

Evaluating and Understanding the Usability of a Pen-based Command System for Interactive Paper

CHUNYUAN LIAO

FX Palo Alto Laboratory, Palo Alto, CA, U.S.A

Department of Computer Science, University of Maryland, College Park, U.S.A.

FRANÇOIS GUIMBRETIERE

Department of Information Science, Cornell University, U.S.A.

Department of Computer Science, University of Maryland, College Park, U.S.A.

To combine the affordance of paper and computers, prior research has proposed numerous interactive paper systems that link specific paper document content to digital operations such as multimedia playback and proofreading. Yet, it remains unclear *to what degree* these systems bridge the inherent gap between paper and computers when compared to existing paper-only and computer-only interfaces. In particular, given the special properties of paper, such as limited dynamic feedback, how well does an average novice user learn to master an interactive paper system? What factors affect the user performance? And how does the paper interface work in a typical use scenario?

To answer these questions, we conducted two empirical experiments on a generic pen-gesture-based command system, called PapierCraft [Liao, et al., 2008], for paper-based interfaces. With PapierCraft, people can select sections of printed documents and issue commands such as copy and paste, linking and in-text search. The first experiment focused on the user performance of drawing pen gestures on paper. It proves that users can learn the command system in about 30 minutes and achieve a performance comparable to a Table PC-based interface supporting the same gestures. The second experiment examined the application of the command system in active reading tasks. The results show promise for seamless integration of paper and computers in active reading for their combined affordance. In addition, our study reveals some key design issues, such as the pen form factor and feedback of gestures. This paper contributes to better understanding on pros and cons of paper and computers, and sheds light on the design of future interfaces for document interaction.

Categories and Subject Descriptors: H5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Graphical user interfaces*

General Terms: Management, Documentation, Design, Human Factors, Evaluation

Additional Key Words and Phrases: Paper-based Interfaces, Pen Gestures, Command Systems

1. INTRODUCTION

Since the introduction of the first Anoto digital pen [Anoto, 2002] in 2002, there has been a renewed interest in paper-based interfaces. Before the availability of the Anoto system, paper-based interfaces such as DigitalDesk [Wellner, 1993], Xax [Johnson, et al., 1993]

Acknowledgement and author addresses

Permission to make digital/hard copy of part of this work for personal or classroom use is granted without fee, provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication, and its date of appearance, and notice is given that copying is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or a fee.

and PaperPDA [Heiner, et al., 1999] required either a complex setting to capture strokes made on paper in real time, or relied on users to scan paper documents for post-processing. The latter approach is somewhat cumbersome and only offers a limited picture of how the strokes were created on paper. In contrast, the Anoto system adopts a highly portable fountain-pen-like digital pen as its interface, and captures not only the shape of the strokes but also the pressure and timing information in real time. The system also provides an ID of the page on which the strokes have been made, making it easy to merge captured data onto the digital version of a printout [Guimbretiere, 2003]. These new features led to the design of several interactive paper systems. For instance, PapierCraft [Liao, et al., 2008] and PaperPoint [Signer and Norrie, 2007] import the captured pen marks from printouts into the corresponding digital documents for active reading and slide annotation, respectively. Extending such 2D pen interaction into 3D space, ModelCraft [Song, et al., 2006] captures annotations on 3D models. The digital pen input can also be integrated with other devices like a mobile projector such as PenLight [Song, et al., 2009] and MouseLight [Song, et al., 2010] to create a highly portable descendent of the original Digital Desk system. The main thrust for these systems was to combine the advantage of paper in terms of display quality, portability, robustness and ease of annotations [Sellen and Harper, 2001] with the advantage of digital media in terms of, for example, easy distribution and processing. Yet to our knowledge, there has been no attempt to characterize empirically how successful the digital pen could be at bridging the gap between the paper and the digital world.

As a step toward better answers to this question, we are presenting the results of two experiments that evaluated the performance of PapierCraft, a generic command system with its application in active reading, which is a task many people still prefer to perform on paper [Sellen and Harper, 2001]. PapierCraft is a good candidate, because it mirrors the command structure of the pen-based command system designed for tablet computers, such as Scriboli [Hinckley, et al., 2005], making the comparison to digital interface easier. Furthermore, Liao et al. has established that when augmented with simple pen-top feedback, PapierCraft's command selection capability is comparable to that of a similar implementation on a Tablet PC [Liao, et al., 2006].

In the first experiment, we extended previous work [Liao, et al., 2006] by considering the overall performance of the PapierCraft system including not only the command gesture itself but also the cost of transition between inking and issuing a command, and the cost of selecting the parameters of the command. Our results indicate

that, within 30 minutes of use, an average participant can master the gesture command system on paper and can achieve a performance comparable to that on the Tablet PC interface in comparison.

In the second experiment, we further broadened the scope of our investigation by asking participants to compare PapierCraft, a dual Tablet PC interface and normal paper setting, when performing the typical active reading tasks [Adler and Van Doren, 1972]. Our results from the post-experiment self-reports are very encouraging. Participants believed the PapierCraft interface combines the advantages of paper and computers, and they especially appreciated the ability to apply digital functions directly to paper without sacrificing the inherent flexibility of paper. When combined with Liao's results [Liao, et al., 2006], our work presents a complete picture of the performance of the PapierCraft interface, and establishes that in the domain of active reading, PapierCraft has the potential to effectively bridge the gap between the paper and digital world.

2. RELATED WORK

2.1 Paper-based interfaces

Paper-based interfaces aim to bring powerful digital functions to physical paper. DigitalDesk [Wellner, 1993] tracks paper documents and projects digital content on top of them. It offers rich digital affordance, but paper flexibility is limited by the non-portable hardware (e.g., the overhead projector and cameras) and complicated system configuration. Similar is Ariel [Mackay, et al., 1995] for augmented engineering drawings. A-Book [Mackay, et al., 2002] relieves such constraint by using a PDA that plays the role of a magic lens on top of the paper document. Intelligent Paper [Dymetman and Copperman, 1998] and Paper++ [Norrie and Signer, 2003] use a separate screen, without the alignment requirement, for rendering multimedia associated with pre-defined hotspots on paper. XAX [Johnson, et al., 1993] and PaperPDA [Heiner, et al., 1999] take a pen-paper-only interface for form-filling through scanned paper. Although more portable, they provide a constrained set of digital functions that might not fit well into the freeform interaction encouraged by active reading.

Recent advancements in digital pens open the doors to interfaces that enable rich digital functions without loss of paper affordance. An Anoto pen uses a built-in camera that observes special dot patterns in paper background to capture handwriting. The pen can capture strokes in real-time and label each stroke with a time stamp and an ID of the page on which the stroke was written. Taking advantage of these features, PADD [Guimbretiere, 2003] allows people to use an Anoto pen to annotate printouts, and automatically merges the annotations into the corresponding original digital files.

Leapfrog's Fly pen [LeapFrog, 2005] and LiveScribe's Pulse pen [LiveScribe, 2007] use the Anoto technology to support paper-based interactions with pre-printed or drawn interfaces. For more flexibility in issuing commands, PapierCraft [Liao, et al., 2008] introduces to paper a generic pen-gesture-based command system with which users can draw pen gestures to select arbitrary paper document content and choose a digital command to be applied. There are also other paper-based gesture commands customized for specific fields. PaperProof [Weibel, et al., 2008] supports a set of gestures for proof-editing Word documents on printouts. Upon synchronization, gestures are interpreted as Word editing commands on the corresponding digital document. CoScribe [Steimle, et al., 2009] permits a "stitching" gesture to attach personal paper book marks to a printout of shared documents and supports collaborative reading. Musink [Tsandilas, et al., 2009] provides customizable gestures for composers to create music scores. We chose the PapierCraft system for our evaluation, because it represents the most generic approach to paper-based interaction and can be easily compared to similar digital systems such as Marking Menu as shown in our previous work [Liao, et al., 2008].

Most recently, other devices, such as mobile projectors, have been combined with the Anoto digital pen. PenLight [Song, et al., 2009] simulates a pen-top projector to augment large architecture blueprints, and MouseLight [Song, et al., 2010] further explores how a hand-held spatially-aware mobile projector could be combined with a digital pen for bi-manual interaction on architecture blueprints. These interfaces aim at large document surfaces, which is usually not the case for active reading, whereby people often use much smaller, letter-sized or A4 printouts. Therefore, we did not use these interfaces for our experiments.

Besides the Anoto digital pen, mobile phones can be used to enable digital interaction on paper. HotPaper [Erol, et al., 2008] facilitates adding multimedia annotations to text-based paper documents and retrieving of the annotations. PACER [Liao, et al., 2010] supports hybrid camera-touch gestures for fine-grained interaction on paper, but it does not capture pen annotations created on paper, so is not suitable to the active reading tasks that we wanted to evaluate. Map Torchlight [Schöning, et al., 2009] utilizes a projector on a cell phone to augment a paper map but does not support pen interactions as does PenLight [Song, et al., 2009].

2.2 Evaluation of Paper-based Interfaces

Our work was originally motivated by Adler's report [Adler, et al., 1998] on the characteristics and requirements of work-related reading. The report highlights how

readers often rely on multiple display surfaces, interleave reading and writing, and rely on the affordance of paper to support search, browsing and collaboration. Later, O’Hara compared reading on paper and on a monitor, pointing out that paper is more efficient than screens for annotating, navigation and spatial arrangement [O’Hara and Sellen, 1997]. Recently, Morris extended O’Hara’s study by including newer devices and software, such as multiple Tablet PCs and large horizontal displays [Morris, et al., 2007]. One key conclusion is that the active reading interfaces should include both horizontal and vertical displays and be flexible enough to configure and support multiple input devices, e.g., a pen, a mouse and a keyboard. The work presented here extends this work by including a newer interactive paper system that combines paper affordance with digital-pen-enabled digital affordance. We did not find any other major comparison evaluations in the literature, except some usability tests on specific paper-based interfaces [Liao, et al., 2008, Song, et al., 2010, Steimle, et al., 2009, Tsandilas, et al., 2009, Weibel, et al., 2008].

3. THE PAPIERCRAFT COMMAND SYSTEM

3.1 Pen-based gestures

PapierCraft has been designed as the primary command system for the Paper Augmented Digital Document system (PADD [Guimbretiere, 2003]), an infrastructure designed to support cohabitation between the paper and the digital world. While the original version of PADD did not consider commands per se, it became clear that a command system was necessary to address the variety of commands required by paper interactions. This requirement motivated the first version of PapierCraft, which explores the feasibility of a paper-based command system where the only feedback is the ink that lay on the paper, written by a digital pen.

