worked on mobile projectors associated with pen and paper systems to avoid these limitations [20, 21]. It is also possible to track a sheet of paper and adapt the projection in realtime so that the projected ink remains on the paper, as done by Holam et al. with their PaperWindows system [10]. This approach, however, requires that the projector covers a large area so that the paper sheet motion is not too restricted. In both cases, the mobility comes at the cost of display resolution. As mobility is not required in our case, we chose a fixed projector.

For the input devices, we consider 3 different solutions: Anoto, optical tracking, and pen tablet.

Anoto

Anoto [11] is a commercial product and, as such, is a mature technology, It is used in most current digital pen and paper systems [22, 9, 14]. Anoto presents the strong benefit of perfect inking because it uses physical ink. Thanks to a dot pattern on the paper, a camera can track the pen position on paper and the strokes trajectories are digitized in high resolution.

We could not use a classical Anoto pen as we need to control the performance of the pen, which is not possible with physical ink. We considered using only the pen tracking feature of the Anoto pen, without drawing physical ink. But it requires a special paper and it can only capture pen traces at a 75Hz [23]. We anticipated that this maximum temporal resolution could be above the Human perception threshold, which would defeat our experiment, as explained in Section 3.4.

Optical tracking

We thus considered using an optical tracking system to track the position of a physical pen. The main benefit of a general optical tracking system is flexibility: optical markers can be attached to any object (i.e. any pen) and the system reports about the 6 dimensional poses of the object (3 translations and 3 rotations).

We experimented with this approach: we used a Natural Point Optitrack system[1] with 4 Flex-13 cameras. We setup the cameras to capture a small volume in order to maximize the spatial resolution. In our setup, the system reported position at sub-millimeter precision. The system temporal resolution was 120Hz. We experimented with a pencil whith the nib covered by tape to prevent physical drawing. We also experimented with a dry erase marker which was cleaned up to remove its ink. We calibrated the transformation between tracking coordinates and display coordinates, we were able to accurately align the nib with the projected ink, as illustrated in Figure 4.1. The 120Hz sampling of the tracking, coupled with the 60Hz update of the projector allowed for a small latency.

The main problem of this approach is detecting if the pen nib is touching the surface or not. This vertical motion of the nib can be very subtle, but its detection needs to be very stable because it defines if digital ink should be layed on the paper or not. With the optical tracking approach, we did not succeed in achieving enough accuracy and stability on this particular problem.

The final system based on a digitizing tablet

As neither Anoto nor optical tracking appeared suitable for our problem, we chose to use a digitizing tablet as in the Docudesk system of Everitt et al. [7]. We put a Wacom Intuos 3 under the sheet of paper, and used the Intuos pen as our digital pen. The main drawback of this approach, compared to the optical tracking approach, is that we cannot control the pen: it has to be the Wacom pen that is detected by the digitizing tablet. As a consequence, we will not be able to experiment on various nib-paper contacts. The contact between the plastic nib of the wacom pen and the sheet of paper is scratchy and provides a feeling similar to a pencil on paper.

The final system is illustrated in Figure 4.2. The apparatus is then composed of :

- A Wacom Intuos 3 digitizing tablet (PTZ1231W, 22')
- A sheet of paper (A4) put on the Wacom

[1] http://www.naturalpoint.com/optitrack/

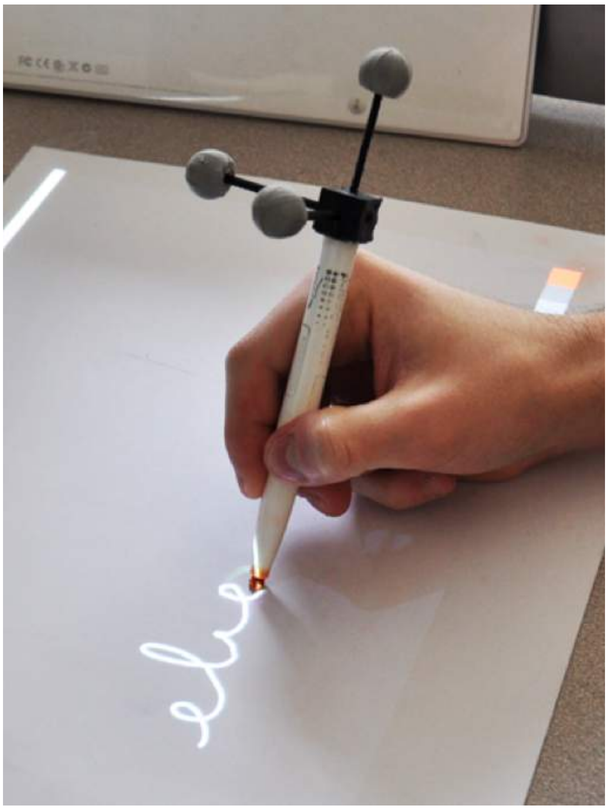


Figure 4.1: The pen with the three optical markers, using the optical tracking device (off-camera)

- A Wacom Intuos 3 Grip pen (17,5cm - 18g) with a plastic nib
- A projector (EPSON EB-G5450WU, 3 LCD 1920x1200)

The temporal sampling of the pen's position is 200Hz. The default Wacom drivers provides the pen positions through the operating system user events. We observed that this software pipeline introduced latency. We thus developed custom software to have a more direct access to the tablet's output. We disabled the default Wacom driver on the computer, and reversed engineered the communication protocol of the tablet so that we could work at the communication bus level (USB). In addition, we synchronized our program with the display refresh, so that we could update the display with the most recent pen position just before the display refresh. We evaluated the final end-to-end latency of our system using the low overhead approach of Bérard et al. [4]; we measured a latency of 50ms.

