Pen + Touch = New Tools

Ken Hinckley, Koji Yatani, Michel Pahud, Nicole Coddington,
Jenny Rodenhouse, Andy Wilson, Hrvoje Benko, and Bill Buxton
Microsoft Research, One Microsoft Way, Redmond, WA 98052
{kenh, mpahud, nicolec, jrodenh, awilson, benko, bibuxton}@microsoft.com; yatani@gmail.com

ABSTRACT
We describe techniques for direct pen+touch input. We observe people's manual behaviors with physical paper and notebooks. These serve as the foundation for a prototype Microsoft Surface application, centered on note-taking and scrapbooking of materials. Based on our explorations we advocate a division of labor between pen and touch: *the pen writes, touch manipulates, and the combination of pen + touch yields new tools*. This articulates how our system interprets unimodal pen, unimodal touch, and multimodal pen+touch inputs, respectively. For example, the user can hold a photo and drag off with the pen to create and place a copy; hold a photo and cross it in a freeform path with the pen to slice it in two; or hold selected photos and tap one with the pen to staple them all together. Touch thus unifies *object selection* with *mode switching* of the pen, while the muscular tension of holding touch serves as the "glue" that phrases together all the inputs into a unitary multimodal gesture. This helps the UI designer to avoid encumbrances such as physical buttons, persistent modes, or widgets that detract from the user's focus on the workspace.

Author Keywords: Pen, touch, gestures, tabletop, tablets, bimanual, systems

ACM Classification Keywords: H.5.2 Information Interfaces and Presentation: Input

General Terms: Human Factors

INTRODUCTION

Direct interaction with displays is rapidly becoming one of the primary means by which people experience computing. This has led to renewed interest in devices with multi-touch input, as well as pen input, manifest in form factors ranging from hand-helds, slates, booklets, desktops, table-tops, and wall displays. The iPhone, iPad, Project Courier [29] (see also [11,20]), Wacom Cintiq, Microsoft Surface, and Smartboard are, respectively, examples of each. But only rarely do systems support both modalities, much less the combination of the two for simultaneous pen + touch interaction [6,38,39]. Meanwhile many efforts continue to consider either touch input in isolation, or pen input in isolation. Yet if dual-mode digitizers with both pen and touch represent the future of displays [15,21], the presence of both modalities should significantly alter our perspective on how to most effectively design for multi-touch, as well as the pen. Hence *simultaneous pen+touch* is a nascent topic that demands further study and exploration.

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, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
UIST'10, October 3–6, 2010, New York, New York, USA.
Copyright 2010 ACM 978-1-4503-0271-5/10/10....$10.00.

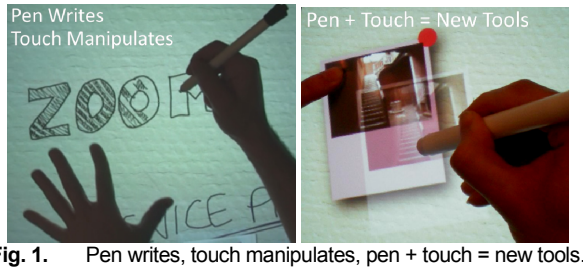

Fig. 1. Pen writes, touch manipulates, pen + touch = new tools.

This argues for a holistic systems-oriented approach, rather than focused evaluation of one-off techniques [17]. Our experience is that trying many ideas [8]— some good, some perhaps bad in retrospect, and some intentionally chosen to highlight conflicting conventions or thorny design decisions rather than hide them— is an excellent way to draw out nuances and gain insights into novel input modalities (e.g. [22,25,37]). Much of the value of pen+touch interaction seems to accrue from the workflow— that is, the flow of interaction across a series of interrelated subtasks— rather than from individual techniques considered in isolation. Exploring a rich application context offers a realistic perspective of how pen+touch influences UI design issues and trade-offs. We inform our design with an observational study of 8 people, and discuss feedback from 11 test users, but do not present a detailed evaluation in this manuscript.

