

# Microanalysis of Active Reading Behavior to Inform Design of Interactive Desktop Workspaces

Matthew K. Hong<sup>1</sup>, Anne Marie Piper<sup>1</sup>, Nadir Weibel<sup>1</sup>, Simon Olberding<sup>2</sup>, James D. Hollan<sup>1</sup>

<sup>1</sup>University of California, San Diego  
9500 Gilman Dr., La Jolla, CA, USA  
{m1hong, apiper, weibel, hollan}@ucsd.edu

<sup>2</sup>MIT Media Laboratory  
20 Ames St., E14-548K. Cambridge, MA 02139, USA  
simon\_o@media.mit.edu

## ABSTRACT

Hybrid paper-digital desktop workspaces have long been of interest in HCI [5, 39], yet their design remains challenging. One continuing challenge is to support fluid interaction with both paper and digital media, while taking advantage of established practices with each. Today researchers are exploiting depth cameras and computer vision to capture activity on and above the desktop and enable direct interaction with digitally projected and physical media. One important prerequisite to augmenting desktop activity is understanding human behavior in particular contexts and tasks. Here we study active reading on the desktop. To better understand active reading practices and identify patterns that might serve as signatures for different types of related activity, we conducted a microanalysis of single users reading on and above the desktop workspace. We describe the relationship between multimodal body-based contextual cues and the interactions they signify in a physical desktop workspace. Detailed analysis of coordinated interactions with paper documents provides an empirical basis for designing digitally augmented desktop workspaces. We conclude with prototype design interactions for hybrid paper-digital desktop workspaces.

## Author Keywords

Active Reading; Microanalysis; Paper-Digital Workspaces

## ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

## General Terms

Human Factors; Design

## INTRODUCTION

Designing hybrid workspaces that support fluid interaction with both paper and digital media is a central area of HCI research. Early seminal visions of an interactive desk to enable simultaneous work with paper and digital information were proposed by Wellner [39] and MacKay and colleagues [20]. More recently researchers have developed many techniques

to support interaction across paper and digital media [12, 19, 35, 33, 38, 41]. One particularly promising focus for digitally enhanced paper documents is to better support active reading [1]. How might observing the coordination of one's body and materials while engaged in active reading provide insights into the design of hybrid workspaces that better support this cognitively demanding task?

Active reading is a critical task of knowledge workers. Knowledge workers spend substantial time in office environments, where information is spread between digital devices (computers, smart phones, tablets, etc.) and physical media (paper documents, whiteboards, books, printed photos, etc.) [24]. Furthermore, collaborative interaction and information sharing with others is often central to the accomplishment of tasks. The introduction of mobile devices, interactive tabletops, multitouch surfaces, and digital paper presents many challenges for designing workspaces that support fluid natural interaction with information spanning the physical and digital worlds.

In this paper we present a detailed microanalysis of reading paper documents positioned on and above a physical desktop. Our analysis looks closely at coordination of the body and materials. We identify opportunities for leveraging naturally occurring behaviors such as body position, pen grip posture, and paper document manipulation to serve as signatures of the micro-task structure of active reading. We use these observations to derive design insights and prototype digitally enhanced interactions with paper documents. Our aim is to learn from people's existing reading behaviors how a digitally augmented system might integrate with and support focus of attention and the immediate task at hand.

This research is part of a larger project to explore novel paper and digital desktop surface interactions. We are iteratively developing a system that tracks body movement, hands on and above the desktop surface, and the presence and orientation of other materials, particularly paper documents, in the workspace. The present paper presents a detailed analysis of behavior during active reading tasks to inform the development of current and future systems aimed at softening the boundaries of paper and digital media.

## RELATED WORK

Considerable prior work guides our analysis of body movements and reading behaviors with paper documents. Behavioral models from experimental psychology provide the foundation for analysis. We briefly review related work and then discuss research on *active reading*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ITS'12, November 11-14, Cambridge, Massachusetts, USA.

Copyright 2012 ACM 978-1-4503-1209-7/12/11..\$15.00.

## Descriptive Models of Manual Behavior

*The left hand knows what the right hand is planning, and the right hand knows what the left hand just did.*

—Frank R. Wilson [40, p. 160]

One of the earliest detailed analyses of manual behavior is Napier and Tuttle’s [26] description of prehensile movements, in which an object is held by a gripping or pinching action between the digits and the palm. Along with the classification of power and precision prehensile grips, they also observed that, “during a purposive prehensile action, the posture of the hand bears a constant relationship to the nature of that activity.” In our analysis of active reading, we observe unique prehensile actions that are consistent with this description, especially regarding pen usage.

Guiard [11] investigated the differential roles of dominant and non-dominant hand in a shared task. Of particular interest is his identification of writing (previously conceived of as a unimanual one) as a bimanual task, showing how the non-dominant hand performs a complementary role of continuously repositioning the paper with respect to the motion of the writing hand. Based on his observation of a range of bimanual activities (such as writing, drawing, and sewing), Guiard concluded that “there is a logical division of labor between both hands that appears to govern the entire range of human bimanual activities,” of which movement of the non-dominant hand both precedes and sets the spatial context for the activity of the dominant hand [11]. Many of the behaviors observed in our analysis consist of asymmetric bimanual movements and can be described within this framework.

Both models of manual behavior (see [40] for a thorough review) share a characteristic property: *the movement of the hands often anticipate and signal the nature of activity*. This parallels gesture in face-to-face conversations, in which hand movements or conversational gestures tend to precede semantically-linked speech utterances [6, 17, 23]. Such identifications promise to aid activity recognition, and an important aspect of this involves examining bimanual interaction.

## Bimanual Interaction

Bimanual interaction with a computer was first introduced by Engelbart and English in 1968. Their NLS system supported simultaneous use of a chorded keyboard in the non-dominant hand and a mouse in the dominant hand [9]. However, it was not until the mid-80s’ that Buxton and Myers [7] first detailed the efficacy of two-handed input. They compared unimanual with bimanual input, demonstrating that two-handed input readily outperforms the former. Although independent of Guiard’s work (published a year later), and providing little discussion of the differential roles of each hand, their research increased awareness of the importance of bimanual input mechanisms. Somewhat later Bier et al. [3], took guidance from Guiard’s work to develop the *Toolglass* metaphor—a semi-transparent palette-like interface controlled by simultaneous but functionally asymmetric two-handed input (see also Kabbash et al. [16] for an informative discussion of two-handed input in a compound task). Inspired by the *Toolglass* metaphor and the affordances of physical paper [30], Song

looked at bimanual interaction techniques that render the non-dominant hand as a frame of reference by controlling the visual feedback projected onto paper [32].