The spatial resolution of the Wacom sensors is rated at 5080lpi[2] (Wacom uses the unit of "line per inch" which is very similar to pixel per inch). The bottleneck of our system in spatial resolution would thus be the display resolution. In order to maximize it, we put the projector as close as possible to the sheet of paper so that we could still get a good focus of the image. With our setup, we created an image of 210x131mm. With a definition of 1920x1200 pixels, this translates to a resolution of 232ppi, or 0.11mm. As the tablet resolution is still much higher than the display resolution, we should be able to identify the exact display pixel that the pen

[2]http://www.wacom-asia.com/intuos3/spec/spec_tablet.html

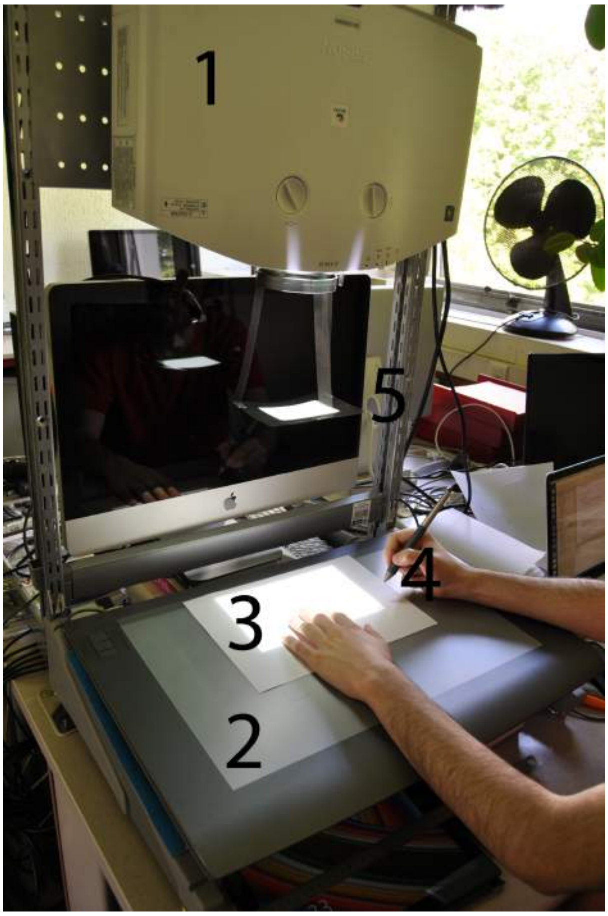


Figure 4.2: The final system: 1. projector, 2. digitizing tablet, 3. sheet of paper, 4. tablet's pen, 5. plastic cache to reduce brightness.

is pointing to. To achieve this result, we calibrated a bi-linear interpolation for the sensor-projection transformation by pointing with the pen at the four corners of the projected image. With this calibration, the ink is projected at a location covered by the surface of the pen's nib, which has a radius of about 0.5mm.

With this projector - paper proximity, the brightness of the generated image is dazzling. We thus added a plastic cache in front of the projector to decrease the brightness. The resulting appearance of the digital ink is very close to that of a pencil or a black pen. This is illustrated in Figure 4.3, although the photograph in this figure does not seem very faithful to the perception of the digital ink in the real world: with a naked eye, the digital ink appeared more similar to the physical inks.

4.2 Parameters degradation

Once our best effort at reproducing physical ink was implemented, we could consider ways to degrade each parameters independently.

To control the offset between the nib and the digital ink, we simply add a fixed offset to the exact nib position computed with the calibrated transformation (see Section 4.1). We need to chose the direction of the offset. We arbitrarily chose to display an offset directly above the

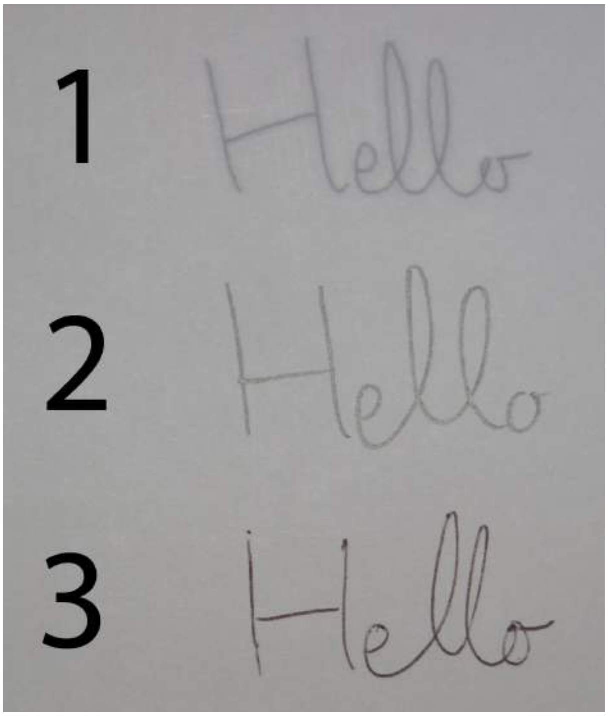


Figure 4.3: Comparison of digital ink (1), pencil (2) and physical black ink (3)

pen nib. This offset can be adjusted at the pixel level. With our display resolution of 232ppi, this translates to an offset resolution of 0.11mm.

To control the latency of our system, we simply buffer the events that we receive and release them to the application once they are old enough. More precisely, each time we read an event on the USB we push it in a queue and date it with the reception date. For the minimal latency condition, we simply pop out of the queue all the events at each display refresh. To simulate an artificial latency of $l$ms, we pop out the event only if it is at least $l - l_{min}$ms old at the date of the display refresh, where $l_{min} = 50ms$ is the minimal latency of our system presented in Section 4.1. This process is illustrated on Figure 4.4.

To control the spatial resolution, we can rely on the high 5080lpi spatial resolution of the digitizing tablet. However, any spatial resolution finer than the display resolution of 232ppi will not be perceivable by the user. We can simulate any spatial resolution below 232ppi by simulating a grid at the desired resolution, and associating the actual pen position reported by the tablet to the closest grid point.

To control the temporal resolution, we need to adapt the digitizing tablet native resolution of 200Hz. One way to decrease this frequency is to only process an event out of N, with N an integer. But this only enables the discrete set of frequencies at $200/N$Hz. This would put a strong constraint of the step size used to adjust the probe in the JND protocol (see Section 3.4). We chose instead to allow any value for the temporal resolution by estimating the position of the pen that the tablet would provide if it was sampling at any desired frequency in the range [0-200]Hz. The estimated position is computed by a linear interpolation of the two enclosing actual samples from the tablet.

The Figure 4.5 shows the result of the degradation of four different parameters of our pen.