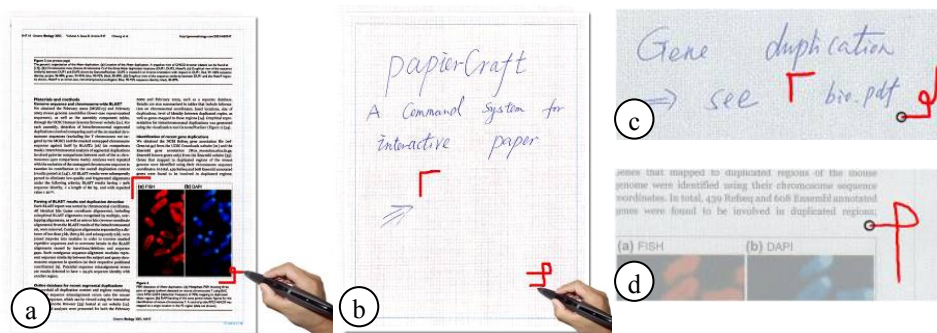


Figure 1. (a) An example of the Copy command in PapierCraft. (b) A Paste command. (c) A Hyperlink command. (d) A Hyperlink Target command. The components of this figure are excerpted from [Liao, et al., 2008].

PapierCraft drew from work done on pen-based computer interfaces like Scriboli [Hinckley, et al., 2005] to create complete command syntax with a digital pen on paper. A PapierCraft gesture command mainly consists of two components, namely the *command scope gesture* and *command type gesture*. A command scope gesture is the stroke(s) used to select the document content to be manipulated at a fine granularity, which may range from individual words to arbitrary document regions. A command type gesture is the stroke(s) utilized to specify the digital operation (i.e. command type) to be applied to the document content. To use the interface, a user first prints a digital document onto Anoto-augmented paper. The system automatically registers each hardcopy page with its original digital version. The user can then mark with a digital pen on the printouts to digitally manipulate the document content. For example (Figure 1- a), a user can draw a pair of cropping marks as a command scope gesture to select a figure in the printout, a pigtail as a separator and a marking menu as a command type gesture, of which direction indicates one of the eight commands to be used (in this example, “right” is for command “Copy”). Subsequently, the user can paste the figure into a separate paper note with a “paste” command (Figure 1-b). Using the stroke timestamps and page IDs, PapierCraft can carry out all operations in the context of the original digital documents, and their outcome can be seen on the digital document viewer provided by the infrastructure or be printed out for the next round of interaction.

PapierCraft supports various other gesture commands. For instance, a user can draw a “hyperlink” command to create a hotspot on top of a document segment (e.g., “bio.pdf” in Figure 1-c), and then link it to another page specified with a “hyperlink target” command (Figure 1-d). Consequently, pen tapping within the paper hotspot area will bring up the digital version of the linked page on a nearby connected computer, if any. PapierCraft also supports the interaction between paper and the web. A user can underline a word on a printout and then draw a “Google” command. The selected word is then forwarded to an associated computer for web search. More details can be found in [Liao, et al., 2008].

3.2 Pen-top feedback

When performing digital functions on the paper interface, users often require feedback to discover the supported commands, identify mistakes early or support application-specific tasks such as finding a keyword. Although the feedback can be conveyed by a nearby display such as a phone or a laptop, this will require users to switch back and forth between their pen and the display. To address this issue, we proposed pen-top multimodal

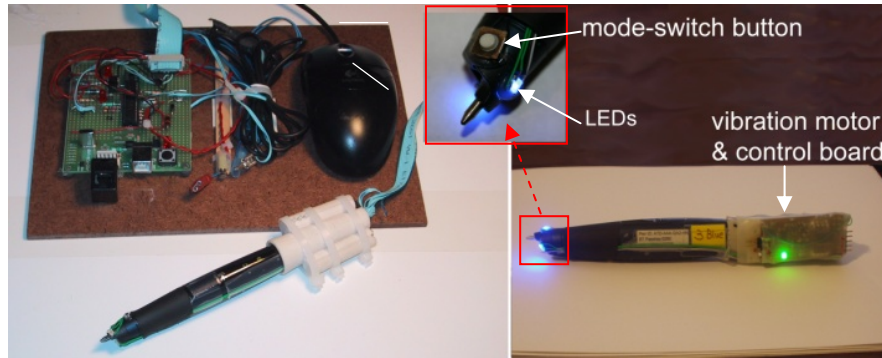


Figure 2. The feedback pens based on the Logitech digital pen. (Left) the first version (Right) the second version.

feedback based on a combination of visual, tactile and auditory feedback to support PapierCraft interactions [Liao, et al., 2006]. As shown in Figure 2-left, the first prototype was built around a digital pen, which is augmented by LEDs, vibration motors and speakers (simulated by a nearby computer). Those augmenting parts are connected to a control board via a long ribbon cable.

The feedback pen provides real time feedback for mode switch, marking menu navigation (e.g. boundary crossing, item names and item ambiguity warning), confirmation of menu selection, as well as specific tasks such as keyword search within a paper document (see Figure 3). More details can be found in [Liao, et al., 2006]. The system proved to be effective in increasing user performance. In particular, the system allows users to reach on paper a level of performance similar to what could be observed

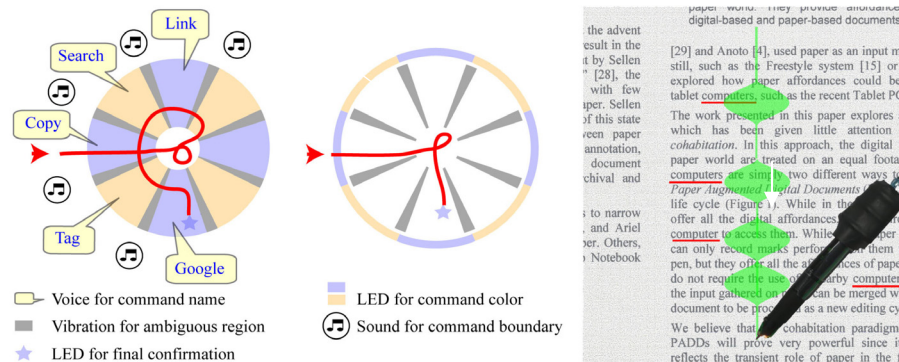


Figure 3. The illustration of feedback scheme supported by the feedback pen (reproduced from [Liao, et al., 2006] with permission). (Left) A novice user starts a gesture (in red) from left, make a pigtail, draw a half circle around the cross to navigate the marking menu via feedback by voice, sound, vibration, and LEDs, and finally select command “Google.” (Middle) An expert selects the same command, receiving feedback for final confirmation. (Right) Feedback for keyword search (“computer(s)” in this example). While the user draws a line in the page margin from top to bottom, the pen’s LED lights up when the pen tip gets close to an occurrence. More details can be found in [Liao, et al., 2006].

on a similar interface running on a Tablet PC [Liao, et al., 2006].

Unfortunately, the first prototype requires the pen to be tethered (Figure 2-left), which limits the overall usability of the pen in more realistic scenarios. To address this problem, we designed a new wireless version of the system (Figure 2-right). The new system uses a Bluetooth module to establish a link between the controlling computer and a micro-controller on the pen, which manages the LEDs and the vibrator and senses the mode switch button. All these feedback parts are integrated into a small clip-on board and are powered by a small battery. The controlling computer receives strokes data directly from the Bluetooth pen and button press from the pen-top board. Upon processing this data, the computer sends commands to the pen-top board for visual and tactile feedback and simulates audio feedback if needed.

4. EVALUATION OF PEN-GESTURE PERFORMANCE

In our previous work [Liao, et al., 2006], we conducted an experiment on how the proposed feedback mechanism improves the discovery of available commands and early error corrections. To do so, we only considered the case of a simple underline as the command scope gesture. Of course, real world performance is influenced by other factors, including the fact that people often alternate between inking and gesturing, and that users use a variety of command scope selection tools while reading a document. To address this problem, we decided to empirically compare the overall command issuing performance of PapierCraft to a similar implementation on a Tablet PC. Unlike the previous feedback pen experiment [Liao, et al., 2006], we did not consider the non-feedback case in this one due to two reasons. First, the previous study [Liao, et al., 2006] has established the value of the pen-top feedback in helping novice users discover and learn the PapierCraft interface with only a pen. Second, this new test involves more complex tasks in which we believe a non-feedback pen is inherently error-prone, for example inking-gesturing mode switch. As a result we believe that a non-feedback pen user, without referring to any other tools, would have much higher cognitive load to remember and easily use the interface.

In those two tested conditions, participants were required to draw designated pen gestures to select specific document content and switch pen modes. In the Tablet PC condition, the selected documents were rendered on a Tablet PC, on which participants drew pen gestures using a stylus. The resulting interaction feedback was presented on the Tablet PC screen. In the PapierCraft condition, the same gestures were performed on printed documents via a PapierCraft feedback pen, which conveys multimodal feedback

for the interaction. To be as realistic as possible, we used an actual document as the basis for our task, alternated inking and gesturing, and varied both the command scopes and types across the experiment.

4.1 Experiment Tasks

In a typical active reading scenario, users often interleave reading, annotating and manipulating document content. More and more paper interactions are interleaved with computer-based interactions, for example, to perform a search or create new content [Morris, et al., 2007]. While the exact interleaving of these activities vary between tasks and between individuals, as far as pen interaction is concerned, this cycle can be abstracted as a five-step cycle including 1) annotation, 2) switching the pen to a “command” mode, 3) drawing a command, 4) switching the pen back to an “annotation” mode, and 5) resuming annotation. To simulate this cycle in our experiment, we asked participants to first write a letter in the “annotation” mode, and then to draw a gesture in the “command” mode, followed by another letter in the “annotation” mode. We believe these simplified steps comprehensively characterize the key interactions involved in active reading.

Accordingly, we used the stimuli as exemplified in Figure 4 (left) for the trials. The stimulus indicates the starting letter “a,” the shape of the command scope gesture (the pair of cropping marks), the name of a faked command type “Papaya,” and the ending letter “b.” Note the command types in the test were deliberately chosen not to be associated with any normal user interface commands to avoid possible bias against specific command types. The stimulus was rendered on a screen in front of participants.