Two-handed mode switching techniques have been explored extensively, especially involving combined use of pen and touch interactions. Li et al. [18] conducted an experimental analysis of five different mode switching techniques for pen-based user interfaces and concluded that pressing a button with the non-dominant hand yields the best performance in terms of speed, accuracy, and user preference. This study stimulated a body of research [4, 15, 32] that further confirmed the importance of understanding bimanual interaction and the asymmetric division of labor between hands. Brandl [4] describes several examples of bimanual pen plus multitouch gestures, assigning pen to the dominant hand and multitouch to the non-dominant hand. Hinckley [15] explored a wider vocabulary of pen+touch techniques that considers the interleaving assignment of pen and touch manipulation to either hand, depending on usage context. Based on certain grip postures used by artists, Song et al. [31] explored mode switching techniques that incorporate various postures and multitouch gestures. Inspired by the way artists manipulate a *conté* crayon, Vogel [37] recently investigated mode switching techniques that leverage various contact points of an interactive prism-shaped crayon with a multitouch table.

The reciprocity between visual attention and the cognitive cost of switching between subtasks has also been discussed in terms of Guiard’s theory of bimanual interaction. For example, exploring bimanual vs. unimanual interactions with a “doll’s head” neurosurgical interface, Hinckley [14] demonstrated the naturalness and cognitive benefits of two-handed manipulation. The use of both hands, afforded by the tangibility of the interface, resulted in a low transaction cost for switching between subtasks. Finally, Bederson [2] asserts that the flow [8] experience in human-computer interaction may be subject to interruptions when visual feedback and the associated need for attention becomes a requirement for task execution. This is something that can be minimized by bimanual interaction.

## Active Reading and Tablet Displays

Research efforts from both experimental psychology and human computer interaction have great significance for informing the design of hybrid paper-digital desktop workspaces. One such domain is Active Reading [1], a form of knowledge work that requires high level of interconnectivity among reading-related tasks to retain a “flow experience” for the reader. Such activity is characterized by fluid transitions between immersive reading (the primary task) and a set of subtasks that support the active reading experience. These secondary tasks involve information creation and organization, including annotation [22], content browsing [27], file organization [21], and cross-referencing [28] between multiple documents. Due to the interlacing of interactions involved, active reading serves as an important testbed for exploring interaction techniques that are cognitively less demanding.

Many prior systems have attempted to support active reading by mimicking paper forms of interaction with digital

| Categories      | Description                                 | Duration          | Contact point      | Symmetry   |
|-----------------|---|-------------------|--------------------|------------|
| Reading         | reading with intense concentration          | long              | hand (both)        | symmetric  |
| Visual Tracking | quick skimming for overview of contents     | moderate          | finger, pen, other | N/A        |
| Pointing        | place-marking gesture for quick references  | short             | finger             | N/A        |
| Piling          | chopping motion to stack group(s) of papers | short             | hand (both)        | symmetric  |
| Flipping        | turning to previous or next page            | short to moderate | hand (single/both) | asymmetric |
| Writing         | producing an explicit marking               | moderate to long  | pen                | asymmetric |
| Annotating      | producing telegraphic markings              | short             | pen, highlighter   | asymmetric |

**Table 1.** Coding scheme evolved from iterative analysis of video corpus. Short interactions usually last 1-5 seconds, moderate interactions are longer continuous activities in the range of 5-15 seconds, and long interactions last 15 seconds to several minutes.

technology. For example, XLibris [29] was one of the first efforts that explored pen-based gestural interactions to simulate a paper-like experience on a tablet display. LiquidText [36] explored bimanual multitouch interactions on a tablet PC along with flexible visual representations to recreate experiences that were previously unique to physical paper, such as viewing two disparate areas of text in a single pane. GatherReader [13] describes a bimanual pen+touch system that supports fluid interleaving interactions between reading and writing without requiring explicit specification or organization of content. Finally, Morris et al. [25] investigated active reading in the context of both horizontal and vertical displays. Still, the tangibility and materiality of paper often outweigh the benefits afforded by computers. Simply mimicking key affordances of paper digitally is unlikely to be sufficient. Detailed understanding of how people interact with paper documents is fundamental to the design of more effective digitally-augmented desktop workspaces.

## METHODS

To better understand the context of active reading, we collected 333 minutes of overhead video data of three PhD students (age 25-35, 1 left handed) at their desk during standard paper reading and reviewing activities. Our research goal is to characterize movements of the body and artifacts during the task of active reading to inform the design of a digital environment. After inspecting the whole video corpus for active reading activities, we selected from each participant's data one representative 30 minute segment where active reading was most salient. Our research team conducted an analysis of the larger corpus as well as a detailed microanalysis of the focused video segments. Through the microanalysis we characterize the nuances of human activity during active reading, often by examining moment-to-moment interaction on a millisecond timescale. Our goal is to understand the richness of body and material coordination and identify interactions that may be augmented in a hybrid paper-digital workspace, including indicators of specific active reading behaviors.

### Coding Scheme

Active reading is a broad category description for behaviors that aid in understanding and retaining written content, and encompasses both writing and reading activities. In order to investigate these dimensions two researchers analyzed the entire video corpus and iteratively developed a common coding scheme, informed by Marshall's previous analysis of annotations [22] by selecting the most redundant actions, activities, and behaviors (Table 1). Inter-coder reliability for this coding

scheme and our data set was calculated with Cohen's Kappa coefficient of 92%<sup>1</sup>.