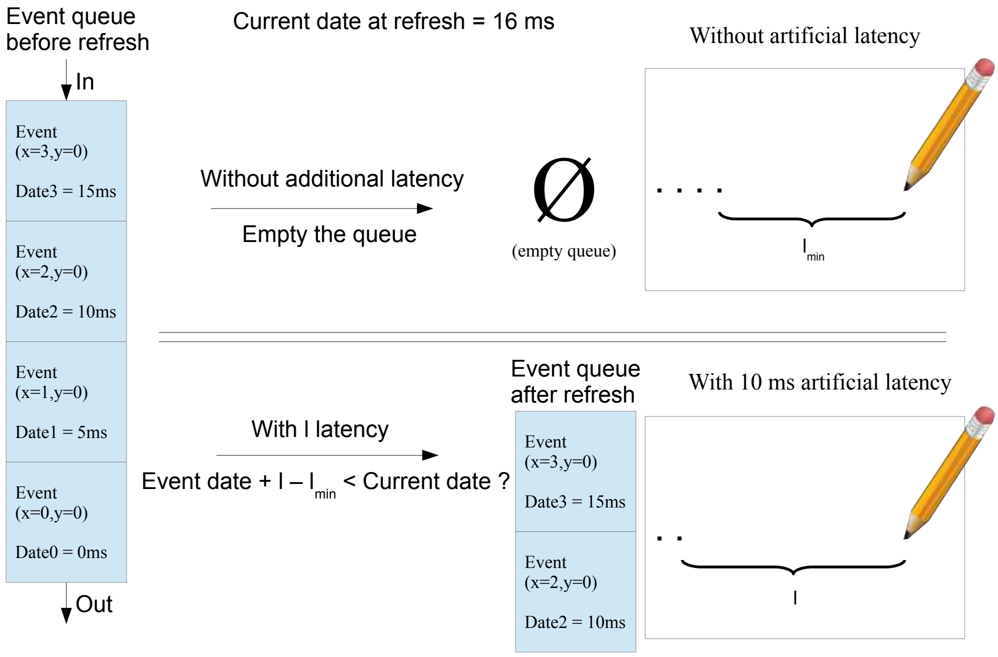


Figure 4.4: Controlling latency. Up: minimal latency ($l_{min} = 50ms$) is achieved by dequeueing all events at each display refresh. Down: a latency of $l = 60ms$ is achieved by popping events only if their receive date plus the desired latency minus the minimal latency of our system is lower or equal to the date of the display refresh.

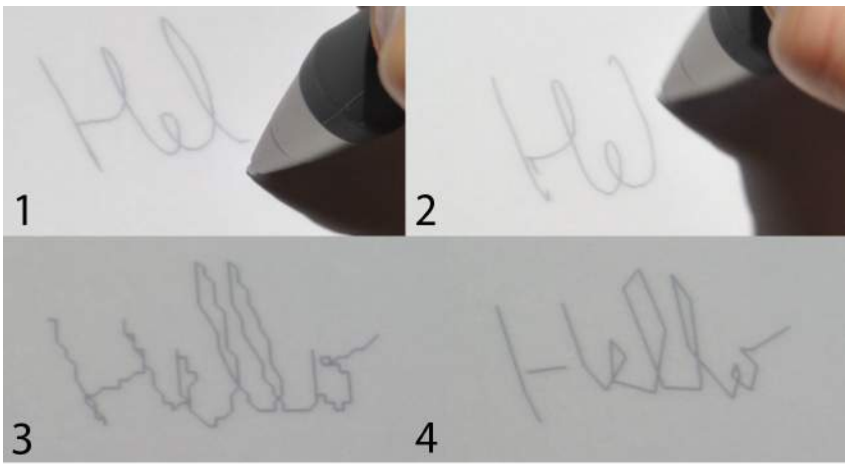


Figure 4.5: Degraded parameters : fixed offset between the nib and the projected ink (1), high latency (2), low spatial resolution (3) and low temporal resolution (4).

— 5 —
User experiments

As introduced in the presentation of our approach (in Section 3), we conducted two experiments. The first experiment was aimed at measuring the general subjective quality of our high performance digital pen. We also focused on the subjective quality of the nib-paper contact, as we could not control this parameter to implement a JND experiment. The second experiment was aimed at measuring the limits of Human perception for four technical parameters of our digital pen.

We ran both experiments as two parts of the same experimental session: participants performed the two experiments in sequence, the session lasting around 45 minutes. Twelve volunteers from our laboratory and from a nearby engineer school participated in the two experiments. Their age ranged from 21 to 36. They were all right handed except one, and they all had previous experience with commercial digital pens.

5.1 Subjective experiment

Participants first performed the subjective experiment. The goal of this experiment was to get the subjective rating of the participants on 3 different pens: a physical pen considered as an ideal pen, a commercial digital pen, and our high performance digital pen. We measured general preferences, but we also focused on the preference for the nib-surface contact.

Our hypothesis was that the ideal pen would receive the best ratings, the tablet would receive the worst ratings and our system would be between the two but closer to the ideal pen than to the tablet.

Apparatus

The three pens presented to participants are illustrated in Figure 5.2 and are listed below:

- A sheet of paper with a ballpoint pen,
- A Microsoft Surface Pro 2 tablet with a 208ppi display, used with its pen and running the OneNote application,
- Our high performance digital pen.

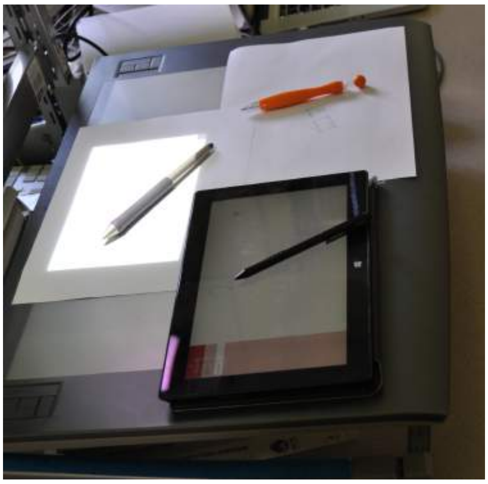


Figure 5.1: The three devices used in the subjective experiment: our high performance digital pen (left), a physical ballpoint pen and paper (top-right), and a Microsoft Surface Pro 2 (bottom-right).