The stimulus was triggered when the participant pen-tapped the “start” button in the margin of the testing document as shown in Figure 4-right, which was either printed on paper or rendered on a Tablet PC. Following the stimulus, the participant first wrote “a”

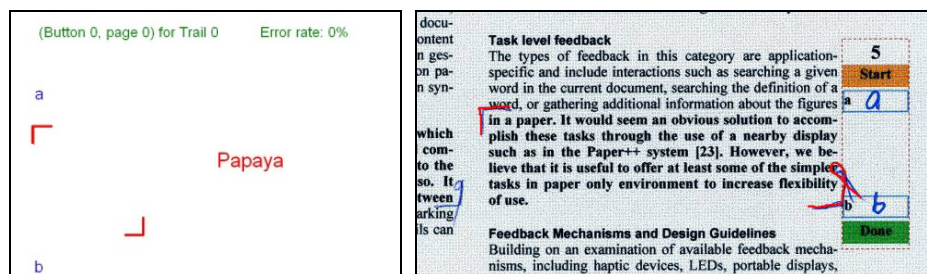


Figure 4. A stimulus (left) and a paper version of the input interface (right). The participant (1) taps the “start” button, (2) reads the stimulus, (3) writes “a” in the annotation mode, (4) draws a cropping mark scope and a marking menu for “Papaya” in the gesture mode (highlighted in red for clarity), (5) writes “b” in the annotation mode, and (6) taps the “done” button to submit the input.

in the designated blank box, drew the cropping marks as the command scope gesture and the marking menu as the command type gesture, and then wrote “b” in the second blank box. The participant can modify the input until satisfied, and then pen-tap the “done” button to submit the input. If the input matched the stimulus, a brief “happy” sound was played, otherwise, a “sad” sound followed, along with a detailed error description. In this way, participants could learn from errors and improve their technique throughout the experiment.

4.2 Experiment Settings

The experiment was conducted in a university laboratory. All stimuli were rendered on a vertical monitor in front of participants, who drew the required command scope and type gestures on either PapierCraft paper sheets or on a Tablet PC in response to the stimuli (see Figure 5).

In the Tablet PC condition, the documents were presented on a slate Tablet PC and all the interactions were performed directly on the screen using a stylus. The size of the documents shown on screen was adjusted to be roughly the same size as a paper copy. Task feedback was presented directly on the screen in the form of a marking menu with a 500ms delay before the menu will pop-up.

In the PapierCraft condition, the documents were printed on letter-sized Anoto paper and all input was performed using the new wireless multimodal pen described above (see section 3.2). The pen was connected, via a BlueTooth link, to the same Tablet PC that serves as a base station. The base station received strokes from the pen, processed them in real time, and then presented stimuli and results on the vertical monitor. It also simulated auditory feedback through a speaker for the feedback pen. In this condition, all task feedback was using the pentop feedback system. Both conditions used the same Tablet PC and gesture recognition algorithms for user input processing.

4.2.1 Digital content used

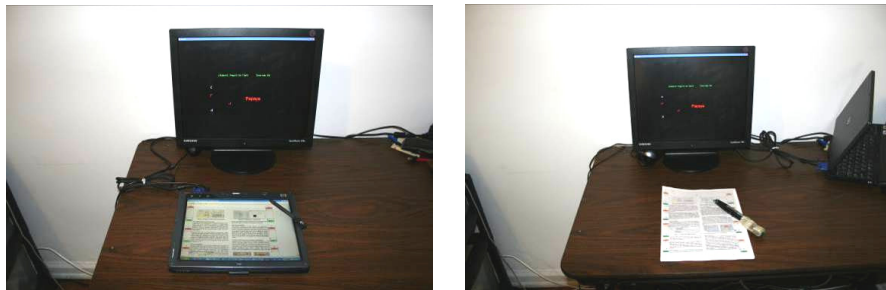


Figure 5. The experiment setting. (Left) The Tablet PC condition. (Right) The PapierCraft condition with the base station nearby. The Tablet PC was used with dual-display configuration.

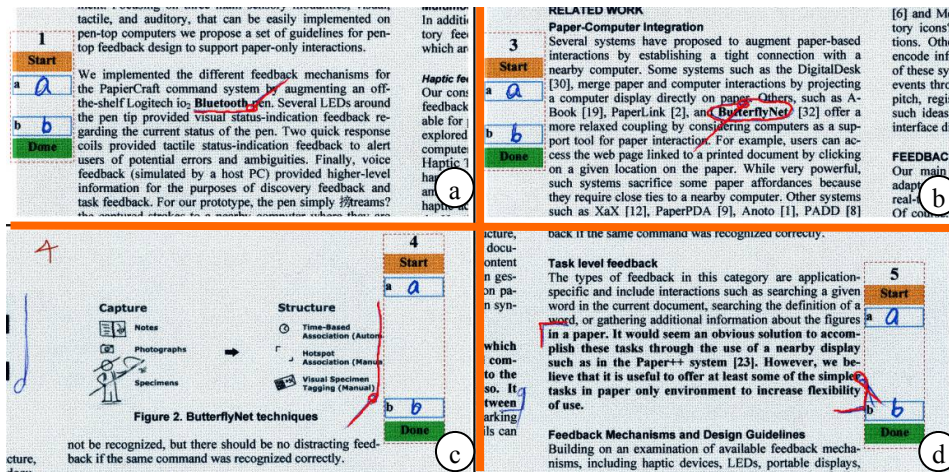


Figure 6. The input pages on paper (a digital version was used for the Tablet PC condition) with 4 scope types. (a) and (b) keyword pages with the underline and the lasso to select keywords. (c) and (d) block pages with the margin bar and the cropping mark to select a paragraph and a figure, respectively. The gestures are highlighted in red for clarity.

We chose several pages from different documents in ACM Digital Library as the study’s input documents, which represent typical technical articles for active reading. To test different command scope gestures, we prepared two types of pages, namely keyword pages and block pages. In the keyword pages (Figure 6 (a) and (b)), participants were asked to use underlines and lassos to precisely select the designated keywords in bold. For generality, we alternately used one-word and two-word targets, and manually chose such words that appear evenly distributed within a page. In block pages (Figure 6 (c) and (d)), participants had to use margin bars and cropping marks to select several lines of text or a figure. Again, we also arranged the target blocks evenly on the pages.

To help participants follow the designed order of the trials, we used an input box next to each target word or block in the document margin (Figure 6). An input box is labeled with an ID and consists of the “start” and “done” buttons, as well as the blank boxes for letters. Participants just followed the IDs to sequentially trigger the stimuli and input the desired annotations and gestures.

4.3 Performance Measures

For quantitative performance measures, we focused on *task completion time* and *error rate*. Task Completion Time is the duration of a trial, from the instant participants tap the “start” button to the instant they tap the “done” button. Error Rate is the percentage of trials that do not match the stimuli. The major error types included: *wrong command scope*, *wrong command type*, *inaccurate selection* and *wrong pen mode*. The inaccurate selection error means that the selected document content is not the requested one, even if

the shape of the command scope gesture is correct. For keyword pages, participants were required to select the exact target words. In contrast, for block pages, it was acceptable if at least 80% of the requested block was selected.

4.4 Experiment Procedure

Each condition began with a training block of a keyword page and a block page, followed by 4 testing blocks of 4 pages (1 for each type of scope gestures, namely *underline*, *lasso*, *margin bar* and *cropping mark*). Each page contained 8 trials, 1 for each of the 8 different command types. Between two consecutive blocks, there was a 5-second break enforced by the experiment program.

For each condition, participants went through 128 trials (4 blocks \times 4 different scopes \times 8 commands per page). We adopted two different sets of background documents and command menus for the two conditions in counterbalanced order, but the training and testing blocks in the same condition used identical menus.

To help participants achieve a near-expert skill level, we gave them a “cheating” card showing the directions of the tested menu items during the first block. Afterwards, the card was removed, and participants needed to refer to the pen-top or the on-screen feedback for unfamiliar menu items. To keep the error rate in check, we showed the current error rate of the last 16 trials on the stimulus screen. Once the rate was above 12.5% (i.e., more than one out of 8 trials were wrong), the error rate number would turn red and participants were reminded to slow down a little. One experimenter sat near participants for observing their behavior, taking notes, giving instructions and answering questions.

4.5 Results

We recruited 12 participants (8 male and 4 female) from the university campus including undergraduate students, graduates students and a visiting scholar. Seven of them had 0.5~7-year experience of using styli on a tablet PC or a PDA. Three participants had used Nokia or Logitech digital pens in the last 3 years occasionally. None of them was familiar with the PapierCraft gesture command system. We compensated each participant 20 dollars for about 90 minutes of his or her time.

After collecting data from participants, we grouped data into sets based on the trials' technique \times block \times scope gesture setting, and then removed all outlier trials with a task completion time more than 3 times standard deviations (SD) away from the mean of its group. We considered the scope gesture in outlier removal, as different scope gesture shapes inherently require different drawing time. In total, 1.82% (56 out of 3072) of the

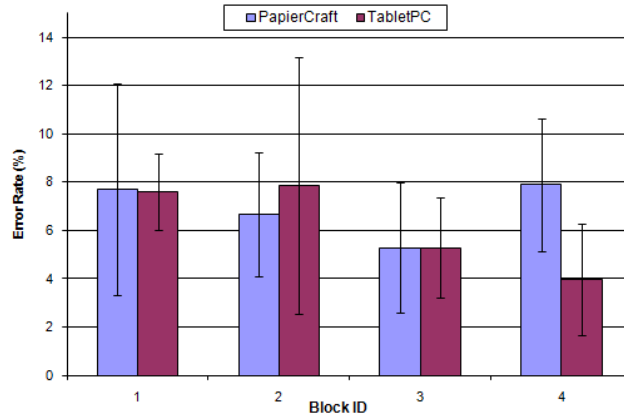


Figure 7. Mean error rate over blocks.

data points was removed. We used Greenhouse-Geisser correction to account for deviations from sphericity and Bonferroni corrections for all post-hoc tests.

4.5.1 Error Rate

We examined the percentage of trials that had at least one type of errors in a block (32 trials). Figure 7 shows the mean error rate per block for each condition. Overall, the mean error rate over blocks remained below 8% in both conditions. The repeated measures technique \times block \times scope gesture 3-way ANOVA found no main effects. There was no interaction found, either.