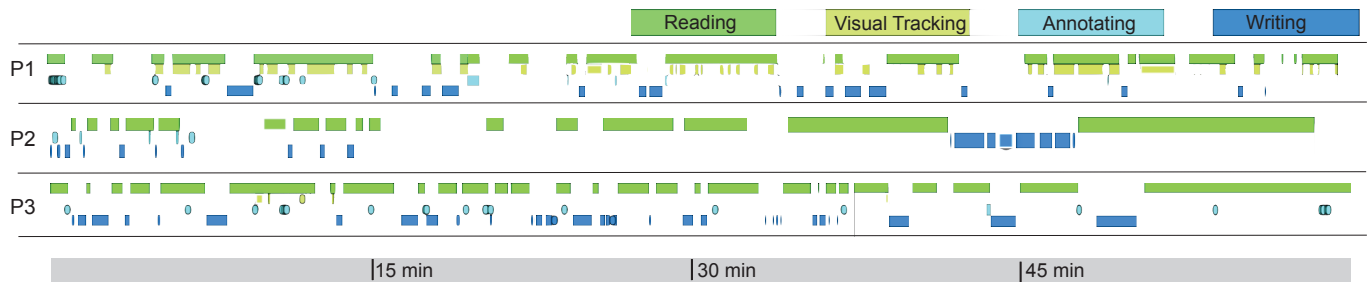
### Video Annotation

Video data was analyzed using ChronoViz [10], a software tool that aids annotation, visualization, and navigation of multimodal time-coded data (see Fig. 1). Each video clip was annotated using the above coding scheme (Table 1), and all annotations were visualized in a single timeline. We also included the acoustic patterns generated in the activities. Thus, adjacent to the annotations, a timeline of an audio waveform was added and incorporated into the analysis. All hand-coded annotations were sorted according to the categories.

<sup>1</sup>Cohen's Kappa coefficient was calculated over a 36 min. session where the two researchers coded a total of 177 events.



**Figure 1.** Analysis of a specific activity in ChronoViz. Multiple categories can be assigned to single events and events of varying duration. Filtering and reordering annotations enables visualization and identification of new patterns of behavior.



**Figure 2.** Distribution of the reading and writing activities for the three participants over one hour of observation. For each participant, four rows of annotations show (top to bottom) reading, visual tracking, annotating and writing. Notice that activities are evenly distributed across the session and that reading and writing are intertwined in a repetitive pattern.

Once sorted, navigation among annotated video segments was feasible within and between categories. Our analysis focused on identifying contextual cues that are consistently present during various active reading activities.

### Microanalysis

Following high-level analysis of the larger video corpus, one 30 minute segment for each of the three participants was analyzed in further detail. As part of this microanalysis, we identified additional dimensions based on body interaction (i.e. single hand vs. bimanual), artifacts handled (i.e. pen, paper, computer), document gestures (i.e. underlining, highlighting, circling), physical gestures (i.e. holding, flipping, tucking), or high-level activity (i.e. searching, browsing).

## RESULTS

The first level of analysis of the video corpus (333 minutes total; P1=160.6 min., P2=71.9 min., P3=100.5 min.) focused on the overall dynamics of reading activity and associated behaviors. In this section we describe the patterns that emerged from this analysis, the temporal distribution of the main reading and writing activities, and the summary statistics of the observed interactions. We then describe the results from the microanalysis.

### Quantitative Analysis

High-level analysis of the video corpus indicates that reading and writing behaviors were fairly evenly distributed across the recorded sessions and among participants. Figure 2 illustrates the distribution of reading and writing for one hour of observation for all three participants. Table 2 illustrates the amount of time each participant spent on writing and reading activities as well as the frequency of actions such as annotating, pointing, piling, flipping and moving. The repetitive pattern of reading/writing was observed across participants. Annotating was a frequent activity and the durations of each contributes to the total time engaged in writing, revealing the importance of general writing interactions interspersed among reading. Flipping was also present across participants, although P3 showed much more flipping activity in conjunction with more writing and annotations. Long reading activities were interrupted by typically shorter periods of writing (between a few seconds and 1.5 minutes). Brief annotations (1-5 seconds) were generally integrated as part of the reading activity, as well as longer visual tracking sessions (2-100 seconds), that were often performed with a finger, a

pen, or a highlighter. Participants varied considerably in use of visual tracking.

### Qualitative Analysis

Building on the high-level analysis of user behavior, we examined the qualitative details of active reading activities. Findings of the analysis are organized into four sections: reading, annotating, navigating, and organizing. These are derived from the coding categories described in the methods section.

#### Reading

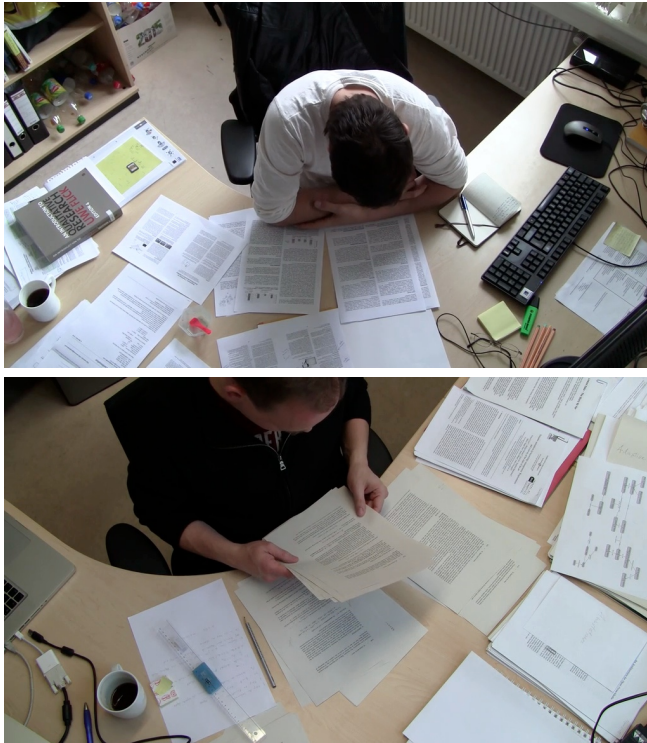
Reading is a rich cognitive activity that requires intense concentration. Although the extent of concentration may vary depending on the goal (e.g., reading for learning, to obtain an overview, find a specific piece of information, etc.), it remains cognitively expensive. Overall, the nature of interaction during reading involved a fixed position and was of longer duration than the rest of the observed activities (see also Table 2).