Task and procedure

Participants were first asked to use the three devices for taking notes. They had 5 minutes to test all three devices in whatever order they wanted, and they could switch devices as often as they wanted. Then, they had to rate each device on a scale from 1 (worst) to 7 (best) on "comfort when writing and quality of the result".

After answering the first set of questions, participants could use our device for one more minute and then answer three questions:

1. Did you see a delay between the pen nib and the ink when the nib was moving?

2. Did you see an offset between the pen nib and the ink when the nib was stopped?

3. Did you see staircases in your curves?

Participants could also make comments at any time about their feelings when writing with any of the device. We hoped that the participants' comments would help us understand their ratings. The three final questions were useful to determine if our device could really approximate the ideal pen of these parameters.

Results

The Table 5.1 shows the subjective ratings provided by the participants for the three devices. These results are illustrated in Figure 5.2.

Some participants expressed some comments on the shortcomings of the tablet: they perceived some "delay", they had "difficulties for pointing with the stylus" that we attribute to the nib-ink offset. But some participants also expressed that they were "pleasantly surprised" by the comfort of writing on the tablet. Contrary to our hypothesis, a lot of participants preferred the smooth contact between the tablet stylus and the glass rather than a physical pen on paper.

Participant	Rating		
	Physical Pen	MS Surface	Our device
1	6	4	5
2	5	7	6
3	5	6	4
4	7	5	4
5	7	4	6
6	5	3	7
7	6	5	5
8	6	5	4
9	5	6	2
10	7	6	5
11	6	5	3
12	7	6	6
mean	6,00	5,17	4.75

Table 5.1: Ratings provided by the participants for the three tested devices, ranging between 1 (worst) and 7 (best).

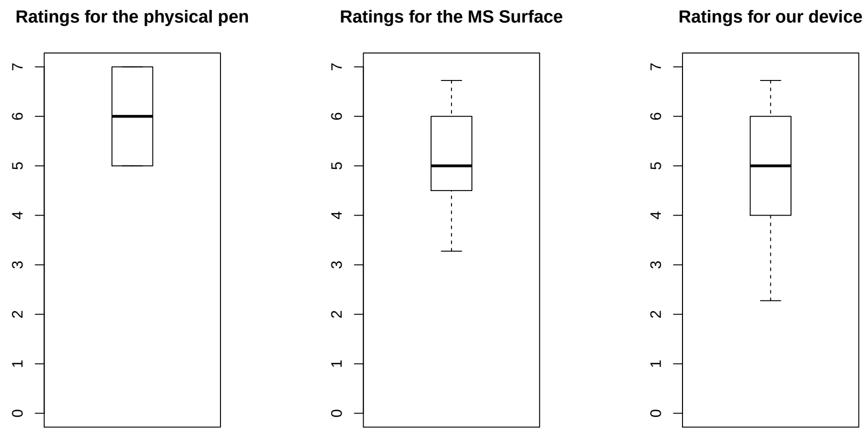


Table 5.2: The boxplot for each device rating : the physical pen, the MS Surface Pro 2 and our device. The boxes are delimiting the first and third quartile. The whiskers delimit the 2.5th and 97.5th percentile. The black line is the median.

They felt that the ruggedness of the ball pen and of the Wacom pen on paper was less comfortable than the smoothness of the tablet. Participants mainly disliked our system because of this ruggedness and because of the shadow that the hand created because of the projected image.

Table 5.3 summarizes the participants' answers to the three questions on our digital pen. 5 participants out of 12 said that they perceived latency in our system, but three specified that it was "very low", "acceptable" or "less than on the tablet".

Did you perceive any...	"yes" (%)	"no" (%)
delay?	42	58
offset?	0.0	100.0
staircases?	17	83

Table 5.3: Participants' answers to the three questions (12 participants) about our digital pen.

Discussion for the subjective experiment

The ratings reported in Table 5.1 confirm our hypothesis that participants preferred the physical pen to the digital pens. However, they rated our device worse than the tablet, despite the fact that the tablet had a pen-ink offset, worse latency, and lower spatial and temporal resolution. This indicates that the parameters that we identified may not be the most important for the subjective quality of a digital pen. The nib-surface contact may play an important role, and we assumed that people would prefer a rugged contact similar to physical pen rather than a smooth contact, but it did not seem to be the case.

The answers to the question about our digital pen indicate that, despite our best efforts, we were not able to implement a system with a latency that was low enough to be imperceptible. This subjective experiment suggests that latencies in the range of 50ms are perceivable. We will show that the objective experiment confirmed this result.

Only 2 participants out of 12 expressed that they saw staircases in the curves drawn with our digital pen. One participant commented that it was very low. We assume that these two people were able to discriminate pixels. Human eyes can distinguish pixels around 280ppi at a distance of a foot[1]. Our device has a resolution of 232ppi. If we consider that a Human eye is able to distinguish 120pixels/degree, then participants looking at the display at a distance of 40cm should not be able to distinguish the pixels. This was mostly the case, but some participants leaned towards the table when writing and came closer to the projected display. This could explain that two participants were able to distinguish pixels and see staircases in the curves. On the tablet, the display resolution was lower than on our device, but the anti-aliasing was used in the rendering of the strokes which made it more difficult to distinguish pixels. This suggests the use of anti-aliasing in our system's rendering. But this should be considered carefully as anti-aliasing is computationally expensive and may increase latency.

The performances of our device appeared to be high enough in terms of calibration (offset), and spatial and temporal resolution, as indicated by the answers to the second and third questions.

5.2 Objective experiment

After answering the questions of the subjective experiment, participants carried out the objective experiment. The goal was to measure the Human limit of perceptions for the four controlled parameters of our device: offset, latency and spatial and temporal resolutions.

[1] http://prometheus.med.utah.edu/ bwjones/2010/06/apple-retina-display/

Apparatus, task and procedure

Only our high performance digital pen was used in this experiment.