We discuss *Manual Deskterity*, a "scrapbook" application [20,23,28] for the Microsoft Surface which serves as a virtual designer's studio: that is, part digital desk and part digital moleskin notebook [8,19]. While a few other efforts do consider pen and touch (e.g. [6,38,39]), we contribute an analysis that significantly advances the topic by identifying nine key design considerations of pen+touch. We also contribute observations of people's use of physical paper and notebooks: there is a rich set of deeply rooted behaviors that people exhibit when working with pen, paper, clippings, pages, and books, with both stylus and bare-handed manipulation, that differentiate between the roles of pen and touch. We leverage this to define the primitives of our interaction language, and to design pen+touch gestures that go well beyond pens and paper to support novel tools.

From an input-theoretic standpoint a key insight of this work is that for multimodal pen+touch interactions, a single direct-touch action (holding an object) integrates three steps: (1) object selection; (2) mode switching [27,31] of the pen from ink strokes to contextual commands that act in reference to the selected object; and (3) the phrasing together of multiple pen

APPLE 1025
Apple v. Smith Interface
IPR2024-01083

and touch inputs into a unitary command [9]. Furthermore the muscular tension of holding an object serves as salient kinesthetic feedback [34] that unambiguously ties all of these things together. This is achieved without recourse to physical buttons, persistent modes, or permanently visible widgets that compete with the workspace for screen real estate and the user's attention.

RELATED WORK

Systems with direct input often employ only one of pen (e.g. [2,25,33]) or touch (e.g. [14,36,37]). Yet an earlier generation of devices, such as the Palm Pilot (1996), supported use of either pen or touch. However, the technology could not differentiate pen vs. finger contact. But a new generation of digitizers is emerging that can sense multi-touch while distinguishing pen from touch, thus enabling pen AND touch rather than pen OR touch [15].

Several research efforts explore the combination of pen and touch. Yee [39] uses single-touch + pen input to support panning a canvas and drawing. Wu [38] describes two examples of pen and touch gestures. Brandl [6] explores bimanual pen + multi-touch techniques that assign the pen to the preferred hand and touch to the nonpreferred hand. We instead tease apart *pen vs. touch, preferred vs. nonpreferred hand assignment,* and *unimanual vs. bimanual interaction*, among other factors. For example, we consider unimanual usage contexts where the user interleaves pen and touch interactions with the preferred hand, and we explore a wider vocabulary of novel pen + touch gestures that afford compound transactions.

Cohen discusses the complementary role of pen gestures and natural language [12]; he treats multimodal input with a probabilistic approach [13]. We instead treat pen+touch in a manner that affords deterministic state-machine-driven GUI's. Also, because pen and touch are both manual input modalities, the nuances of how the two complement one another are more subtle, and we must overcome a legacy of designs that have treated pen or touch interchangeably.

Guiard observes that the hands cooperate [18], so the question is not *Which hand is better?* but rather "What is the logic of the division of labor between the hands?" Likewise, in our research we ask: *What is the logic of the division of labor between pen and touch in UI design?* Our gestures also build on Guiard's observation that the nonpreferred hand frames the action of the preferred hand.

T3 explores indirect bimanual input with a multi-button puck and stylus input [26], including airbrushing along a french curve held in the nonpreferred hand. We explore analogous pen + touch techniques to draw along a straightedge, but our techniques employ entirely button-free direct input. Studies of two-handed input have emphasized how bimanual interactions afford compound tasks, such as navigation-selection tasks [10,26]. The design approach of *chunking and phrasing* [9] proposes aggregation of atomic tasks into higher level "cognitive chunks." We show how pen + touch offers a new arsenal of techniques to support such compound tasks, without recourse to buttons [27].

Frisch describes a set of user-elicited gestures where users often treat pen and touch interchangeably [16]. As a result, the gestures contain many ambiguities. Which pen, touch, or pen+touch gestures should a system support, and why? User-defined gestures yield insights but cannot be taken too literally: people have difficulty envisioning how to use new modalities of which they have little or no experience [17]. Indeed, our exploration of *Manual Deskterity* convinces us that if each input modality offers complete coverage of all interactions, it quickly robs the combination of pen and touch of much of its vigor. Differentiating between pen and touch, rather than treating them interchangeably, offers a consistent and rich *designed* input vocabulary. Nonetheless we build our gestures and techniques on a vocabulary of naturally occurring manual behaviors, as shown below.