*Immersive Reading:* We observed two major postural configurations that indicated immersive reading. In general, the readers inclined to maintain a close proximity to the reading materials, which was accomplished in two ways. The first and most prevalent posture was made after bending and moving their heads closer to the desk. For almost all of these instances, the readers' elbows were on the desk either making a cross-arm posture or providing support with hands to rest their heads (Fig. 3). This particular variation of arm posture

| Activity    | Measure            | P1     | P2     | P3     |
|-------------|--------------------|--------|--------|--------|
| Reading     | % of total time    | 55.94% | 50.04% | 68.97% |
|             | Max (sec)          | 408.3  | 223.6  | 505    |
|             | Min (sec)          | 5.9    | 15.5   | 4      |
| Writing     | % of total time    | 7.77%  | 6.03%  | 12.95% |
|             | Max (sec)          | 77.3   | 22.7   | 97.4   |
|             | Min (sec)          | 3      | 6.2    | 1.6    |
| V. Tracking | % of total time    | 30.83% | -      | 2.35%  |
|             | Max (sec)          | 99.8   | -      | 22.1   |
|             | Min (sec)          | 2.1    | -      | 7.7    |
| Annotating  | occurrences / hour | 24.28  | 8.34   | 38.99  |
| Pointing    | occurrences / hour | 4.11   | 8.34   | 3.00   |
| Piling      | occurrences / hour | 5.23   | 1.67   | 2.40   |
| Flipping    | occurrences / hour | 11.21  | 5.00   | 29.99  |

**Table 2.** Duration and frequency of coded active reading behaviors. Over 50% of time is dedicated to reading activities while writing occurs between 6% and 13% of the time. We normalize frequency counts to the average number of occurrences per hour.



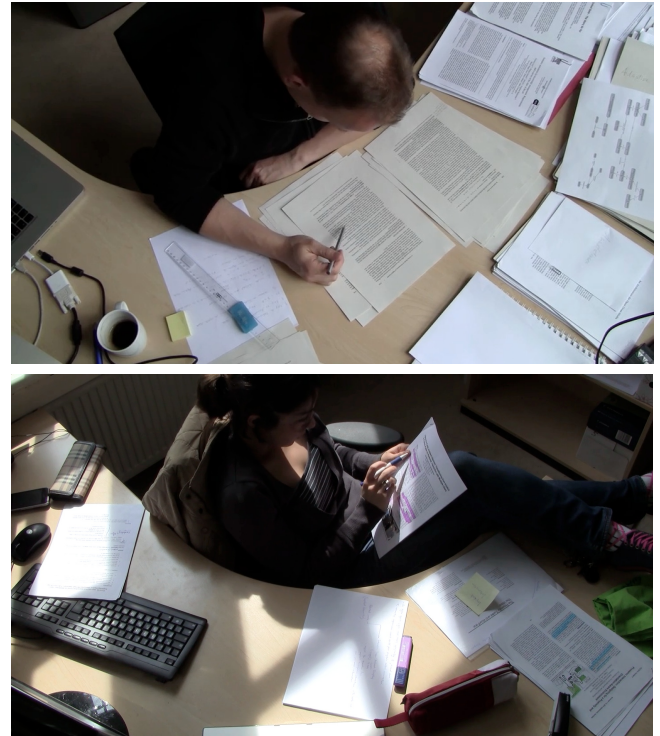


**Figure 3.** Examples of immersive reading. Leaning over with elbows bent on the physical desktop (*top*) and papers brought closer to reader's head (*bottom*)

may counter muscle strain coming from prolonged periods of neck and head flexion [34]. In these configurations, we also observed subtle continuous movements of the hands, such as the persistent fidgeting of fingers or twirling of a pen. The second posture was made while holding the reading materials with both hands in mid-air, and drawing them closer to the reader's head. This posture was also maintained when one participant switched to a leaned-back position by placing both legs on the desk. Detecting this sustained orientation of body and materials through an overhead camera or sensing device (e.g., depth camera) promises to be an effective index of immersive reading. Also, we noticed the overwhelming presence of symmetry in the readers' use of arms and hands, yet another strong contextual cue and identifying feature.

*Visual Tracking:* In the midst of each reading session, readers (especially P1) exhibited visual tracking activities, involving the dominant hand and an object of thin profile (i.e., a pen) or a finger to serve as a transient focal point of visual attention. A typical tracking activity took the form of a hovering gesture, in which the tip of an object was seen as hovering (less than half a centimeter) above parts of the text in a linear fashion. While using a marking tool as the main mode to aid linear reading of the text, such as a pen or pencil (Fig. 4), all participants exhibited a high tendency to switch between visual tracking and annotating or writing activities, leveraging the same tool for different activities.

It was sometimes difficult to determine whether the reader was hovering above or making slight marks on the paper, until we noticed a fundamental visual cue that served to signify



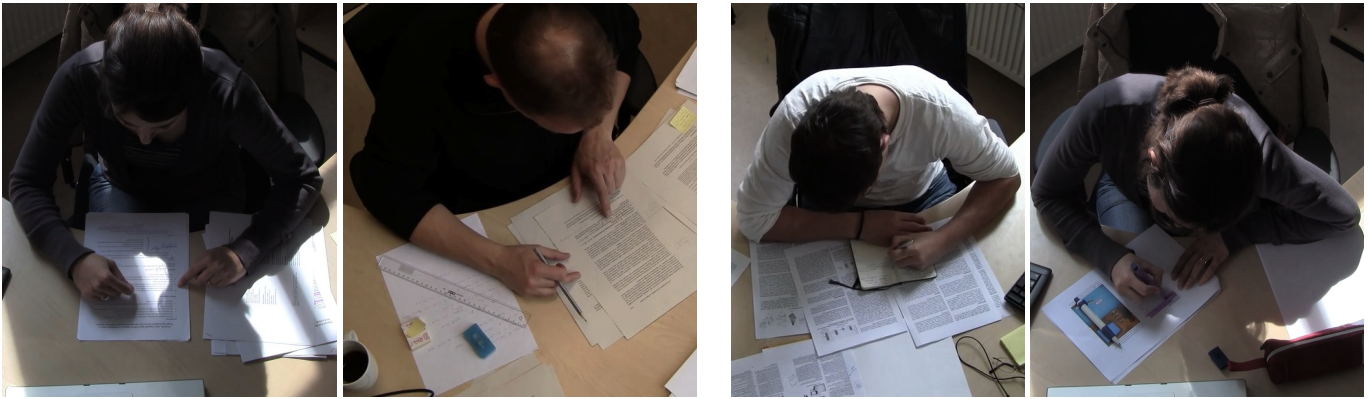
**Figure 4.** Visual tracking with a pen or pencil. The instrument is held loosely towards the top, providing precise index with minimal occlusion. People shift easily from this grip to a closer grip for marking the paper.

visual tracking. For most tracking activities, the grip posture was relaxed and the tip of the object protruded such that it was visible at any orientation and the angle was less acute than when marking. We also observed that the orientation of the pen tip changed relative to the area of text that was being read. Detecting this loose grip posture coupled with the angle and movement over the page will help identify visual tracking.