The task was designed according to the JND protocol as detailed in Section 3.4. Participants executed four JND series in sequence, one for each parameter (offset, latency, spatial or temporal resolution). We expected that participants may be subjected to fatigue after four series, and thus that they may be less efficient in the last series. We thus balanced the effect of fatigue accross parameters by presenting the series in random order for each participant. Participants were told which parameter was tested at the beginning of each JND series, they were thus aware that only this parameter would vary during the series. In each trial of a JND series, participants were asked to recognize the baseline when presented with both the baseline and a degraded version of the baseline. The first trial of a series was very easy, i.e. the probe was set at the maximal degradation, so that participants easily understood the effect of the tested parameter.

The display projected on the sheet of paper is illustrated in Figure 5.2. It had a white background and was divided into a left and a right side by a vertical grey line at its middle. Two small grey rectangles were displayed in each side but close to each other. Participants were asked to write in these rectangles of size 40mm x 21mm. This size was chosen to be big enough to allow the writing of the word “Hello” at a normal size, i.e. similar to the size of the word “Hello” written with a physical pen. Each side of the display also had a dark square. Participants tapped with the pen’s nib inside one of these squares to vote for the side as being the baseline. Finally, the display contained a dark square, in its lower right corner, used as an “erase” button. It was to allow participants to restart a trial only in case of an external problem. It has hardly ever been used.

At each trial, we *randomly* selected one side where the writing was performed with the pen at its baseline, and on the other side the pen’s performances were set at the probe (degraded) level. In other words, the pen’s parameters were dynamically changed as soon as the system detected that the pen location had switched sides.

Participants were asked to write the word “Hello” in the left rectangle then in the right rectangle and then to vote for the one side they thought was running with the pen at its baseline. The word “Hello” was chosen because it is very common and it contains many curves (spatial and temporal resolution have mainly an impact on curves). Participants were asked to write at the same speed and with the same style during the whole experiment. They were given the instruction to write “as if you were taking notes”. This requirement enabled to normalize each participant writing style. It also forbade them to track some paramaters degradation by changing their writing style (e.g. writing quicker to detect higher temporal resolution). As there was no “I don’t know” button, participants were forced to vote for the baseline at each trial, even when they could not perceive a difference between the two sides. As soon as they voted, the writing rectangles were cleared, the next trial was setup and the participants could start writing again. This allowed for a very fast execution of trials.

JND design

We chose the parameters of our JND series after Ng et al. [17]. Each series was actually composed of two interleaved JND series. The series was interleaved to avoid participants tracking their progress through the protocol. The first series began with the probe at the maximum degradation, as explained above. The second series began with the probe at the baseline. Both

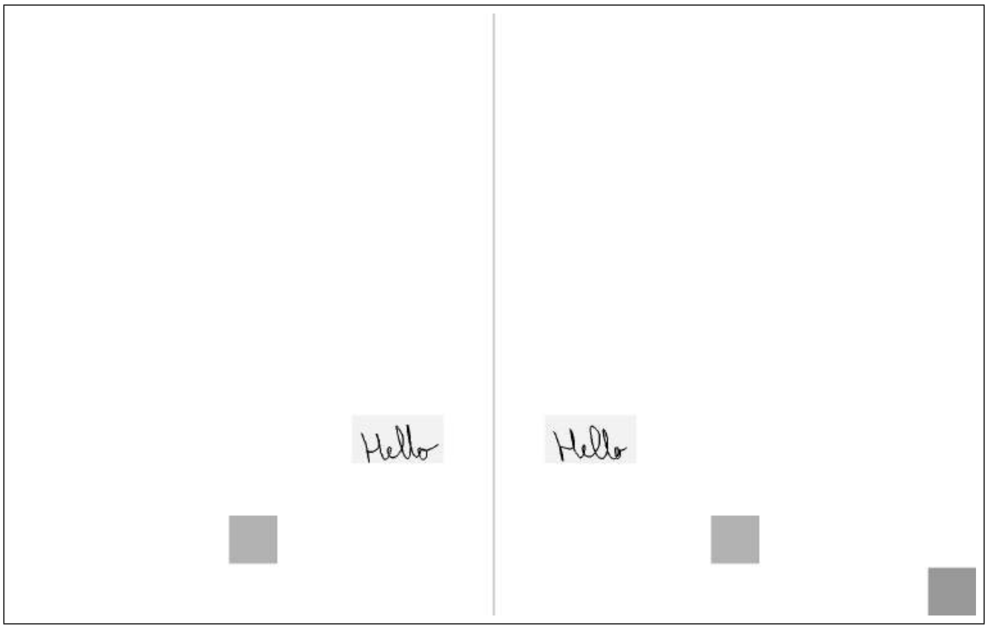


Figure 5.2: One screenshot of the display during the experiment.

series were stopped at their 10th reversal. The initial step size was chosen after pilot experiments to make the JND series converge as quickly as possible. At each reversal the step size was divided by two until it reached a minimum step size defined after pilot experiments. The minimal step size was 0.11mm (1 pixel) for the offset, 2.5ms for latency, 5.8ppi for spatial resolution and 3.75Hz for temporal resolution. At the end of a JND series, the JND was computed as the mean of the values at every reversal, except the first two which were discarded to take into account the initial phase. The JND for one participant for a given parameter was computed as the mean of the two interleaved series.

Results

Parameter	Baseline	Just Noticeable Difference			
		mean	std. err.	95% CI lower	95% CI upper
Offset (mm)	< 0.5	0.90	0.092	0.468	1.45
Latency (ms)	50	78	3.6	62.6	98.8
Spatial resolution (PPI)	232	163	9.8	153	193
Temporal resolution (Hz)	200	58	4.3	43.1	82.2

Table 5.4: Results of the objective experiment for the four parameters: baseline value and JND statistics (mean, standard error, 95% Confidence Interval lower and upper bounds) accross the 12 participants.

The results of the objective experiment are shown in Table 5.4 and illustrated in the Figure 5.5.