4.5.2 The Speed of Drawing PapierCraft Gestures

We first examined users' learning behavior in terms of task completion time. The repeated measures technique \times block \times scope gesture 3-way ANOVA showed a main effect of block ($F(3, 33) = 44.54, P < .01, \eta_p^2 = .80$), reflecting a normal learning behavior. There was no main effect of technique, but these results are qualified by a technique by

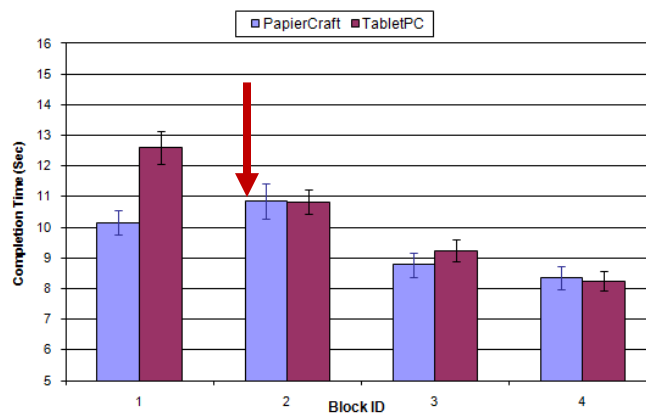


Figure 8. The Block by Technique interaction in mean task completion time (95% confidence interval). The red arrow indicates a slight increase in completion time in the PapierCraft condition due to the removal of the cheating card.

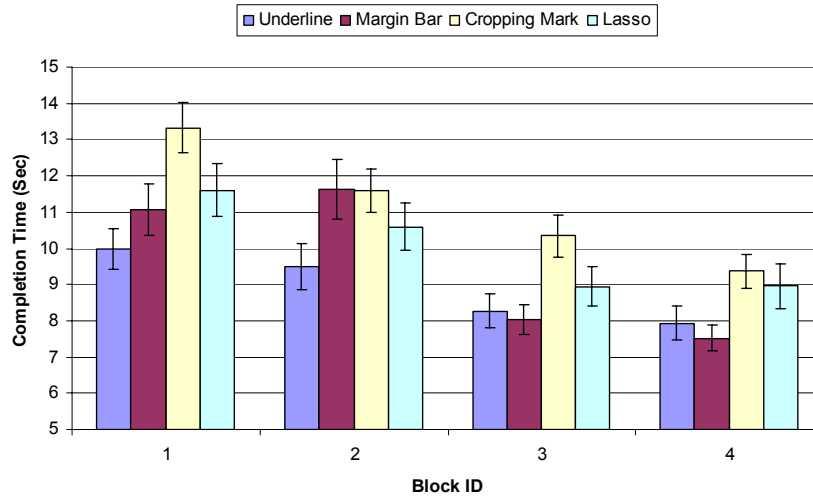


Figure 9. The Block by Scope Gesture interaction in terms of mean task completion time (95% confidence interval)

block interaction ($F(3, 33) = 6.16, P < .01, \eta_p^2 = .36$), suggesting a different learning curve for each technique, as shown Figure 8. Based on our observations, we believe it might have been caused by the removal of the cheating card at the beginning of block 2. Due to the paper's weak visual feedback, participants tended to use the card much more than in the Tablet PC condition, and thus suffered more from the card removal. Because of the above interaction, we also examined the task completion time of the last block. A paired-samples T test $t(11) = -.86, p = .41$ showed no significant difference between PapierCraft ($M = 8155, SD = 3543$) and Tablet PC ($M = 8734, SD = 3449$). This result suggests that novice users of PapierCraft can learn the paper interface as effectively as the Tablet PC interface.

Scope gesture was found to be a main effect ($F(3, 33) = 21.97, p < .01, \eta_p^2 = .67$). This effect is qualified by a block by scope gesture interaction, ($F(9, 99) = 5.57, p < .01, \eta_p^2 = .34$), suggesting a different learning curve depending on the scope gesture used. As Figure 9 suggests, the margin bar costs slightly more time than the underline at the beginning of the test, since participants were unfamiliar with this new mark. However, after the practice of three blocks, there was no statistically significant difference between the two in the fourth block. In addition, we did not find significant interaction between technique and scope gesture. There was no (technique \times block \times scope gesture) interaction, either.

4.5.3 More Discussion about Error Rate

To obtain more insights into the users' performance, we further divided the errors into four categories: *wrong command scope*, *wrong command type*, *inaccurate selection* and

others (including *wrong pen mode*). The average number of errors per user in each category (out of 4 blocks \times 32 trials = 128 trials) is presented in Figure 10. Paired-samples T-tests showed that participants made significantly more Command Type errors over the four blocks with PapierCraft ($M = 3.50$, $SD = 2.20$) than Tablet PCs ($M = 0.33$, $SD = .89$), $t(11) = 5.27$, $p < .01$. This is consistent with the novice user performance reported in the previous work [Liao, et al., 2006], in which Liao et al. attributed the difference to the stronger visual feedback of the Tablet PC screen. Also interesting is the fact that participants made marginally less Target Selection errors with PapierCraft ($M = 2.83$, $SD = 1.59$) than with Tablet PCs ($M = 4.83$, $SD = 3.21$), $t(11) = -2.17$, $p = .053$. Based on our observation and post-test interview with participants, one explanation might be that the Tablet PCs have lower display quality and a smoother writing surface than PapierCraft, which make the selection of small document content more difficult. The parallax on the screen of the Tablet PC could be another cause.

The above analysis of error rate and task completion time has demonstrated that users are able to learn to draw PapierCraft pen gestures at computer-comparable accuracy and speed after a short period of time of training, but it might take a longer amount of time to establish a stable skill level that is robust to interferences such as time pressure. We leave the latter issue for future longitudinal experiments. Based on this gesture level study of user experience, we further examined the application level user experience of PapierCraft, in order to understand how well these gestures are used for typical tasks in active reading.

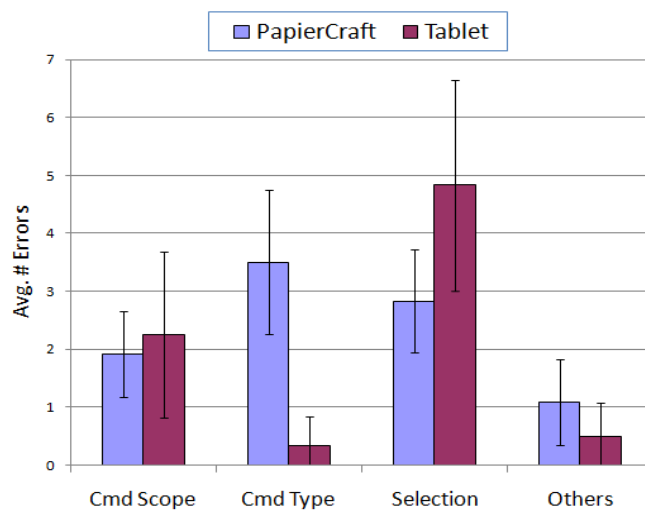


Figure 10. The average number of errors in the four categories (95% confidence interval)

5. EVALUATION OF PAPIERCRAFT FOR ACTIVE READING

Active reading often involves both paper and computers. Paper is used for reading and annotating, for it is easy to annotate, convenient to navigate and users can easily compare documents side by side. At the same time, people use computers for digital affordance such as searching a given word or indexing notes. We hypothesized that using PapierCraft to enable digital affordance directly on paper would improve user experience. To test this hypothesis, we examined how PapierCraft can be integrated into a comprehensive set of active reading activities, going beyond the study of gestures themselves in the last experiment.

We began with an exploration of the active reading technology. We projected the design space along 3 main design vectors: 1) the overall physical quality of the display for reading documents; 2) the in-situ digital functions provided and 3) the number of display surfaces provided. Figure 11 illustrates the relative positions of three representative techniques, namely normal paper, dual-display Tablet PCs and PapierCraft, in this design space. Paper offers a great display surface for both text and annotations, and makes it easy to create multiple displays, but paper lacks any digital functionalities. A Tablet PC offers a full range of digital affordance but only provides one display, making navigation either within or between documents more complicated. For a fair comparison in terms of the number of displays, we therefore opted for two connected Tablet PCs for our experiment. Furthermore, the Tablet PC screen is known to be sub-optimal for writing [Morris, et al., 2007]. PapierCraft can leverage all the advantage of paper while offering some of the digital affordance provided by Tablet PCs. Of course, it is often difficult to evaluate the relative influence of the limited feedback offered by the PapierCraft system or the overall feeling for the pen on the screen. This is why we

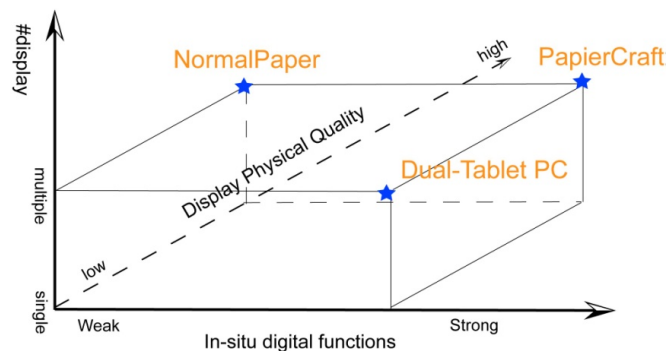


Figure 11. The design space of active reading interfaces.

decided to conduct an empirical comparison of the three systems in the context of active reading: standard paper, PapierCraft and two Tablet PCs linked together. By comparing PapierCraft with normal paper, which has no in-situ digital functionality, we were able to study the effect of the digital capability of PapierCraft; by contrasting PapierCraft to the Tablet PCs, we were able to examine the impact of paper affordance on active reading interaction. Together, these two comparisons would allow us to better understand how well PapierCraft might bridge the gap between paper and digital affordance.

5.1 Device Setting

When considering the setting for the experiment, our goal was to simulate a typical knowledge worker configuration. If we look at the paper setting as a starting point, users will be using several printouts for reading and often a notepad for taking notes. But they might also use a laptop for composing text and accessing external resources available through the web [Morris, et al., 2007]. Beside the inherent reading quality of paper, the multiple displays are compelling, as they simplify navigation between documents [O'Hara and Sellen, 1997]. Therefore, we decided that for each condition, we should provide at least two display surfaces with pen input for reading and annotations and a laptop for composition. The three conditions *normal paper*, *PapierCraft* and *Tablet PC* are shown Figure 12.