DESIGN STUDY USING A PAPER NOTEBOOK

We conducted a design study with a paper notebook to gain insight as to how people work with physical tools and pieces of paper. We asked each participant to illustrate their ideas for a hypothetical short film by pasting and annotating clippings in a paper notebook. To simulate a slate computer where the user could move between pages, we provided a paper notebook (7" x 10" when opened) as the authoring space. Users sat at a table stocked with tools such as pens, glue, tape, and scissors, as well as print-outs of inspirational materials (20 visually salient web pages).

Eight people (6 male, 2 female, age 9–53 years, all right-handed) participated in the study. We taped the sessions and looked for patterns in how users worked, gestured, held objects, or structured their working space. From this exercise we observed behaviors (B0-B9) that informed specific gestures as well as design attributes of our system.

B0. Specific Roles. Participants wrote with the pen, and arranged clippings with their fingers, often using both hands (Fig. 2a, 2b). Users did not move objects with the pen. While seemingly obvious, we call these behaviors out here because they conflict with the one-handed, unimodal methods of interaction that typically encumber users on devices such as Tablet PC's.

B1. Tuck the Pen. People tucked the pen between their fingers to interleave writing and moving clippings around with their preferred hand (Fig. 2a), or flipping pages (Fig. 2f). Participants were remarkably adept at rapidly interleaving the pen and fingers in this manner.

B2. Hold Clippings. People temporarily held clippings in place with a finger of their nonpreferred hand (Fig. 2a).

B3. Hold while Writing. Participants tended to hold a clipping with their nonpreferred hand while writing about it with the pen. This was often the case even if the clipping was glued into place, or if the annotations were to the side of the clipping (Fig. 2b).

B4. Framing. A common hand posture was to hold thumb and index finger to "frame" clippings (Fig. 2b). This appeared to be a behavior to *focus attention on a source object and reference annotations to it*, rather than one strictly necessary to hold it in place.

B5. Scraps. Participants cut clippings from the inspirational materials while holding a sheet in their nonpreferred hand, above the notebook. Users at times adopted the resulting scraps when they fell onto the page (Fig. 2c).

B6. Extended Workspace. Participants arranged the workspace with the notebook proximal to their body, while reaching above it to access tools and materials. A surround of clippings, unused pages, and tools formed an extended workspace beyond the notebook. (Fig. 2d).

B7. Piling. All users formed piles of "interesting" clippings while holding other items in the nonpreferred hand.

B8. Drawing along Edges. While not a common behavior, a couple of people did employ clippings as a constraint for the pen, to draw a border around an item (Fig. 2e).

B9. Hold Page while Flipping. Participants kept their place in the notebook by holding a thumb or fingers of the nonpreferred hand on a page while flipping to distal pages with the preferred hand (Fig. 2f).

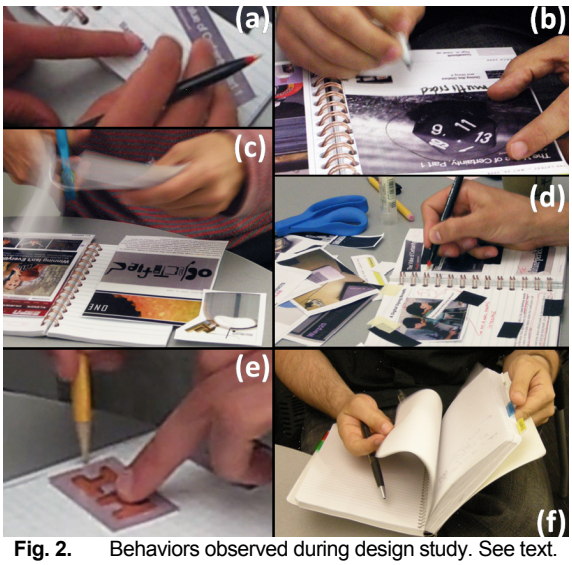

Fig. 2. Behaviors observed during design study. See text.

The observations above contribute naturally occurring behaviors, with both pen and bare-hand manipulation, that exhibit a clear differentiation between the roles of pen and touch. Terrenghi also observes a rich set of bimanual behaviors resulting from physical manipulation of clippings [35]. These stand in contrast to the results of user-elicited gestures for pen and touch [16], which might otherwise lead us to treat many pen and touch inputs interchangeably. Our observations suggest it would be mistaken to do so. By the same token, our goal is not to mimic paper. Pen+touch should go beyond paper, but the best foundation for pen+touch gestures likely lies in behaviors that people already exhibit when working with pen, paper, and books.