All the participants were able to track very small offsets. They could distinguish the offset even when the ink appeared at less than 1mm from the nib. The baseline in Figure 5.5 is dashed

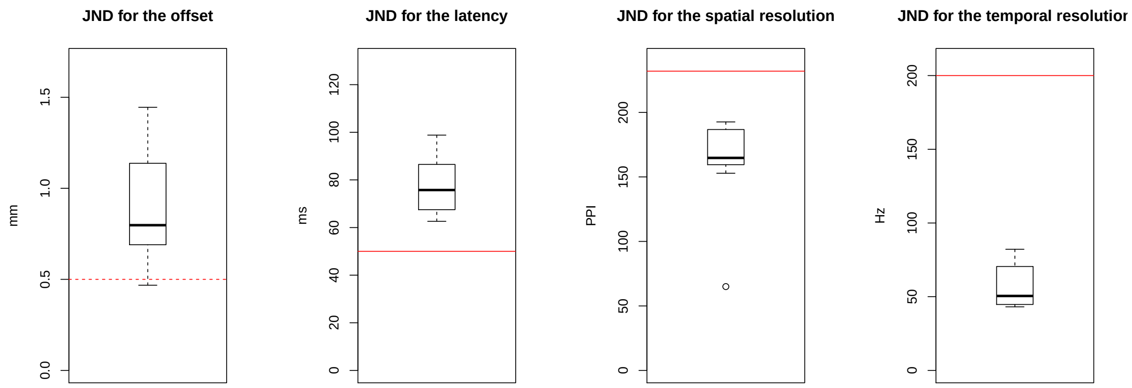


Table 5.5: The boxplot for each parameter : offset, latency, spatial and temporal resolution. The boxes are delimiting the first and third quartile. The whiskers delimit the 2.5th and 97.5th percentile. The black line is the median. The dots are data considered as outliers. The horizontal line is the baseline.

because we may be more precise than 0.5mm, but under this value, the pen nib thickness is hiding the ink so that participants are unable to see the offset.

For the latency, the participants were able to track latencies down to 78ms. Looking at the JND series showed that the probe often came at level of the baseline, which indicates that our baseline was not below the limit of Human perception, as discussed in the section JND for absolute perception thresholds.

The JND for the temporal resolution shows a large amount of variability, which we relate to the variability in writing speed accross participants: the effect of a low temporal resolution of the pen is more visible when the writing is fast.

Discussion for the objective experiment

Despite our best efforts to reduce the latency of our device, we could not get our baseline below 50ms. Imperceptible latency remains a very difficult challenge for interactive systems, as demonstrated by Ng et al. [17, 16]. The subjective experiment indicated that our baseline was above the limit of Human perception. The objective experiment confirmed this. Tracking the latency JND series of participants, we found that the probe was close to the baseline, often reaching the baseline. This is illustrated by one series on Figure 5.3. As we explained in Section 3.4, using the JND protocol to measure an absolute threshold of Human perception requires that the baseline is sensibly lower than this threshold. Even though the average JND for latency is at 78ms, this can be interpreted as the minimum perceivable *difference* of latency being around 28ms (78ms - 50ms). But with a lower baseline, participant may well have perceived lower levels of latency. However, this experiment shows that the threshold of Human perception for the latency when writing is under 78ms. It corroborates the results of Ng et al. [16] who showed that people were able to perceive latency down to 40 ms when scribbling. We consider the task of scribbling as similar to the task of writing in terms of cognitive load and attention.

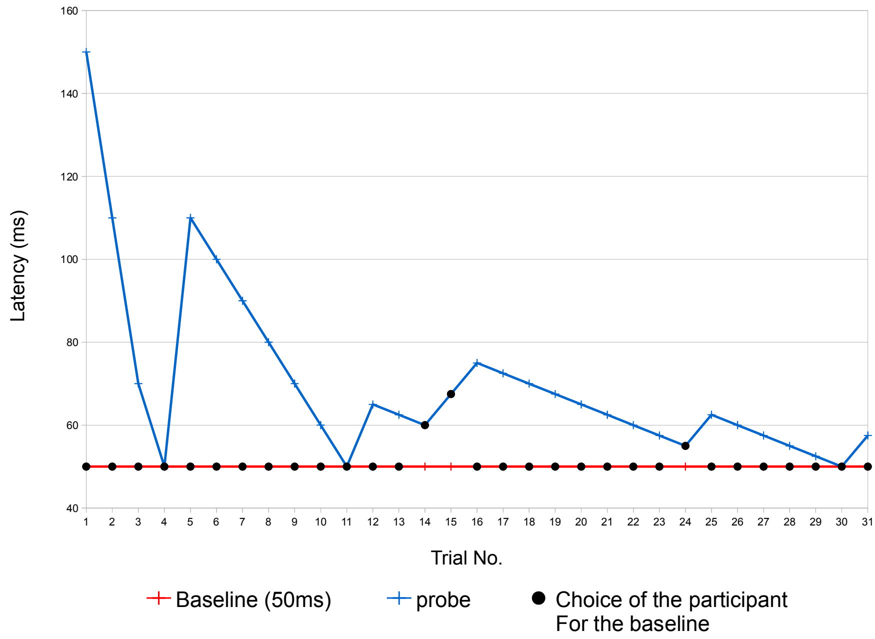


Figure 5.3: JND series of one participant for the latency. The probe is close to the baseline or even reaches it.

The experiment revealed that the results are dependent of the participants' writing style. In particular, the writing speed affects the perception of the temporal resolution, and the writing posture affects the perception of the spatial resolution. We did not give any instruction to the participants for the writing speed or the posture as we were aiming at a natural behavior: we asked them to write as they do when taking notes.

Still, the results of the experiment are interesting because they provide an image of the required efficiency for a digital pen on the population of the study. The upper bounds may be the most useful, as a device that satisfies the requirement of the upper bounds should be ideal for a major part of the whole user population.

The JND for offset is close to the baseline, but it was not possible to get a lower baseline for our digital pen: its offset was already unmeasurable because it was smaller than the width of the pen's nib. The JND series for offset reveals that digital pens should be carefully calibrated as offset as small as 0.47mm are perceivable by users.

The JND for spatial and temporal resolution are quite far away from their respective baselines, which indicates that our experiment did reveal a threshold of Human perception: digital pens that provide 193ppi of spatial resolution and 82.2Hz of temporal resolution should be indistinguishable from an ideal (physical) pen for these parameters.