Paper Condition

In the Paper Condition, participants used a normal ink pen, letter-sized printouts, and several blank white sheets as scratch paper. Information transfer from one display to another requires participants to either write or type the information to be transferred.

PapierCraft Condition

In the PapierCraft Condition, participants used the multimodal feedback pen described above in conjunction with Anoto-enabled printouts and scratch paper. The laptop processed the pen strokes in real time and executed pen gesture commands accordingly. To limit the influence of learning behavior on our results, we decided to use a limited version of the PapierCraft system with only 3 commands (Figure 13): Web search, Copy,



Figure 12. Hardware configurations: (Left) normal paper (Middle) PapierCraft (Right) Tablet PC

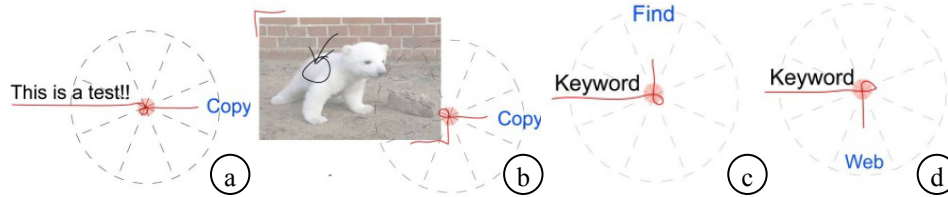


Figure 13. The gesture command set. (a) (b) “Copy” for texts and pictures. (c) “Find” for keyword search within documents. (d) “Web” for web search or opening a web page

and Find. Furthermore, we simplified the command semantic in the following way: the copy command supports two scopes, *cropping mark* for pictures and *underline* for text (Figure 13-a, b). We further automated the copied data type selection: if the percentage of text area within the command scope was higher than a threshold, the system categorized the selection as pure text, otherwise as a bitmap image in the clipboard of the laptop, from which it is easily pasted into a Word document. For the find command (Figure 13-c), we used the pen-top feedback to indicate the positions of hits: first, the system said out loud the rough position of the next hit, and then the LED on the tip of the pen showed the precise line as described in [Liao, et al., 2006]. Finally, the Web command (Figure 13-d) only supported the underline scope, and if the selected text was part of a URL, then the system automatically figured out the full URL and opened the web page in a browser on the laptop. Otherwise, it would issue a search on the selected words using Google.

In addition to the above gesture commands, we provided an annotation management facility, the Snippet Reviewer (see Figure 14), which is similar to XLibris [Schilit, et al., 1998]. As participants were annotating the printed documents, the program identified annotations that looked like an underline or a circle, and then generated snippets composed of the selected text and a snap-shot picture of the document segment. Participants could use this application to quickly access the information they annotated and copy it directly into the editing program.

Tablet PC condition

In the Tablet PC condition, participants were provided with two slates (Tablet PC without a keyboard). One was the NEC Versa LitePad Tablet and the other was the HP TC1100 Tablet. Both were equipped with a 10.4 inch XGA (1024 x 768 pixels) display. Participants could use these slates for both reading and note-taking. The computer interface on these slates supported the exact same pen gestures and dynamic information rendering (e.g., opening a browser or running the Snippet Reviewer on the laptop) as the PapierCraft condition. The difference regarding the system feedback was: 1) on the Tablet PC, a popup marking menu showed up after a 500ms pause after a pigtail crossing;

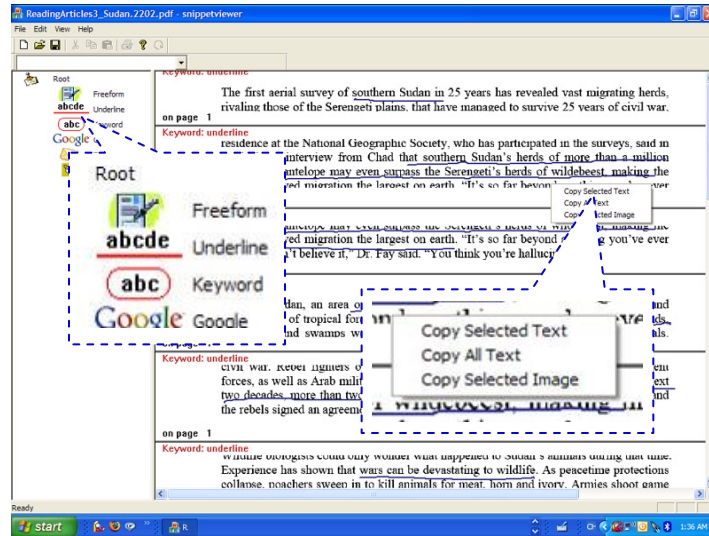


Figure 14. Snippet Reviewer. Users can review the annotated document content and copy the underlined or circled text via the popup menu

2) during a keyword search, the Tablet PC interface automatically flipped to the hit page, and highlighted the target word. With this stronger visual feedback, this condition was expected to engender better user experiences than the other conditions for this search task.

5.2 Experimental Tasks

Given our focus on the active reading activities, we first considered a synthesis task in which participants were asked to use the interface to read a reasonably long article (e.g., 4 pages), write a summary and answer detailed questions about what they read. This task had been used extensively in past experiments [Morris, et al., 2007, O'Hara and Sellen, 1997] and is believed to be a good representation of active reading, because it encourages participants to process information as they read and requires a variety of display-related interactions, such as annotating, note-taking, navigating, browsing and copy/pasting. However, the summarization of a text is a free format task, so there is no guarantee that participants would perform specific interactions, such as copy/paste and web searches. To address this problem, we decided to add a directed tasks section, encouraging participants to search specific keywords on the web, open URLs, annotate and compare figures, and find words within documents.

5.2.1 Summary Tasks

Participants were asked to read one 4-page article of about 1100 English words in length. The articles were selected from the New York Times to ensure that average people could understand them and finish reading within 10-15 minutes. The articles' difficulty levels were roughly equal, judged from our pilot test. Users were encouraged to annotate the text as they saw fit during their reading. To ensure that participants could finish all of the

tasks in 2.5 hours, we simplified the summarization procedure by asking participants to first write a thesis statement (one or two sentences) for the whole article and then one summary sentence for each paragraph in the order they appear in the text. Although no final, formal summary was generated, all of the important sub-tasks of active reading, such as reading and annotating, took place in our task. Our pilot tests showed that, no matter whether participants wrote the final summary or not, they exhibited similar reading, annotation and navigation interactions. Because this task relies heavily on paper quality, we hypothesized that paper and PapierCraft would do well in this task.

5.2.2 Directed Tasks

For these tasks, we asked participants to 1) perform a web search for an unfamiliar word; 2) follow a link to check document-related information on the web; 3) review illustrations; and 4) search for a keyword in the document. These tasks reflect digital affordance, so we hypothesized that the tablet PC and PapierCraft condition would fare better during these tasks.

In the **web search task**, participants were to first locate a given word in the article, and then drew a star mark in the margin next to the word, underlined the word, and searched the meaning of the word on the internet. Upon finding the definition, they should copy a one-sentence

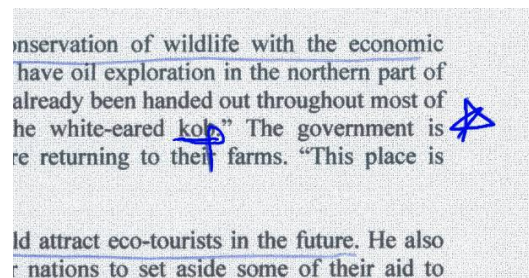


Figure 15. An example of a web search task for a keyword "kob". The participant first drew a star, then underlined the keyword, before looking up the word on the web. Marks are highlighted for clarity.

explanation of the word from Wikipedia into their notes (see Figure 15 for an example). There were 3 keywords spread over 3 pages for this task. The star mark and underline aimed to emulate the interweaving of reading and searching.

In the **link following task**, participants accessed a web URL embedded in the article (see Figure 16). This link provided further information about the subject matter of the article, and participants were asked to copy one picture from the web page to the summary document that they were editing. The web pages were randomly selected but always relevant to the article. There were three such hyperlinks in three pages.

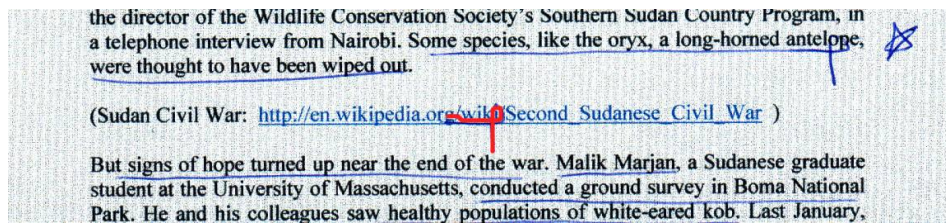


Figure 16. An example of the link following task. Participants could use the “web search” command (highlighted in red for clarity) to select any part of the URL to open the web page.

For the **illustration review task**, participants were asked to describe a given picture or illustration using the most effective means including using text, sketches, excerpts and annotations. For example, for Figure 17, they were asked to find a mountain, a river and a dam and briefly describe the targets’ relative locations in the picture. Participants were also asked to compare 2 figures to find and describe their similarities and/or differences. For example, for Figure 18, they were asked to identify and roughly describe the locations of three differences between the two pictures displayed on different pages. This task aimed to test user experiences of information transfer and multiple-display operations.

For the **keyword finding task**, participants were asked to answer 3 questions about specific details in the text by searching for keywords in a document. For example, with the question, “*What did Michael Leinbach say about the launch,*” participants were expected to first locate keywords (e.g. Michael Leinbach) within the article and then find the answer to the question in the context paragraph.

To limit knowledge transfer between the summary and keyword finding tasks (e.g., short-term memory of the location of a keyword that was just read), we used a different article on a similar topic for the keyword finding task. Choosing a relevant



Figure 17. One figure used in the illustration review task. Participants were asked to identify the relative position of entities like the mountain, river and dam.



Figure 2. (left) International Space Station in 2008

The new mission includes three spacewalks to maintenance and test techniques for cleaning a tray of the station's power supply. That joint 10 feet is



Figure 5. Previous configuration of the International Space Station

Figure 18. An example of figures for page comparison. (Left) a figure on page 2 (Right) another figure on page 4. Participants were asked to identify 3 differences.

second article maintained a somewhat continuous thought flow and reduced possible cognitive load caused by sudden topic switching. To limit the testing time, we set a two-minute threshold on each question. Participants had to stop even if they did not find the answer.