Visual tracking was also often accompanied by an indexical gesture, performed with the non-dominant hand (Fig. 5, left). While reading, participants placed their index finger on the margins to the left of a particular line or paragraph of interest, and moved up or down in relation to the area of text being read. This is an example of using the non-dominant hand as a frame of reference for guiding visual attention and dominant hand activity, functioning as a place marker. Attending to the placement of the non-dominant hand provides important cues for recognizing activities and focus of attention.

#### *Annotating*

Marshall's extensive study [22] of annotated materials reveals that annotations provide a permanent frame of reference for guiding the reader's visual attention. It is also known to aid encoding activities and memory. According to Marshall, the major functional roles of annotations include procedural signaling for future attention, place marking and aiding memory, in situ locations for problem solving, and visible traces of reader's attention. We observed annotating practices in our video corpus that fit these descriptions. In the remainder of this section we outline explicit (short notes) and telegraphic annotations (underlining, highlighting, circling).



**Figure 5.** Reader uses the non-dominant hand in coordination with the dominant hand to focus attention on segments of the document (*left images*), while they use the non-dominant hand to control the position and orientation of paper while writing (*right images*).

*Short Notes:* In our data short notes were written on a notebook, margin of a paper, or a Post-It. Marginal and Post-It notes were written in relation to the spatial location of a specific paragraph, line, or word of interest, rendering manual search and navigation less difficult. This was apparent, for example, when one participant (P1) quickly identified and selected a document with a Post-It note from a large pile of similar looking documents.

*Underlining, highlighting, and circling:* Other forms of annotations, including underlining, highlighting and circling (telegraphic marking), also increase the saliency of particular areas of text. Underlining activities were often preceded, or followed by visual tracking. Unlike tracking, annotating activities involved the use of a tripod grip, which is a characteristically rigid and tense grip posture. Calculating pen-to-surface angle in addition to pen grip posture may help a system discriminate between tracking and annotating activities. A common pen-to-surface angle was a consistent signature of annotation instances.

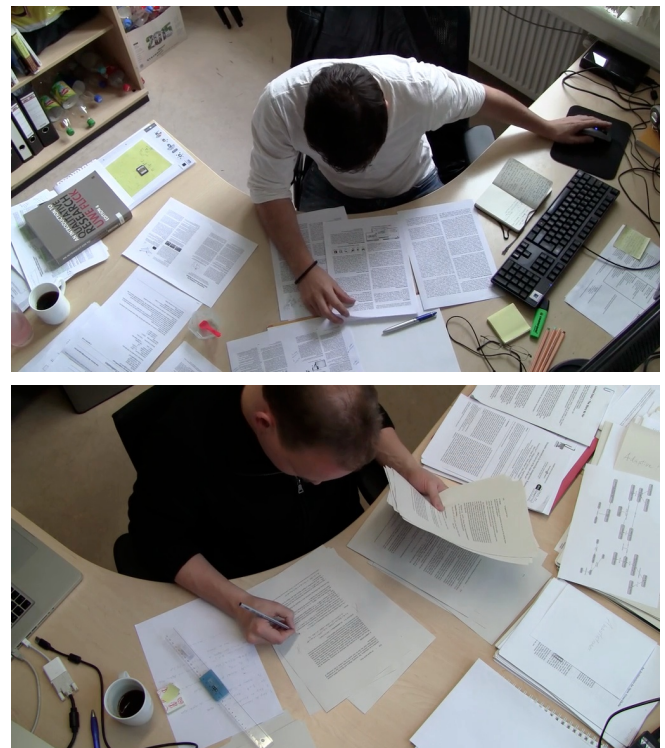
*Dominant vs. Non-dominant hand:* We also observed a fundamental difference in the relationship between the dominant and non-dominant hands during annotating activities. According to Guiard’s description of asymmetric bimanual interaction [11], the dominant hand moves relative to the motion of the non-dominant hand, which is an instance of setting a frame of reference. Consistent with Guiard’s observation, we found all annotating activities to be carried out by an asymmetric division of labor between the hands, in which the non-dominant hand served as an anchor point by holding down the paper while the dominant hand performed the functional role of the annotating activity (Fig. 5, right). This interaction was also preserved when one participant wrote a brief note on a small Post-it. Yet again we note that detecting the location and posture of the non-dominant hand provides important cues about readers’ attention and frame of reference.

#### Navigation

We observed multiple instances of paper navigation behaviors that were typical of content browsing, searching, and cross-referencing activities within and between source documents (Fig. 6). These behaviors were typically observed in the beginning of reading activities as well as in transitional phases in which the reader switched between different

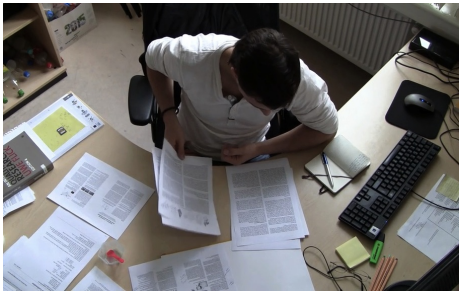
reading materials. They include holding, flipping, and tucking (Fig. 7), which all involved a coordinated asymmetric division of labor between the dominant and non-dominant hands. Many of these activities are afforded by the tangibility of paper and its physical properties of being thin, lightweight, and flexible [30]. We see many opportunities in tracking the position of paper documents with respect to one’s hands as a way of indexing attention and activity.

*Holding:* A holding gesture was defined as any instance of holding or lifting an object (i.e. paper) from the surface of the desk. This gesture marked the beginning of most paper navigation activities. Almost all navigational activities were

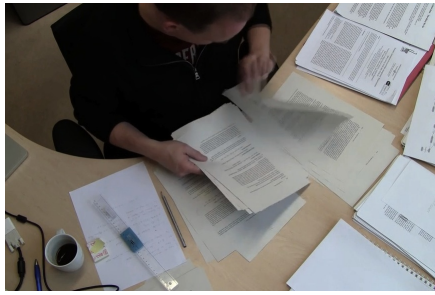


**Figure 6.** Readers hold documents or objects while cross-referencing between multiple source documents. Preserving the non-dominant hand’s position on a mouse (*top*) or holding a group of papers (*bottom*) may indicate intention to return to a previous state or frame of reference.