— 6 —
Discussion

6.1 Parameters

In the work presented here, we selected five parameters of a digital pen as parameters that may have a strong influence on the usability of the pen for inking. The parameters are pen-nib contact, offset, latency and spatial and temporal resolution. There is no guarantee, though, that other parameters not studied here don't have a strong influence of the pen usability. Our choice is a result of our experience and the discussions with other researchers. We also consider and rejected some parameters that we discuss below.

The weight and the shape of the pen probably have an influence on its usability. However, it does not seem relevant to include these parameters in our study. These parameters influence physical pens as well as digital pens, and it seem quite feasible to replicate a good physical design on a digital pen.

In the physical world, a pen stroke is not a perfect line with constant thickness and constant color. The thickness and contrast of the stroke depends on the pen speed and the pressure on the inking surface. In our device, we used a simple invariant line to model the ink. On the contrary, the One Note application used in the subjective study models a nicer ink with varying thickness depending on the pen pressure. Our initial hypothesis was that the strokes appearance would not have an influence on users *performance* in inking. The subjective experiment revealed that the MS Surface pen was preferred to our system despite having a spatial offset, higher latency, and lower spatial and temporal resolutions. This may indicate that the influence of the strokes appearance may be stronger than what we anticipated, in particular for the user perception of the quality of the pen.

6.2 Task

Our study focuses on handwriting when taking notes. We can expect that the results would be different for other tasks involving a stylus. The offset, for example, has probably a much stronger influence than latency for the task of directly pointing at objects. But latency is well known to strongly degrade performances in dragging tasks.

When considering latency alone, Ng et al. have showed that the perceived latency depends on the task [16]. They showed that for a scribbling task, participants were able to track latencies down to 40ms. But for a dragging task, less demanding in cognitive load and attention, people tracked latencies down to 6ms.

Our hypothesis was that current digital pen technologies do not emulate physical pens well enough that they would allow efficient and comfortable inking. We wanted to evaluate the efficiency by ranking productions, but we explained in Section 3.3 why this approach was not applicable. We decided to measure the limits of Human perception with the idea that what you can't perceive can't be a hindrance. However, if a user *can* perceive a non ideal parameter, this does not directly imply that his performance or comfort will be impacted while performing the task. Hence the perception thresholds revealed by our objective do not strictly define the requirements for ideal digital pens: pen with lower performance may be sufficients for inking without degradation of users performance. This is quite improbable, though, as the literature has many examples of the opposite [12, 15, 25].

6.3 Comparison with commercial devices

Parameter	Mean JND	Typical Performance of commercial devices
Offset (mm)	0.90	2.5
Latency (ms)	78	130
Spatial resolution (PPI)	163	264
Temporal resolution (Hz)	58	60

Table 6.1: Mean JND and performance of recent commercial devices, the values were taken either from the Apple iPad Air or the Microsoft Surface Pro 2.

Table 6.1 opposes the mean JND resulting from the objective experiment and the performance of two recent commercial devices. The offset is problematic: participants were able to track offset smaller than a millimeter. On the Surface Pro 2, after careful calibration, some area of the screen exhibit offsets in the order of 2.5mm. Here again, participants expressed a concern about this offset during the subjective experiment. Moreover, the hindrance of offset is probably aggravated for tasks that require accuracy on the landing phase of the pen, such as pointing at the graphical user interface's buttons and menus.

Even though our baseline latency did not allow us to measure a threshold of perceivable latency, we can state that ideal pens should have a latency lower than 50ms. This is not the case for the commercial tablets. Indeed, some users expressed a "delay" problem when using the Surface Pro in the subjective experiment. Our studies confirm that latency is a major problem of current devices.

The spatial resolution for the devices is actually the display resolution, as the sensors are more accurate than the displays. The iPad air can display at 264ppi and the Microsoft Surface Pro 2 at 208ppi. My study shows that for a writing task, these resolutions are high enough to be imperceptible.

In terms of temporal resolution, when taking notes, our study showed that people are not able to distinguish frequencies beyond 58Hz. Thus, a tablet sampling at 60Hz should be sufficient when taking notes. However, we anticipate that for drawing large curves at high speed, the perception threshold will be higher and a 60Hz sampling will not be enough. We observed the problem on the Surface Pro: taking notes did not present problems in term of spatial resolution, but drawing high speed loops showed clear segments in the curves.

— 7 —

Conclusion and future work

In this work, we studied the performance of digital pens in order to bring to light their limitations compared to physical pens.

We created a new high performance digital pen which performances make it more similar to a physical pen than current commercial devices. We attempted to compare productions done with a physical pen and with our device under different parameters. These efforts revealed that the subjectivity of drawings evaluation does not provide a metric that is accurate enough to measure an effect of device performance.

We thus followed a different approach and ran two experiments to study user's perception of digital pens. The first study showed that people are pretty comfortable with the smooth glass surface of a tablet and they don't like the ruggedness of our device's plastic nib on paper. We also assume that the appearance of the digital ink of our system is not satisfying as they preferred the varying thickness of the strokes on the tablet.

In the second study, we ran a set of Just Noticeable Difference experiments in order to detect the limit of human perception for the four parameters we chose to control (offset, latency, spatial and temporal resolutions). It revealed that the pen's nib to ink offset is troublesome. Users are able to track offsets below 1mm for a writing task, and other tasks such as pointing may be even more demanding. The latency is also a troublesome parameter: users were able to perceive the minimal 50ms latency of our system, which is sensibly lower than current commercial systems. However, current commercial digital pen already outperform Human capabilities in terms of spatial resolution. Their temporal resolution appears to be sufficient for the writing task, but this may not be the case for drawing.

Our work shows that the performance level of current commercial tablets does not reach the Human perception limits. Improvements will be required for these digital pen to offer a comfort in writing that is similar to the one of physical pens. These shortcomings may explain why people usually prefer a physical pen and paper rather than a tablet when taking notes.

Our studies focused only on the differences in writing with a physical pen or a digital pen. We did not take into account other differences between paper and tablets. In future work, we plan to investigate the other affordances of paper such as flexibility or the size of the display. Indeed, it is possible to spread more than one sheet of paper on a desk and to arrange them the way we want, but a tablet only offers a limited screen size and usually it can only display one page at a time. Tablets are also fragile and not flexible.