5.3 Procedure and Measures

Considering the possible cross-participant variance, it was judged best to use a within-subject experiment with 3 counter-balanced hardware conditions. For each condition, the testing began with a training session to help the participant familiarize with this condition. In an effort to encourage the most natural behavior as possible, participants were told that they had the flexibility to choose whatever tools they thought most effective and efficient to finish the tasks. An experimenter sat near participants for observing their behavior, taking notes, giving instructions and answering questions.

During the testing session, participants first read a given article (the specific articles were counter-balanced), and then typed the summary on the laptop. Then, they proceeded to the directed tasks shown on the laptop. They referred back to the reading material, performed the appropriate operations, and composed answers to the questions on the laptop. When all three conditions were finished, participants filled out a self-report questionnaire to rate their user experience of each task for the different interfaces. Questions were grouped into two categories, covering six paper-friendly tasks (*read, annotate, note*¹, *navigate/browse, spatially arrange* and *compare*), in which paper affordance played a key role, and five digital-friendly tasks (*use of annotation, share, retrieval, information transfer* and *keyword search*), in which in-situ digital functions make a difference. For instance, we used a question, “*Do you think it is easy to navigate*

¹ “annotate” means free form handwritings within the reading material; “note” means those on separate paper sheets.

and browse (e.g., flip pages back and forth) the documents and notes? (1 for 'Very Difficult' and 7 for 'Very Easy')” to elicit the users’ experience of the navigation and browsing tasks. A detailed list of questions can be found in [Liao, 2009]. For each question, we asked participants to rate each technique on a 1~7 Likert scale (1 for most negative experience).

We also asked participants to elaborate on the reasons for their choices, in order to gain insights into why and how participants scored a specific interface for a task. We did not examine the exact task completion time or error rate, as we focused on the high level user experience, not on the low level interaction techniques. The whole experiment was designed to be finished within 2.5 hours and participants received \$30 for their participation.

5.4 Participants

We recruited 12 (6 male and 6 female) participants from the university campus. 11 of them were at the age of 21~30, and one in his thirties. One was a college student and all others were graduate students. Four participants had background in humanity (e.g. psychology and education) and 8 majored in computer science or electronic engineering. Seven participants had 0.5~4-year experience of using styli on a tablet PC or a PDA. One used an Anoto digital pen for note-taking for 2 years, and two knew the concept of digital pens. None of them had used a PapierCraft-like system before.

5.5 General Results

We used the Likert scale as a continuous variable through the following statistic analysis. In the following discussion, except if otherwise noted, Greenhouse-Geisser correction was used to account for deviations from sphericity. All post-hoc tests used Bonferroni

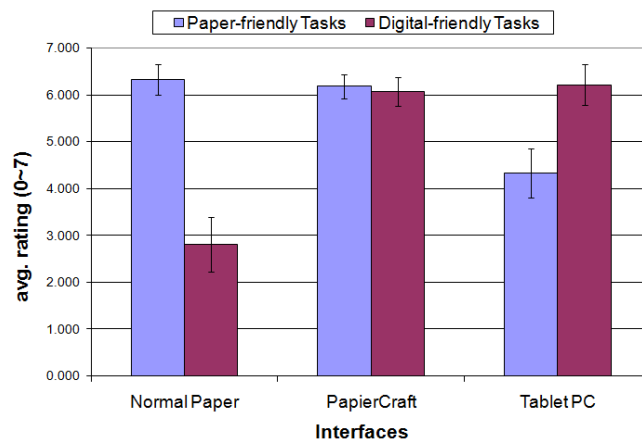


Figure 19. Averaged user ratings on the three interfaces per category (95% confidence interval).

corrections.

To evaluate the user experience while using each technique with respect to the task categories, we averaged every participant's responses within each (technique \times category) block. This data is shown in Figure 19. We ran a repeated measures technique \times category two-way ANOVA on the results, and found category a main effect ($F(1, 11) = 9.32, p < .05$ and $\eta_p^2 = .46$), and technique also a main effect ($F(2,22) = 26.95, p < .01$ and $\eta_p^2 = .71$). There is a strong category by technique interaction, $F(2,22) = 142.70, p < .01$, and $\eta_p^2 = .93$.

Because of this interaction, we then looked at each technique separately. For PapierCraft, the paired-samples T test ($t(11) = .47, p = .65$.) showed no significant difference between paper-friendly tasks ($M = 6.18, SD = .46$) and digital-friendly tasks ($M = 6.07, SD = .54$), which suggests that PapierCraft can provide a very good user experience in both task categories. In contrast, for normal paper, the paired-samples T test found that participants rated it significantly lower in digital-friendly tasks ($M = 2.82, SD = 1.03$) than in paper-friendly tasks ($M = 6.33, SD = .58$), $t(11) = 13.99, p < .01$; similarly, the Tablet PC was rated significantly lower in paper-friendly tasks ($M = 4.33, SD = .92$) than in digital-friendly tasks ($M = 6.22, SD = .76$), $t(11) = -6.26, p < .01$. This result highlights the combined advantages of PapierCraft in terms of both paper and digital affordance. To better understand the user responses and the pros and cons of each

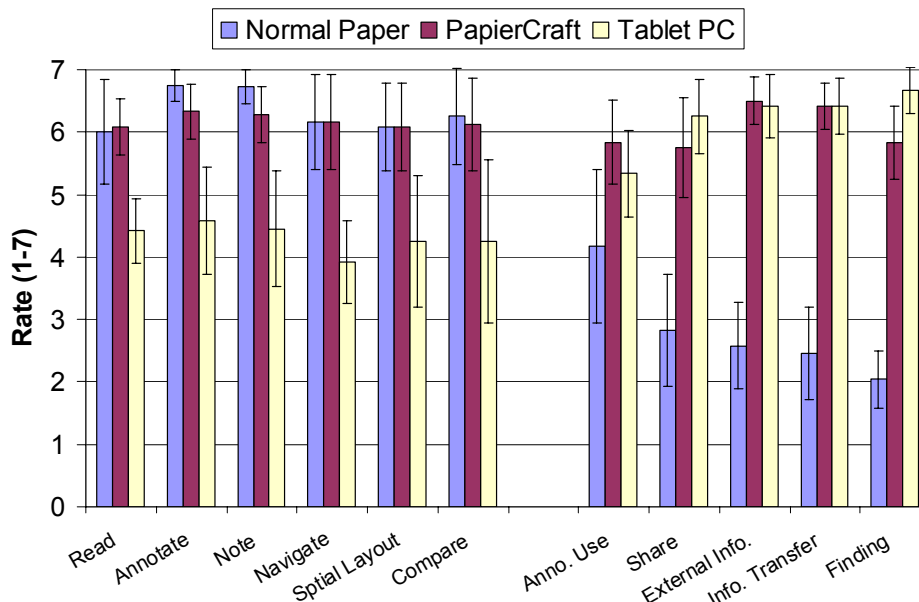


Figure 20. The mean and confidence interval of the user experience of the experiment tasks. (Left) Paper-friendly tasks (Right) Digital-friendly tasks.

Subtasks	Found Main Effect: Interface Type			Normal Paper		PapierCraft		Dual-Tablet PC	
	F	p	η_p^2	M	SD	M	SD	M	SD
Read	F(2, 22) = 15.02	< .01	.58	6.00	1.48	6.08	.79	4.42	.90
Annotate	F(1.19,13.12) = 21.44	< .01	.66	6.75	.45	6.33	.78	4.58	1.51
Note	F(1.29,12.91) = 17.50	< .01	.64	6.73	.47	6.27	.79	4.46	1.64
Navigate/Browse	F(2, 22) = 22.09	< .01	.67	6.17	1.34	6.17	1.34	3.92	1.17
Spatial Layout	F(2, 22) = 13.18	< .01	.55	6.08	1.24	6.08	1.24	4.25	1.87
Compare	F(1.03, 11.35) = 10.68	< .01	.49	6.25	1.36	6.13	1.32	4.25	2.30

Table 1. Results of the repeated measures 1 way ANOVA on ratings of the paper-friendly tasks

interface, we investigated specific tasks one by one, as follows.

5.6 Discussion about Paper-friendly Tasks

Our data shows a consistent pattern across all questions covering paper-friendly tasks (Figure 20-left). By running a repeated measures one-way ANOVA on the ratings of each task, we found technique a main effect ($p < .01$), with both the paper and the PapierCraft condition rated significantly better ($p < .01$) than the tablet condition in post-hoc pairwise comparison for all the six tasks. The detailed statistic analysis results are summarized in Table 1. They are consistent with the results reported by [O'Hara and Sellen] but seem to contradict the results presented by Morris et al, who reported that recent computing devices, such as Tablet PCs, have proved to be comparable or even superior to paper in reading, annotating and note-taking [Morris, et al., 2007]. This inconsistency might be attributed to the different hardware and experiment tasks that were used in our experiment. We now look at each aspect in order to better understand the differences between the three techniques.

With respect to reading, the main complaint about PapierCraft was the diminished contrast caused by the Anoto pattern. This is a known design issue, especially when using a pattern printed by a laser printer. This problem could be avoided altogether by using a human-invisible yet infrared-absorbent printing ink [Yousaf and Lazzouni, 1995], although the cost will be higher and the digital pen may need some changes to recognize the special ink. We also observed that most participants tended to move and hold the paper



Figure 21. A participant held the Tablet PC for a comfortable reading posture.

documents at an angle during reading. This affordance is clearly very important, since two participants tried to emulate it while using the Tablet PCs (Figure 21), although they admitted that doing so was a little awkward with the Tablet PCs. Lighter and thinner displays will be the key for such interaction. This observation is indeed consistent with Morris' work, which indicates "In contrast to the quick and somewhat effortless ergonomic adjustment in the paper, tablets, and vertical conditions, the horizontal condition required more thought and effort for 75% (9) of our participants " [Morris, et al., 2007].

New reading devices such as Amazon Kindle [Amazon, 2010] and Apple iPad 2 [Apple, 2011] could alleviate, to a degree, this inconvenience with their slimmer and lighter form factor, but they still can hardly beat paper in terms of ergonomics and spatial flexibility, due to their relatively heavy, thick, small and rigorous form factor.