DESIGN CONSIDERATIONS FOR PEN AND TOUCH

Informed by the above observations, as well as current research and industry trends in direct input devices, we can now show how the design space of pen + touch interaction raises many interrelated factors and considerations. The most obvious, of course, is the factor of ***pen vs. touch input*** itself. Without feigning completeness, the following tableau summarizes the interaction properties shared by pen and touch of particular relevance to our discussion here. We do not characterize such properties as "pros" and "cons," as has been attempted elsewhere [6], to accentuate our belief that almost any property of a device can be advantageous in interaction design. For example, accidental palm contact can seen as a problem, or it can viewed as an opportunity to correctly orient menus for each of multiple table users [7].

PROPERTY	PEN	TOUCH
Contacts	**1 point** *A single well-defined point.*	**1-10+ contact regions** *with shape information.*
Occlusion	**Small (pen tip)** *But hand still occludes screen.*	**Moderate** *("fat finger") to* **Large** *(pinch, palm, whole hand)*
Precision	**High -** *Tripod grip & lever arm for precision, writing, sketching.*	**Moderate**
Hand	**Preferred hand**	**Either hand / Both hands**
Elementary Inputs	**Tap, Drag, Draw Path, Crossing**	**Tap, Hold, Drag, Pinch, 2-finger Hold (thumb + index, Fig. 2b)**
Inter-mediary	**Mechanical Intermediary** *Takes time to unsheathe the pen. Pen can be forgotten.*	**None: Bare-Handed Input** *Nothing to unsheathe or lose. No lever arm. No buttons.*
Acquisition Time	**High** *(first use: unsheathe pen)* **Moderate** *on subsequent uses: tuck pen between fingers (B1).*	**Low** *No mechanical intermediary to acquire.*
Activation Force	**Non-Zero** *Tip switch/ minimum pressure.*	**Zero.** *Contact area (a proxy for pressure) often can be sensed.*
False Inputs	**Palm Rejection** (while writing) *Palm triggers accidental inputs, fingers drag on screen, etc.*	**Midas Touch Problem** *Fingers brush screen, finger rests on screen while holding it.*

Fig. 3. Summary of design properties for pen vs. touch.

We can now briefly map out the main design considerations that we believe the design space of pen+touch raises:

1. *Pen vs. Touch:* The type of input, as summarized above.
2. *Differentiated vs. Interchangeable task assignment:* For a given input task, does the interface differentiate between pen and touch, or are they interchangeable?
3. *Nonpreferred vs. Preferred hand assignment:* Which device is assigned to which hand, and under what circumstances?
4. *Unimodal vs. Multimodal input:* Inputs can be articulated using either unimodal pen, unimodal touch, or multimodal (simultaneous) pen+touch. How are each of these used to best advantage in an interface design?
5. *Mobile vs. stationary use:* Desktop, table, and wall direct input devices are necessarily stationary, but form-factors such as slates and booklets [11,20] transition between mobile and stationary use (the user can pick them up, or put them on a desk). How can pen+touch support a common conceptual model to enable *graceful degradation* between stationary and mobile usage?
6. *Unimanual vs. bimanual interaction.* For what tasks and usage scenarios does the user interact using one hand, versus both hands? Can the user complete core tasks one-handed? Does the interface leverage the full potential of human hands, as well as simultaneous pen + touch input, when working with both hands?

7. *Interleaved vs. Simultaneous inputs:* Does an interaction require simultaneous pen+touch, or can users perform it by interleaving a series of touch-pen-touch interactions, with the user tucking the pen as needed?
8. *Ink vs. Command mode:* Are movements of the device left as an ink stroke (or finger smudge)? Or are they instead interpreted by the system as a gesture (or direct manipulation) that acts on the underlying object?
9. *Elementary inputs vs. Phrases:* Does the design consider just the compatibility of the input device to individual interactions, or does it support a workflow across multiple tasks, or subtasks, that therefore afford phrasing together of actions into compound tasks?