(a) Holding



(b) Flipping



(c) Tucking

Figure 7. Examples of paper navigation.

preceded by the holding gesture, since any paper, or stacks of papers, had to be picked and held in mid-air before it could be acted upon. For example, page turning was most likely preceded by a holding gesture, in which an edge of a page was picked up and held in advance before the reader turned the page. Importantly, its duration was short-lived, except a few occasions in which a paper was held in “stand-by” mode during an ongoing activity.

However, in a few cases, a paper was moved or manipulated without being lifted from the surface of the desk. One example was a dragging gesture, in which a paper, once read, was dragged by the participant (P3) across the desk and away from the center space. Minor positional adjustments were made via panning and rotating gestures.

Long term usage of the hold gesture was observed during cross-referencing or content browsing activities. We observed one such behavior (P2) that seemed to be characteristically slow and fine-grained, indicating a decision making process. This behavior was observed when readers have narrowed down their search to a smaller number of items. During this process, one or more papers were held in mid-air for close inspection, and readers looked back and forth between the papers on the desk and the ones they were holding.

*Flipping:* A flipping gesture was used when a reader turned over a page in order to navigate between the previous or next page. As mentioned earlier, a typical page flip started with a holding gesture, or a pinch grip of the edge of a paper, followed by two different variations of the flipping gesture depending on the state of the document. When the papers were stapled together, readers typically exhibited a bimanual asymmetric interaction. In general, the dominant hand picked up the tip of the paper (hold gesture), passing it over to the non-dominant hand to turn the page.

In contrast, a single paper sheet was flipped single-handedly, typically using the non-dominant hand. In both cases, the non-dominant hand was used to perform the flipping gesture, which was typically due to the fact that the dominant hand was occupied with a marking tool for performing annotating activities. However, there was another difference in the execution of the flipping gestures. With stapled papers, readers turned the page by rotating the entire forearm. With unstapled pages, the same gesture was performed by only rotating the wrist. This difference seemed to emerge as a result of the use of staples, which limit the paper’s range of motion.

*Tucking:* Another consistent use of the holding gesture involved tucking. This gesture seemed to cue within-document cross referencing or search activities, in which readers were often seen as making quick comparisons between multiple pages. Instead of making a full page turn, readers searching for a particular page used one hand to lift the edges of a paper while using the other to hold onto multiple pages in “standby”. During this phase, one reader (P1) used her “holding hand” as a place marker by inserting one or two fingers between several prematurely turned pages. This action seemed to facilitate within-document search or referencing processes by allowing the reader to easily return to one of the target search pages while accessing those pages in full view.

#### Organization

After an extensive study of desktop organization behaviors, Malone [21] concluded that documents are organized into piles and files. We observed various multimodal cues that appear to signify piling activities, including the stacking motion, spatial layout of document piles, acoustic pulse patterns, and searching behaviors.

*Stacking:* A stacking gesture was always followed by a formation of a paper pile. It may be followed by a stapling activity that binds multiple single pages into one group. The stacking gesture was consistently identified as a symmetric bimanual interaction, and all readers performed the gesture in three simple steps: first, multiple papers were gathered together and held in mid-air with both hands in a vertical fashion; second, while maintaining the vertical hold gesture, papers were let loose from the fingers to slide down until they touched a flat surface; and finally, further adjustments were made with both hands to ensure alignment in the vertical and horizontal directions. This gesture leverages gravity and the flatness of the desk surface to yield a near-perfect alignment of a group of papers.

*Spatial layout:* While the most actively read documents are placed in the immediate vicinity of the reader, other documents and piles were arranged to the periphery. In addition, a dense region of piles and documents often formed towards the non-dominant side of the reader (see Fig. 6). This particular arrangement seemed to be influenced by the placement of the reader’s personal computer (both laptop and desktop), positioned so as to be able to use the mouse with their dominant hand. Since piles are by no means arranged in systematic order, recognizing the spatial location and activity history of

each pile will provide additional cues to the organization of physical documents in space.

*Search:* Because of the nature of piling activities, with papers stacked on top of each other, finding a particular document or page can be a serial and time consuming process. When looking for a particular document from a pile of papers, participants typically performed the search by removing one paper at a time, beginning from the top of the pile, until they found the paper of interest. Interestingly for the purpose of recognizing this activity, we noticed that a secondary pile was created by virtue of the activity, later to be combined with the previous pile once the reader completed the search. Participants also demonstrated well-coordinated asymmetric bimanual interaction. Although we have not observed any particular role bias between the dominant and non-dominant hands, there was clearly an asymmetric division of labor between hands in which one hand was in charge of lifting while the other did the moving.

*Audio waveform:* We observed an interesting phenomenon while analyzing acoustic pulse patterns generated from the above activities. To our surprise, most peaks in the audio waveform were associated with paper manipulation activities (see Fig. 1 for an example of stacking behavior aligned with a peak in the audio signal). We now suspect that there may be unique ranges of frequency amplitudes associated with various paper activities. Detecting and processing such audio signals may provide additional cues and signatures of the structure and organization of activity on the desktop.

## IMPLICATIONS FOR DESIGN

When, where, and how digital technology should be used to augment active reading are central questions of our research. We argue that detailed understandings of the structure and patterns of active reading activities are needed in order to be able to augment them in effective and natural ways. The contribution of the current work is a microanalysis of active reading, focusing primarily on identifying patterns and practices that might serve as activity and context signatures. By applying the outcomes of our analysis as part of a system able to recognize and track body movement and activity on and above the desk, we aim to develop feedback mechanisms that support novel augmented active reading experiences. While the integration of those findings in the system we are developing is part of our future work, in this section we detail the implications of active reading activity recognition, and we sketch four possible interactions that can be implemented by leveraging the recognized activity.

### Recognizing Activities

We identified various types of activities involved in active reading. One central task involves immersive reading. It requires intense concentration and may be susceptible to the slightest of interruptions. Annotation is another central task commonly associated with immersive reading, serving to assist subsequent activity. We identified several peripheral tasks that provide support: visual tracking, underlining and highlighting, cross-referencing between multiple documents, content browsing, and document organization. Of particular

importance for design, we identified how the position and movement of the body and hands with respect to the material workspace might serve as signatures for these activities.

*Sustained Body Positioning:* A salient cue for automatically identifying immersive reading is the presence of a sustained body position (leaning over desk, elbows on desk, or feet propped up). Another feature of this activity is the proximal distance of a document to the reader's head. From a design perspective being aware of this state is particularly important since it is particularly vulnerable to interruptions.