So far as screen resolution, some emerging devices like iPhone 4 [Apple, 2010] have already a DPI of 326, which is comparable to normal printing DPIs and thus provides better user experience than what we had in the experiment. Nevertheless, it is very hard to make a larger display (say Letter size) of such a high DPI, since the existing mobile devices are not computationally powerful enough to handle the large number of pixels (e.g. ~5M pixels). Furthermore, none of these mentioned devices supports annotation as well as paper, although this capability is very important to active reading user experience.

With respect to navigation, spatial arrangement and page comparison, while participants' evaluations were clearly contrasted, their comments reflected a mixed feeling. In particular, they complained that the software navigation buttons (25 x 25 pixels) on our slates were too small. This design issue could be addressed with hardware buttons. Participants also pointed out that while printouts make it easy to spread information around, they also make it difficult to keep the workplace organized. In particular, one could easily lose page number information or could spend time rearranging a given pile. In that respect, digital content is much better. Participants also commented on the advantage of having two digital displays for the comparison task. This result confirmed the findings of O'Hara [O'Hara and Sellen, 1997] that multiple page operations can be better served with large and/or multiple displays. In this direction, the dual-display E-Book reader proposed by Chen et al. [Chen, et al., 2008] is on the right track among purely digital solutions. However, the currently available devices still prevent users from arranging two tablets as easily as paper, which was suggested by

previous research by Morris [Morris, et al., 2007]. Further, participants were also concerned about the cost of the two computers.

With respect to annotations and note taking, we noticed that in our experiment, none of participants took any notes on blank paper or on the Tablet PC, presumably because they preferred typing on the laptop when it came to composing text, as observed by Morris et al. [2007]. As a result, one should assume that the evaluation on note-taking was mainly based on participant's experience of annotating. Overall, participants preferred the normal ink pen to the Tablet PC stylus and the PapierCraft digital pen. The main complaint of participants came from the poor performance of a Tablet PC for writing. Participants reported that the Tablet PC pen was too small and slippery, and the pen was unable to pick up small hand movements. This is a known issue, and more research is required. Participants also complained about the limited screen real estate in the tablet condition, which re-confirmed Morris' findings [Morris, et al., 2007].

Another issue recently emerges that many new mobile reading devices including iPhone [Apple, 2010] and iPad [Apple, 2011] use capacitive input via fingers and do not provide built-in support for pen-based interaction. This limitation causes inconvenience for active reading, which often requires precise in-text annotation. To address this issue, people created some special pens like BoxWave Capacitive iPad Stylus (<http://www.boxwave.com>) and Pogo Sketch (<http://tenonedesign.com/sketch.php>) that mimic the touch of a finger, but these pens do not deal well with palm rejection and thus the users have to hold their hand in the air while writing or write at the bottom of the screen, which is awkward for long time use.

Regarding PapierCraft, participants commented that the pen was too big and unbalanced and thus made it difficult to hold the mode-switch button while drawing a pen gesture. These issues are not surprising when considering the design of our prototype, and could be easily addressed by a better industrial design.

5.7 Discussion of Digital-friendly Tasks

Similar to the analysis of paper-friendly tasks, we ran a repeated measures one-way ANOVA on the ratings of the digital-friendly tasks. The statistic results are summarized in Table 2. With the exception of Annotations Use and Management, our data also show a consistent pattern (Figure 20-right): for each task, we found the factor technique a main effect ($p < .01$), with both the Tablet PC and the PapierCraft condition rated significantly better ($p < .01$) than the paper condition in post hoc pair-wise comparison.

Subtasks	Found Main Effect: Interface Type (except *)			Normal Paper		PapierCraft		Dual-Tablet	
	F	p	η_p^2	M	SD	M	SD	M	SD
*Anno. Use	F(1.28,14.13) = 3.78	= .06	.26	4.17	2.17	5.83	1.19	5.33	1.23
Share	F(2, 22) = 26.23	< .01	.71	2.83	1.59	5.75	1.42	6.25	1.06
Ext. Info. Retrieval	F(1.19,13.11) = 80.11	< .01	.88	2.58	1.24	6.50	.67	6.42	.90
Info. Transfer	F(1.10, 12.09) = 66.23	< .01	.86	2.46	1.31	6.42	.67	6.42	.79
Keyword Finding	F(2, 22) = 107.05	< .01	.91	2.04	.81	5.83	1.03	6.33	.49

Table 2. Results of the repeated measures 1 way ANOVA on the ratings of the digital-friendly tasks.

Participants enjoyed the ability to automatically capture notes and share them through a digital channel. They also found the in-situ access to digital information in either the Tablet PC or PapierCraft systems very convenient, noting that in the paper setting, they would have to carefully transcribe the information and it was too much of a hassle to look up a word on paper. Interestingly, we noticed that two participants started to use the web search command for unfamiliar words while working on the summary task, despite the fact that the command was supposed to be used later for the directed tasks. This suggests that the function is well accepted and easily learned by users.

With respect to information transfer, one of the most important features of the PapierCraft and TabletPC condition was probably the copy-paste command between displays. This was the most frequently used command in the experiment for both the summarization and the directed tasks (see Figure 22). Interestingly, users tended to use copy-paste whenever possible, even though sometimes it might take more time to do pen-keyboard switching and draw the command gesture than to simply type a few words. This could be explained by the fact that participants often used copied text as a context for composition. When we compared the text selected by the copy command (identified by the drawn copy gestures and logs) against the resulting summary, we found that many

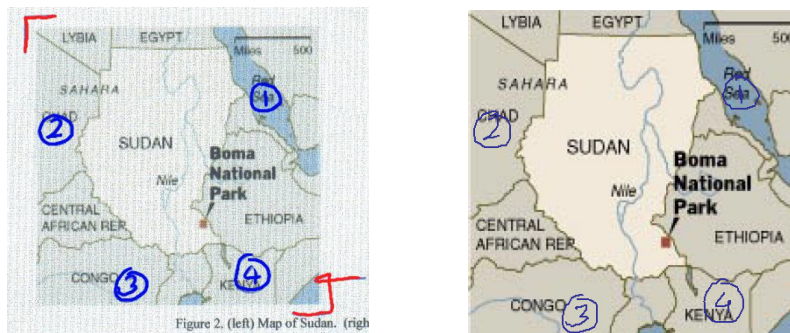


Figure 22. An example of pictures copied from paper with annotations, in which the participant intuitively described the positions of the POI (point of interest) with hand-written numbers. (Left) the marks on paper, highlighted for clarity (Right) the segment copied into the laptop

times, the two were actually different, since participants had rephrased text and/or changed word order. The benefit of this practice is that users did not have to frequently switch focus between their summary and the original article during the composition phase. This phenomenon also confirms the popularity of the copy-paste function during summarization in Morris's work [Morris, et al., 2007]. This function was also used extensively during the directed task section: 97.2% of quiz questions for which a picture copy-paste could be used were actually answered with copied pictures with annotations.

Participants liked the keyword search facility offered in the PapierCraft and Tablet PC condition. While PapierCraft offered a somewhat limited capability when compared with the Tablet PC, participants found it useful compared to the paper condition. Another issue with PapierCraft was the unwanted pen ink created when a user scanned the pen in the margin to locate the next target line. These marks could often be distracting to users. This problem can easily be addressed by modifying the firmware of the pen to provide tracking information when either the tip of the pen is pressed against the paper or hovers over the paper [LiveScribe, 2007] with the command button pressed. Participants would also have liked a more salient feedback for the search in line with the work of Song et al. [Song, et al., 2009], in which the pen can project data directly on to the paper document.

Finally, as indicated at the beginning of this section, with respect to annotation use and management, we did not find a significant difference between techniques ($p = .06$). This result was a bit of a surprise, since we had thought the automatically generated digital snippets in the PapierCraft and the Tablet PC condition should have significantly improved user performance in composing the summaries. Several issues might explain this result. First, participants complained that our visualization did not provide enough context for them to process the annotations efficiently. Like the design of XLibris [Schilit, et al., 1998], our design only increased the size of the bounding box of an annotation cluster (see Snippet Reviewer in Figure 14, section 5.1) to include about 2 text lines, which proved insufficient. The position and the layout of annotations within the original pages should have been retained to help users restore contextual information. This feature was missing in the experiment application, but has been implemented in other PapierCraft applications [Liao, et al., 2008]. Another possible cause might be that participants often use underlining as a rough reminder of an interesting part in the text, which is not necessarily the most important part. This resulted in unwanted text extractions that were confusing and thus were abandoned by users. It is interesting to note that one participant mentioned that he was much more careful in drawing regular

underlines with PapierCraft and the Tablet PC than with paper, because he knew that the precise underline could improve the extraction. This response suggests that with time, users can change their regular behavior to adapt to new features for useful functionalities.

6. CONCLUSION AND FUTURE WORK

We have presented two empirical experiments that evaluate the effectiveness of PapierCraft, an interactive paper system, in terms of bridging the gap between paper and computers. Our first experiment indicated that after a short training time (~30min), PapierCraft users can achieve a performance comparable to Tablet PC users in using the gesture-based command system. Our second experiment showed PapierCraft can merge digital and paper affordance in support of active reading, and demonstrated positive user acceptance of such an interactive paper system.

We also learnt some key design issues about PapierCraft. For example, users commented on the uncomfortable feedback pen, the distracting background pattern, the limited scope of the gesture recognizer, and undesired gesture marks left on paper. These findings urge us to continue to refine the feedback pen and the gesture processing mechanism.

Beyond the PapierCraft system, we obtained insights for more general interaction in active reading. First, our experiments confirmed that the feedback pen and a static display like paper are able to support interaction as efficiently and effectively as more traditional Tablet PC-like interfaces. Thus the feedback pen technique could be used for the Amazon Kindle interface, which cannot refresh its display as fast as regular screens and can save power by updating the display content as little as possible. Second, our experiments involved carefully designed testing tasks that aim to capture the typical active reading activities and avoid confounding factors as many as possible. We feel such designs could be useful to inspire future evaluation methods for other active reading interfaces such as Amazon Kindle and Apple iPad.

Our experiment also showed the qualified paper advantages in document navigation and spatial layout due to the difficulties in managing a large set of independent pages, and indicated the limitation of single display Tablet PCs. This encourages dual or tri-display reading devices, such as the one reported in [Chen, et al., 2008], which would offer an optimal number of displays. The user feedback also suggested the importance of striking a better balance between emphasizing the annotated document content and retaining enough easy-to-access context information for the annotations.