We do not have answers to all of these questions. But our enumeration of these considerations builds on the contributions of Brandl [6] to further map out the design space of pen and touch. Furthermore we can now articulate how our approach falls into place along these dimensions:

- Unimodal pen vs. touch have differentiated effects in the interface (per observation *B0*). We assign ink mode to the pen, while multi-touch gestures articulate commands: *the pen writes, and touch manipulates.*
- By tucking the pen (per *B1*), the user can efficiently interleave pen and touch inputs with the preferred hand for mobile, unimanual usage scenarios;
- Designing core tasks for unimanual touch serves mobility while also enabling stationary bimanual interaction that *instead assigns these tasks to the nonpreferred hand*, thus supporting more efficient interaction as well as advanced gestures;
- These benefits are derived while retaining the possibility of bimanual interactions with multimodal pen and touch: *pen + touch = new tools.*
- For simultaneous pen+touch, holding a finger on an object (*B2*) integrates object selection, transition to gesture mode, and the phrasing together of multiple pen or touch inputs into a compound gesture [9,34]. That is, pen strokes in reference to an object while the user holds it are recognized as gestural commands, rather than ink strokes that would otherwise mark the object.

With the foundation for our approach now set, we turn our attention to our prototype application and the novel interaction techniques that it supports.

APPLICATION SCOPE AND MOTIVATION

Manual Deskterity is intended primarily as a research vehicle to explore pen + touch, which we believe has many potential applications. Nonetheless we emphasize practical tools for note-taking and mark-up noted by previous work (e.g. [20,23,28]). For example, several papers emphasize reading [1,24] in conjunction with writing, annotation, selecting, copying, arranging, and aggregating objects, both in digital [2,38] and paper-based knowledge work [4,24,35]. In this context reading as a pure "consumption" activity is a myth; it is inseparable from the annotation and creation of new work artifacts. Furthermore, jotting notes, or annotating a book, take place in a surround of tasks and activities that embody the user's higher level workflow [1,24]. What other information is read and cross-referenced with the sheet of paper the user is currently writing on? How does the user structure and access this workspace? Our observations of user's extended workspaces (B6), the frequency of reaching into the space above the notebook to retrieve objects (also B6), and the facility with which users flip back and forth between distal pages of a notebook (B9) serve as specific examples. Since hand and stylus are closely tied to how people work with books and documents, to realize pen and touch techniques along these themes *Manual Deskterity* supports both whole-screen and dual-screen [11,20] views. Pen+touch techniques may afford digital workspaces with a feel closer to the informal way that people actually work with sets of active documents [1,24] and scraps of information [28].

Our design study suggests a few additional behaviors of interest in the context of simultaneous pen + touch, such as holding items while acting on them with the pen (B2, B3, B4), cutting operations (B5), and employing clippings as a constraint for the pen (B8). These and the considerations discussed above motivate the features that we elected to explore using pen, touch, and pen+touch interactions.

Implementation

We use a Microsoft Surface with a custom infrared LED pen, activated during contact via a tip switch. Our pen is much brighter than hand/finger contacts, so it is trivial to identify as the brightest spot in the image. The software is written in C# with WPF and the Microsoft Surface SDK.

A potential limitation of pen+touch is the "palm rejection" problem: the user may rest his hand on the screen while writing, potentially leading to unintended operations. We treat touches with a large contact area as incidental. This is sufficient for prototyping pen + touch techniques, but robust handling of incidental contacts remains an important problem for future work. Our approach does not solve it.

CORE TASKS: PEN WRITES, TOUCH MANIPULATES

We drive our core interactions by touch interactions including zooming, flipping pages, moving and selecting objects, and creating new objects (e.g. digital post-it notes). For writing, only the pen produces ink strokes (although in some contexts the finger "smears" colors, as discussed later). Notwithstanding the exception of finger-painting, for these core tasks *the pen writes, and touch manipulates*, period. This makes the entire canvas available for immediate annotation with the pen, while pan/zoom, page navigation, and object manipulation are also immediately available via touch, without any explicit mode switches. Of course, we say *the pen writes* as shorthand for any graphological act that leaves a mark, whether it comprises sketching, annotating, cross-hatching, or handwriting.

Manipulating, Zooming, and Selecting Objects

The user can arrange objects such as photos and post-it notes on the canvas via direct manipulation (single touch drags

objects, while pinching objects rotates and scales them). Two-finger pinch on the canvas zooms the entire page, allowing detailed edits. Since these core tasks are driven by unimanual touch, users can perform these actions with the preferred hand (e.g. for mobile usage scenarios), or with the nonpreferred (while the preferred hand remains ready to go with the pen) for bimanual usage scenarios.