*Hand Positioning and Contact Points:* The position and document contact points of the hands provide considerable information about the focus of attention. Most activities center around gestures that involve hands in contact with paper documents. Knowing which hand is in contact with which document is critical, but additional resolution at the paragraph or word level is also necessary. Sustained contact between hands and documents may indicate place-marking and holding, both of which provide a spatial frame of reference that guides attention to facilitate peripheral activities (e.g. visual tracking, cross-referencing, searching). Furthermore, visual tracking and cross-referencing are accompanied by a place-marking gesture (Fig. 5, left) while cross-referencing and searching are accompanied by a place-holding gesture (Fig. 6).

*Positioning of Non-Dominant Hand:* Detecting the location and document contact points for the non-dominant hand working in coordination with the dominant hand provides a different view into the person's frame of reference for their current activity. The non-dominant hand serves as a pointer, index, and guide to direct attention, possibly at a higher level or with different temporal qualities than the dominant hand. Using the non-dominant hand to temporarily hold a group of documents or other object may preserve a past frame of reference to which the person intends to return.

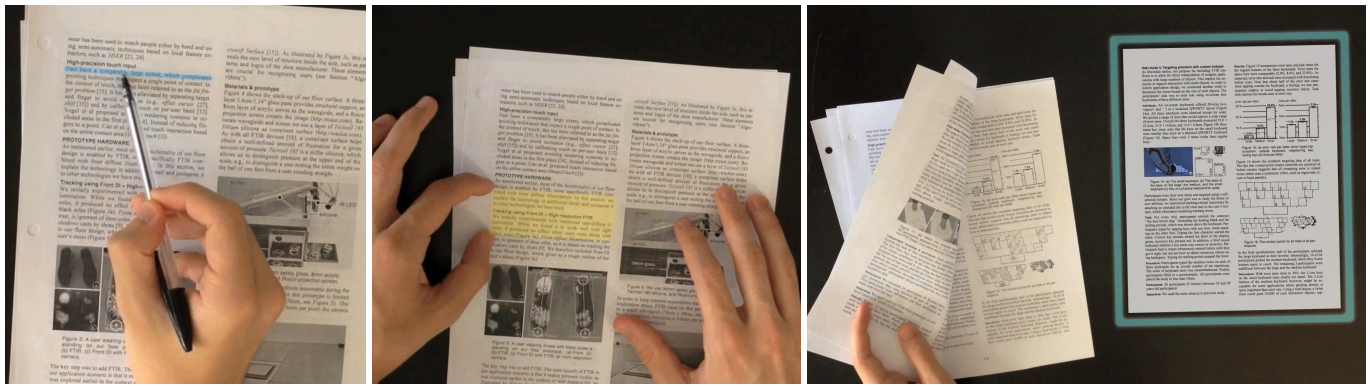
*Pen Grip Posture:* While both visual tracking and annotating activities frequently involve the use of pens, the associated postures differ (i.e. relaxed vs. tripod grip). The angle of the pen-to-surface may also differ during visual tracking and annotating. These different configurations may be used to identify when participants switch between writing and tracking, subsequently providing different user interface facilities depending on the activity.

*Acoustic Properties:* The stacking gesture used in piling activities has a unique acoustic property. Other paper navigation techniques such as flipping pages and searching through a pile of documents also seems to generate unique frequency amplitudes. In combination with tracking document movement and placement via an overhead camera, processing these acoustic signals may provide a robust method of helping to detect grouping or navigation actions without additional user input. More work is needed to explore the specific acoustic properties of each interaction.

### Envisioned Interactions

The microanalysis of active reading behavior and identification of signatures of component activities provides a basis for





**Figure 8. Envisioned interactions. Left:** Visual tracking with a pen or finger temporarily highlights text on the paper document. **Center:** Pinch-to-select gesture persistently highlights blocks of text. **Right:** Tucking a page provides an auxiliary view on the digital surface.

the design of a range of novel interaction techniques to augment interactive desktop workspaces. We have prototyped several interactions based on our microanalysis of active reading (see Fig. 8). Note that while we plan to implement these interactions as part of the desktop workspace we are currently developing, the figures are mockups that give concrete design ideas to demonstrate how our results may inform augmented paper-based interactions.

**Visual Tracking:** To provide a supplementary visual cue to guide attention while engaged in active reading, the tip of the pen or the finger used to track the information on the paper document can be augmented with a focused high-intensity light cast on the entire line of text. This can function as a transient focal point of attention, with the highlight fading away as the reader moves his pen or finger away from the line to another line, and the focused light following (see Fig. 8, left).

**Pinch-to-Select:** To create an artificial highlight of the reader's visual attention while they work elsewhere, readers can use the index and thumb finger, similar to the "pinch-to-zoom" gesture on mobile phones, with their non-dominant hand next to a portion of text they want to highlight or select. The selection results in a persistent highlighting that can remain on the page while reading the document (see Fig. 8, center). Selected text can be used also in other interactions with the system, such as copying and transferring.

**Drag to Copy:** The reader can drag selected text or an image from the paper document onto the interactive workspace. Flicking an image with one finger can send the image to the nearest display monitor in the direction of the flicking motion.

**Parallel View:** To support within-document cross-referencing, a duplicate page is projected onto the desktop when readers flip a page that is tucked in between their fingers more than once. This enables reading two pages in full view regardless of the source location of each page (see Fig. 8, right).

## CONCLUSION

This paper describes the process of active reading in detail as a foundation for designers of augmented desktop workspaces. Our research team is iteratively developing a novel augmented desktop workspace that uses a depth camera to track

movement of the body and hands on and above the desktop surface, captures video of activity, and can project onto the desk surface. The main contributions of the work reported here are a detailed microanalysis of active reading, identification of potential signatures of component activities, and concrete examples of how these insights should drive interaction design. We stress the importance and significance of asymmetric bimanual interaction and the need for digital augmentations that fit naturally with everyday active reading practices and support fluid interaction with both paper and digital media. The broader impact of our analysis is to inform the development of novel interaction techniques based on the recognition of specific features of active reading components, opening up new avenues for exploring augmentation of active reading on and above the desktop.