These design issues and general insights suggest the directions in which our future research should proceed. Our future work involves better feedback mechanisms, generalizing PapierCraft, mixed-media multiple display interfaces and a longitudinal test. First, our experiment suggested that the PapierCraft user may benefit from higher fidelity visual feedback. This could be achieved by using the latest LiveScribe digital pen [LiveScribe, 2007], which has a built-in LCD screen on the pen. More advanced solutions include using mobile projectors or high-resolution cell phone screens. Although existing work such as PenLight [Song, et al., 2009], MouseLight [Song, et al., 2010] and Paper++ [Norrie and Signer, 2003] has shown promising results, several research problems still remain unsolved, including how to achieve a stable and precise projection without any complex hardware settings that interfere with the common pen-paper interaction, how to reduce the distraction of eye focus switching between a pen and a screen and, when multiple visual feedback channels are available (e.g., the LEDs, projectors, and cell phones are all used), how to choose the best channel for optimal user experience.

The second interesting direction is to generalize the PapierCraft interface, avoiding being tied to the Anoto pen and the dot-pattern paper for the reason that they may not be available or allowed (depending on the circumstances), especially on some legacy documents. Therefore, it would be critical and valuable to study other techniques for real-time document recognition and tracking and pen tracking that do not rely on any special markers or patterns within paper documents *and* still retain the original pen-paper flexibility. Recent advance in camera-phone-based interactive paper [Hull, et al., 2010, Liao, et al., 2010, Nakai, et al., 2006] could help us explore this direction.

Our experiment has already involved multiple paper and electronic displays. This situation will continue and include more and more displays (e.g., interactive paper, projectors, cell phones and screens). The seamless integration of these displays becomes important for providing users with a unified and consistent experience of document interaction. The fluidity of the cross-device interaction is the core obstacle between current models and this goal. For instance, it needs to be determined how one can easily migrate on-going applications and documents across displays without worrying about the different naming mechanisms, access controls, formats, form factors, DPIs, resolution, input devices (e.g., mouse vs. pen) and so on. There has been active research in these fields [Carzaniga, et al., 2001, Hinckley, et al., 2004, Johanson, et al., 2002]. We will further study more general frameworks to support the mixed displays described above.

Finally, one limitation of our active reading test is that the experiment only measured how the interactions in that task could be better supported by PapierCraft, rather than

how well active reading (e.g. the quality of the produced summary) would be improved. In the future, we would like to look into this issue for a more comprehensive evaluation. Moreover, to go beyond the laboratory-scale test reported in this paper, we will deploy the system in a real scenario and conduct a longitudinal test to examine the user experience of PapierCraft in long term. This approach will help us better understand how PapierCraft fits into people's existing workflow, how it interacts with the computer interfaces and how paper-computer bridging changes user behavior in long run.

7. REFERENCE

- ADLER, A., A. GUJAR, L. B. HARRISON, K. O'HARA and A. SELLEN (1998), A diary study of work-related reading: design implications for digital reading devices. *Proceedings of CHI'98*, pp. 241-248.
- ADLER, M. J. and C. VAN DOREN (1972), *How to Read a Book*. 1972, New York, NY: Simon and Schuster.
- AMAZON (2010), Kindle <http://www.amazon.com/kindle>.
- ANOTO (2002), Development Guide for Service Enabled by Anoto Functionality.
- APPLE (2010), iPhone <http://www.apple.com/iphone>.
- APPLE (2011), iPad 2 <http://www.apple.com/ipad>.
- CARZANIGA, A., D. S. ROSENBLUM and A. L. WOLF (2001), Design and evaluation of a wide-area event notification service. *ACM Trans. Comput. Syst.*, 2001. **19**(3): pp 332-383.
- CHEN, N., F. GUIMBRETIERE, M. DIXON, C. LEWIS and M. AGRAWALA (2008), Navigation Techniques for Dual-Display E-Book Readers. *Proceedings of CHI'08*, pp. 1779-1788.
- DYMETMAN, M. and M. COPPERMAN (1998), Intelligent Paper. *Proceedings of EP'98*, pp. 392 - 406.
- EROL, B., EMILIO ANTUNEZ and J. J. HULL (2008), HOTPAPER: multimedia interaction with paper using mobile phones, in *Multimedia'08*, Editor^Editors. ACM: Vancouver, British Columbia, Canada. p. 399-408.
- GUIMBRETIERE, F. (2003), Paper Augmented Digital Documents. *Proceedings of UIST'03*, pp. 51 - 60.
- HEINER, J. M., S. E. HUDSON and K. TANAKA (1999), Linking and messaging from real paper in the Paper PDA. *Proceedings of UIST'99*, pp. 179 - 186.
- HINCKLEY, K., P. BAUDISCH, G. RAMOS and F. GUIMBRETIERE (2005), Design and analysis of delimiters for selection-action pen gesture phrases in scriboli. *Proceedings of CHI'05*, pp. 451-460.

HINCKLEY, K., G. RAMOS, F. GUIMBRETIERE, P. BAUDISCH and M. SMITH (2004), Stitching: Pen Gestures that Span Multiple Displays. *Proceedings of AVI'04*, pp. 23 - 31.

HULL, J. J., X. LIU, B. EROL, J. GRAHAM and J. MORALEDA (2010), Mobile image recognition: architectures and tradeoffs. *Proceedings of ACM HotMobile'10*, pp. 84-88.

JOHANSON, B., G. HUTCHINS, T. WINOGRAD and M. STONE (2002), PointRight: experience with flexible input redirection in interactive workspaces. *Proceedings of UIST'02*, pp. 227 - 234.

JOHNSON, W., H. JELLINEK, J. LEIGH KLOTZ, R. RAO and S. K. CARD (1993), Bridging the paper and electronic worlds: the paper user interface. *Proceedings of CHI'93*, pp. 507 - 512.

LEAPFROG (2005), Fly Pen, <http://www.leapfrog.com>.

LIAO, C. (2009), PapierCraft: A Paper-based Interface to Support Interaction with Digital Documents, PhD thesis, University of Maryland. 2009

LIAO, C., F. GUIMBRETIERE, K. HINCKLEY and J. HOLLAN (2008), PapierCraft: A Gesture-Based Command System for Interactive Paper. *ACM ToCHI*, 2008. **14**(4): pp 1-27.

LIAO, C., F. GUIMBRETIERE and C. E. LOECKENHOFF (2006), Pentop feedback for paper-based interfaces. *Proceedings of UIST'06*, pp. 211-220.

LIAO, C., Q. LIU and B. LIEW (2010), PACER: A cameraphone-based paper interface for fine-grained and flexible interaction with documents. *Proceedings of CHI'10*, pp. 2441-2450.

LIVESCRIBE (2007), LiveScribe pen, <http://www.livescribe.com>.

MACKAY, W. E., D. S. PAGANI, L. FABER, B. INWOOD, P. LAUNIAINEN, L. BRENTA and V. POUZOL (1995), Ariel: augmenting paper engineering drawings. *Proceedings of CHI'95*, pp. 421 - 422.

MACKAY, W. E., G. POTHIER, C. LETONDAL, K. BØEGH and H. E. SØRENSEN (2002), The missing link: augmenting biology laboratory notebooks. *Proceedings of UIST'02*, pp. 41 - 50.

MORRIS, M. R., A. J. B. BRUSH and B. MEYERS (2007), Reading Revisited: Evaluating the Usability of Digital Display Surfaces for Active Reading Tasks. *Proceedings of IEEE Tabletop'07*, pp. 79-86.

NAKAI, T., K. KISE and M. IWAMURA (2006), Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. *Proceedings of 7th Int'l Workshop DAS'06*, pp. 541-552.

NORRIE, M. C. and B. SIGNER (2003), Switching Over to Paper: A New Web Channel. *Proceedings of Web Information Systems Engineering'03*, pp. 209-218.

O'HARA, K. and A. SELLEN (1997), A comparison of reading paper and on-line documents. *Proceedings of CHI*, pp. 335 - 342.

SCHILIT, B. N., G. GOLOVCHINSKY and M. N. PRICE (1998), Beyond paper: supporting active reading with free form digital ink annotations. *Proceedings of CHI'98*, pp. 249 - 256.

SCHÖNING, J., M. ROHS, S. KRATZ, M. LÖCHTEFELD and A. KRÜGER (2009), Map torchlight: a mobile augmented reality camera projector unit, in *CHI'09*, Editor^Editors. ACM: Boston, MA, USA. p. 3841-3846.

SELLEN, A. J. and R. H. R. HARPER (2001), *The Myth of the Paperless Office*. 1st ed. 2001: MIT press.

SIGNER, B. and M. C. NORRIE (2007), PaperPoint: a paper-based presentation and interactive paper prototyping tool, in *TEI'07*, Editor^Editors. ACM: Baton Rouge, Louisiana. p. 57-64.

SONG, H., FRANCOIS GUIMBRETIERE, TOVI GROSSMAN and G. FITZMAURICE (2010), MouseLight: Bimanual Interactions on Digital Paper Using a Pen and a Spatially-aware Mobile Projector, in *CHI'10*, Editor^Editors. p. (to appear).

SONG, H., T. GROSSMAN, G. FITZMAURICE, F. GUIMBRETIERE, A. KHAN, R. ATTAR and G. KURTENBACH (2009), PenLight: combining a mobile projector and a digital pen for dynamic visual overlay. *Proceedings of CHI'09*, pp. 143-152.

SONG, H., F. GUIMBRETIERE, C. HU and H. LIPSON (2006), ModelCraft: capturing freehand annotations and edits on physical 3D models. *Proceedings of UIST'06*, pp. 13-22.

STEIMLE, J., O. BRDICZKA and M. MÜHLHÄUSER (2009), CoScribe: Integrating Paper and Digital Documents for Collaborative Knowledge Work. *IEEE Transactions on Learning Technologies*, 2009. **2**(3): pp 174-188.

TSANDILAS, T., C. LETONDAL and W. E. MACKAY (2009), Musink: composing music through augmented drawing. *Proceedings of CHI'09*, pp. 819-828.

WEIBEL, N., A. ISPAS, B. SIGNER and M. C. NORRIE (2008), Paperproof: a paper-digital proof-editing system. *Proceedings of CHI '08* pp. 2349-2354.

WELLNER, P. (1993), Interacting with paper on the DigitalDesk. *Communications of the ACM*, 1993. **36**(7): pp 87 - 96.

YOUSAF, M. and M. LAZZOUNI (1995), Formulation of an invisible infrared printing ink. *Dyes and Pigments*, 1995. **27**(4): pp 297-303.