The user can select multiple objects via finger-taps that incrementally add individual objects to the current selection. Dragging an item that is already selected drags it *and all other selected items* while maintaining their relative spatial relationship. Dragging an item that is *not* already selected drags *only that item*; this enables multi-touch dragging of items to multiple different locations.

Pen always writes vs. pen drags selected items: In our system, the pen (when used as a unimodal input) always writes on items, even if they are selected. Beyond recourse to our observation of *specific roles (B0),* why is this so? We experimented with moving selected objects with the pen to support a precise dragging mode (as proposed by Brandl [6]). However, we found that users would frequently intend to write on a selected item with the pen, but drag it by mistake. The feedback for selection is a prominent pink highlight that surrounds the object. But this feedback is delivered through the visual channel, so it is not particularly salient to the user [34]. How might this be improved? We also tried dragging the item *only when the user holds it with the finger*. The muscular tension of holding a finger on the object provides more salient feedback, and as a result dragging with the pen feels more natural to users. However, this is now a *multimodal pen+touch* gesture, which has many more possibilities; we will return to these shortly.

Interchangeable inputs in common controls: On selection, a radial menu appears at the upper right corner of objects, with options such as *copy, delete, clear annotations,* etc. Initially, our radial menus required use of the pen, but we found that all test users expected this menu to be operable via touch as well. Although we have so far advocated differentiation between the roles of pen and touch, this is a good example of a context where users clearly expect pen or touch inputs to have the same effect in the interface. We still advocate the principle that *the pen writes and touch manipulates*, but the lesson here is that an exception must be made for common controls such as menus, close boxes, sliders, and buttons that represent limited spatially-multiplexed contexts where pen and single-touch inputs should indeed be interchangeable.

PEN + TOUCH YIELDS NEW TOOLS

Next, we explored how the expanded input vocabulary afforded by the combination of pen and touch can support new tools. Our design uses a few primitives for all pen + touch operations. The richness of the gestures arises from how the primitives are combined. For the pen the primitives we use are tap, drag-off, crossing, or drawing a stroke. For touch, we employ single-finger tap, single-finger hold (as seen in our design study, behavior B2), holding with thumb and forefinger (B4), and crossing. We do not implement all combinations, but rather support a sufficiently rich set of operations, with semantics that map well to our application domain, to illustrate the expressiveness of our approach.

The pen+touch techniques described below all use fingers of the nonpreferred hand hold an item, while the pen acts in reference to the item. This builds on the tendency that we observed for users to hold clippings with the nonpreferred hand while making pen markings in reference to them (B3), and it also corresponds to Guiard's principle that the nonpreferred hand sets the frame of reference for the preferred hand [18]. Thus, we construct gestures that allow for non-physical digital effects, yet remain grounded in people's naturally occurring behaviors with physical paper.

Stapler: Grouping Items into a Stack

To support piling (observation B7), we implemented stapling items into a stack. The user can finger-tap-select multiple items, and then staple them together by holding an item and tapping it with the pen. The item that the user performs the "staple" (hold-and-tap) gesture on moves to the top of the stack, so that it represents the entire pile.

Tap-selecting a series of items and stapling them together enables quick tidying of a messy work surface into a few piles. This transaction separates the identification of the items to stack from the decision of which item should become the representative item on the top of the resulting stack. Performing the pen+touch gesture on the representative item keeps the user's attention focused on it, at exactly the moment the user makes this decision, which (based on our informal user studies) corresponds well with users' mental model of the task. While piling has been explored by other systems (e.g. [2]), what is unique here is the facility of creating stacks, with a representative image on top, via a dedicated pen+touch gesture. Test users rated this as one of their favorite tools, and particularly liked the ability to easily specify which object is on top of the stack.

We could have used tapping with a second finger for stapling. Why use the pen instead of touch? Although our test users did not ask for this, the question once again highlights the tradeoff of differentiated vs. interchangeable pen and touch inputs. Because touch requires zero contact force, we have observed that accidental taps are common. That is, holding items and tapping with a finger is prone to accidental activation in a way that holding plus tapping with the pen is not. Using pen and touch for this gesture therefore feels more distinct to the user, while also more effectively leveraging the strengths of each input modality.