## REFERENCES

1. A. Adler, A. Gujar, B. L. Harrison, K. O'Hara, and A. Sellen. A diary study of work-related reading: design implications for digital reading devices. In *Proc. CHI '98*, pages 241–248, 1998.
2. B. B. Bederson. Interfaces for staying in the flow. *Ubiquity*, pages 1–1, Sept. 2004.
3. E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. Toolglass and magic lenses: the see-through interface. In *Proc. SIGGRAPH '93*, pages 73–80, 1993.
4. P. Brandl, C. Forlines, D. Wigdor, M. Haller, and C. Shen. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *Proc. AVI '08*, pages 154–161, 2008.
5. V. Bush. As we may think. *The Atlantic Monthly*, 176(1):101–108, 1945.
6. B. Butterworth and G. Beattie. Gesture and silence as indicators of planning in speech. *Recent advances in the psychology of language: Formal and experimental approaches*, 4:247–360, 1978.
7. W. Buxton and B. Myers. A study in two-handed input. *SIGCHI Bull.*, 17(4):321–326, Apr. 1986.
8. M. Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*, volume 8:1. Harper Perennial, 1991.

9. D. Engelbart and W. English. A research center for augmenting human intellect. In *Proc. AFIPS '68*, pages 395–410, 1968.
10. A. Fouse, N. Weibel, E. Hutchins, and J. D. Hollan. Chronoviz: a system for supporting navigation of time-coded data. In *Extended Abstracts of CHI 2011*, pages 299–304, 2011.
11. Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model, 1987.
12. F. Guimbretière. Paper Augmented Digital Documents. In *Proc. UIST '03*, 2003.
13. K. Hinckley, X. Bi, M. Pahud, and B. Buxton. Informal information gathering techniques for active reading. In *Proc. CHI '12*, pages 1893–1896, 2012.
14. K. Hinckley, R. Pausch, and D. Proffitt. Attention and visual feedback: the bimanual frame of reference. In *Proc. I3D '97*, pages 121–ff., 1997.
15. K. Hinckley, K. Yatani, M. Pahud, N. Coddington, J. Rodenhouse, A. Wilson, H. Benko, and B. Buxton. Pen + touch = new tools. In *Proc. UIST '10*, pages 27–36, 2010.
16. P. Kabbash, W. Buxton, and A. Sellen. Two-handed input in a compound task. In *Proc. CHI '94*, pages 417–423, 1994.
17. A. Kendon. Gesture. *Annual Review of Anthropology*, 26:pp. 109–128, 1997.
18. Y. Li, K. Hinckley, Z. Guan, and J. A. Landay. Experimental analysis of mode switching techniques in pen-based user interfaces. In *Proc. CHI '05*, pages 461–470, 2005.
19. C. Liao, F. Guimbretière, K. Hinckley, and J. Hollan. PapierCraft: A Gesture-Based Command System for Interactive Paper. *ACM TOCHI*, pages 1–27, 2008.
20. W. MacKay, G. Velay, K. Carter, C. Ma, and D. Pagani. Augmenting reality: Adding computational dimensions to paper. *Communications of the ACM*, 36(7):96–97, 1993.
21. T. W. Malone. How do people organize their desks?: Implications for the design of office information systems. *ACM Trans. Inf. Syst.*, 1(1):99–112, Jan. 1983.
22. C. C. Marshall. Annotation: from paper books to the digital library. In *Proc. DL '97*, pages 131–140, 1997.
23. D. McNeill. So you think gestures are nonverbal? *Psychological Review*, 92(3):350–371, 1985.
24. T. Moran and S. Zhai. Beyond the desktop metaphor in seven dimensions. *Designing integrated digital work environments*, pages 335–354, 2007.
25. M. Morris, A. Brush, and B. Meyers. Reading revisited: Evaluating the usability of digital display surfaces for active reading tasks. In *Proc. TABLETOP '07*, pages 79–86, 2007.
26. J. Napier and R. Tuttle. *Hands*. Princeton Science Library. Princeton University Press, 1993.
27. K. O'Hara and A. Sellen. A comparison of reading paper and on-line documents. In *Proc. CHI '97*, pages 335–342, 1997.
28. K. O'Hara, A. S. Taylor, W. M. Newman, and A. Sellen. Understanding the materiality of writing from multiple sources. *Int. J. Hum.-Comput. Stud.*, 56(3):269–305, 2002.
29. B. N. Schilit, G. Golovchinsky, and M. N. Price. Beyond paper: supporting active reading with free form digital ink annotations. In *Proc. CHI '98*, 1998.
30. A. J. Sellen and R. H. Harper. *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA, 2003.
31. H. Song, H. Benko, F. Guimbretière, S. Izadi, X. Cao, and K. Hinckley. Grips and gestures on a multi-touch pen. In *Proc. CHI '11*, pages 1323–1332, 2011.
32. H. Song, F. Guimbretière, T. Grossman, and G. Fitzmaurice. Mouselight: bimanual interactions on digital paper using a pen and a spatially-aware mobile projector. In *Proc. CHI '10*, pages 2451–2460, 2010.
33. J. Steimle, O. Brdiczka, and M. Mühlhäuser. CoScribe: Integrating Paper and Digital Documents for Collaborative Knowledge Work. *IEEE Trans. on Learning Technologies*, 2009.
34. L. Straker, C. Pollock, R. Burgess-Limerick, R. Skoss, and J. Coleman. The impact of computer display height and desk design on muscle activity during information technology work by young adults. *Journal of Electromyography and Kinesiology*, 18(4):606 – 617, 2008.
35. A. Tabard, W. Mackay, and E. Eastmond. From individual to collaborative: the evolution of prism, a hybrid laboratory notebook. In *Proc. CSCW '08*, 2008.
36. C. S. Tashman and W. K. Edwards. Liquidtext: a flexible, multitouch environment to support active reading. In *Proc. CHI '11*, pages 3285–3294, 2011.
37. D. Vogel and G. Casiez. Conté: multimodal input inspired by an artist's crayon. In *Proc. UIST '11*, pages 357–366, 2011.
38. N. Weibel, A. Ispas, B. Signer, and M. C. Norrie. PaperProof: A Paper-Digital Proof-Editing System. In *Proc. CHI '08*, 2008.
39. P. Wellner. Interacting with paper on the digitaldesk. *Communications of the ACM*, 36(7):87 – 96, 1993.
40. F. Wilson. *The Hand: How Its Use Shapes the Brain, Language, and Human Culture*. Vintage Series. Vintage Books, 1999.
41. R. Yeh, J. Brandt, J. Boli, and S. Klemmer. Interactive Gigapixel Prints: Large, Paper-based Interfaces for Visual Context and Collaboration. In *Proc. Ubicomp 06*, 2006.