We also focused on the task of writing and drawing during note taking. But tasks that are less cognitively demanding allow users to perceive more subtle shortcomings of the digital pen performances. We thus plan to conduct similar studies on simpler tasks, with the idea that it may exist absolute bounds on Human perception for each of the parameters we studied. If it is the case, then the goal of creating an ideal digital pen may actually be achievable.

8
Acknowledgements

I thank François Bérard my supervisor who gave me a lot of ideas and ways to improve my work. I thank also Yann Laurillau who helped me communicate with the Wacom through the USB port and Thomas Vincent who lent me the Microsoft Surface Pro 2 for my study.

This work has been partially supported by the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01) funded by the French program Investissement d'avenir

Bibliography

[1] Helmut Alt, Bernd Behrends, and Johannes Blömer. Approximate matching of polygonal shapes. *Annals of Mathematics and Artificial Intelligence*, 13(3-4):251–265, 1995.

[2] Esther M Arkin, L Paul Chew, David P Huttenlocher, Klara Kedem, and Joseph SB Mitchell. An efficiently computable metric for comparing polygonal shapes. In *Proceedings of the first annual ACM-SIAM symposium on Discrete algorithms*, pages 129–137. Society for Industrial and Applied Mathematics, 1990.

[3] Dana H Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122, 1981.

[4] François Bérard and Renaud Blanch. Two touch system latency estimators: High accuracy and low overhead. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS)*, pages 241–250. ACM, 2013.

[5] Francois Berard. The magic table: Computer-vision based augmentation of a whiteboard for creative meetings. In *IEEE International Workshop on Projector-Camera Systems (PROCAMS)*, 2003.

[6] Barrett Comiskey, JD Albert, Hidekazu Yoshizawa, and Joseph Jacobson. An electrophoretic ink for all-printed reflective electronic displays. *Nature*, 394(6690):253–255, 1998.

[7] Katherine M Everitt, Meredith Ringel Morris, AJ Bernheim Brush, and Andrew D Wilson. Docudesk: An interactive surface for creating and rehydrating many-to-many linkages among paper and digital documents. In *Horizontal Interactive Human Computer Systems, 2008. TABLETOP 2008. 3rd IEEE International Workshop on*, pages 25–28. IEEE, 2008.

[8] Pasi Fränti, Alexey Mednonogov, Ville Kyrki, and Heikki Kälviäinen. Content-based matching of line-drawing images using the hough transform. *International Journal on Document Analysis and Recognition*, 3(2):117–124, 2000.

[9] François Guimbretière. Paper augmented digital documents. In *ACM Symposium on User Interface Software and Technology (UIST)*, UIST '03, pages 51–60, New York, NY, USA, 2003. ACM.

[10] David Holman, Roel Vertegaal, Mark Altosaar, Nikolaus Troje, and Derek Johns. Paper windows: interaction techniques for digital paper. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 591–599. ACM, 2005.

[11] Anoto : http://www.anoto.com.

[12] Ricardo Jota, Albert Ng Ng, Paul Dietz, and Daniel Wigdor. How fast is fast enough? a study of the effects of latency in direct-touch pointing tasks. In *ACM Conference on Human Factors in Computing Systems (CHI)*. ACM, 2013.

[13] Christian Kaernbach. Simple adaptive testing with the weighted up-down method. *Attention, Perception, & Psychophysics*, 49(3):227–229, 1991.

[14] Chunyuan Liao, François Guimbretière, and Ken Hinckley. Papiercraft: a command system for interactive paper. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 241–244. ACM, 2005.

[15] I Scott MacKenzie and Colin Ware. Lag as a determinant of human performance in interactive systems. In *ACM Conference on Human Factors in Computing Systems (CHI)*, pages 488–493, 1993.

[16] Albert Ng, Michelle Annett, Paul Dietz, Anoop Gupta, and Walter F. Bischof. In the blink of an eye: Investigating latency perception during stylus interaction. In *ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1103–1112. ACM, 2014.

[17] Albert Ng, Julian Lepinski, Daniel Wigdor, Steven Sanders, and Paul Dietz. Designing for low-latency direct-touch input. In *ACM Symposium on User Interface Software and Technology (UIST)*, pages 453–464. ACM, 2012.

[18] Sara Rosenblum, Patrice L Weiss, and Shula Parush. Product and process evaluation of handwriting difficulties. *Educational Psychology Review*, 15(1):41–81, 2003.

[19] Bill N Schilit, Gene Golovchinsky, and Morgan N Price. Beyond paper: supporting active reading with free form digital ink annotations. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 249–256. ACM Press/Addison-Wesley Publishing Co., 1998.

[20] Hyunyoung Song, Tovi Grossman, George Fitzmaurice, François Guimbretiere, Azam Khan, Ramtin Attar, and Gordon Kurtenbach. Penlight: combining a mobile projector and a digital pen for dynamic visual overlay. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 143–152. ACM, 2009.

[21] Hyunyoung Song, Francois Guimbretiere, Tovi Grossman, and George Fitzmaurice. Mouselight: bimanual interactions on digital paper using a pen and a spatially-aware mobile projector. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2451–2460. ACM, 2010.

[22] Jürgen Steimle. Designing pen-and-paper user interfaces for interaction with documents. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pages 197–204. ACM, 2009.

[23] Jürgen Steimle. Survey of pen-and-paper computing. In *Pen-and-Paper User Interfaces*, Human-Computer Interaction Series, pages 19–65. Springer Berlin Heidelberg, 2012.

[24] Salvatore Tabbone, Laurent Wendling, and Karl Tombre. Matching of graphical symbols in line-drawing images using angular signature information. *Document Analysis and Recognition*, 6(2):115–125, 2003.

[25] Robert J. Teather, Andriy Pavlovych, Wolfgang Stuerzlinger, and I. Scott MacKenzie. Effects of tracking technology, latency, and spatial jitter on object movement. In *IEEE Symposium on 3D User Interfaces (3DUI)*, pages 43–50. IEEE Computer Society, 2009.

[26] Pierre Wellner. Interacting with paper on the digitaldesk. *Communications of the ACM (CACM)*, 1993.