X-acto Knife: Cutting Items (and Tearing Items)

The user can turn the pen into an X-acto knife by holding an object and fully crossing it with the pen. That is, the pen stroke starts outside the object, crosses through the interior of the object, and finishes on the exterior of the object. The pen stroke within the item can follow any path, allowing intricate cuts if desired (Fig. 4). When the pen exits the item, both the cut and the scrap piece appear on the page, following the real-world action where we observed scraps falling onto the work

surface (design study B5). Users found the technique natural and compelling, even though not all saw a need for it in their daily work. Most users liked the scrap cuts, and expected both pieces to remain: as one user asked, "How would it know which half I want?"

Fig. 4. Example X-acto cuts created by one test user.

To probe the semantics of pen vs. touch in analogous gestures, we also implemented tearing items by holding an object with a single finger, and then crossing the item with another finger. This tears the item along the line connecting the entry and exit points of the finger. This is similar to cutting, but produces a different visual affordance. This technique demonstrates how touch can sometimes be used to sneak a nuance of expression into a transaction, by applying a different look or different default command parameters. On the other hand, in our system this precludes using the touch gesture for a different command, such as layering [14]. Should the semantics of analogous pen+touch vs. touch+touch gestures be similar or contrasting?

In this regard, our test users did not perceive tearing photos as a particularly useful operation; for example some users suggested that tearing could support different semantics, such as deleting the object. But they also did not suggest that both versions of the gesture should provide the same X-acto cut behavior. The lesson is that users perceive different semantics; most people do not expect analogous pen+touch and touch+touch gestures to do the same thing.

Carbon Copy: Drag-Off with the Pen

The user copies an object by holding it with a finger (B2) and then "peeling off" a copy with the pen (Fig. 7). This gesture is similar to Frisch's copy gesture [16], but here we identify the interaction pragmatics as well as why this gesture differs from a *Copy* command on a context menu, for example. Here we seen an example of how the user-elicited gestures methodology can yield fertile ground for suggesting plausible gestures (so long as we keep in mind that users are not designers [17]).

Once the pen drags away by a minimum distance, a semi-transparent copy of the object appears attached to the pen. As the pen continues to drag, the object becomes opaque. The user continues dragging the object to "paste" it at the desired location. Annotations on the object are also copied.

Our system also supports *Copy* via the object's context menu. Drag-off with the pen phrases together the entire transaction (select, copy, and drag to final position) into a single cognitive chunk [9] via the muscular tension of the nonpreferred hand holding down the original item. We found that this corresponds well to users' mental model of duplicating items– they not only want to copy the item, but also place the duplicate at a particular location. By contrast, *Copy* from the context menu divides this select-copy-position transaction into multiple steps, enabling one-handed copying at the cost of syntactical complexity [9].

Holding an item and peeling off a copy with the pen is a good example of a pen+touch technique that ostensibly violates the principle that *the pen writes and touch manipulates*, because here the pen drags the copy. However, those principles apply to pen or touch as *unimodal inputs*. The transaction is consistent with the principle that guides our multimodal gestures: *pen + touch yields new tools*. Our test users found the gesture natural and effective. Furthermore it is grounded in people's naturally occurring behaviors with physical paper, such as holding an item and making pen strokes in reference to it (Guiard [18], and our design study observations B2, B3).

Ruler: Using an Object as a Straightedge

The user can employ an object as a straightedge by holding down the object with the thumb and index finger, like the framing gesture observed in our design study (B4, Fig. 2b). The user can then stroke along the object with the pen constrained to its border, as inspired by observation B8. When two fingers come down on an object, an animation starts that increases the transparency of the item and adds a dotted line around its border. Informal test users who tried early versions of the system suggested that items should become mostly-transparent in this manner. This enables users to see the relationship of the straightedge to the underlying surround of other objects and strokes.

The Ruler uses content as its own tool, introducing a subtle duality between content and tool into our system. We could add a dedicated ruler, but this would necessitate acquiring the ruler before drawing a straightedge. Using an object as its own straightedge also supports sketching techniques such as adding a *drop shadow* or *outer glow* to an item. To be clear, however, we are not arguing that the system should not include a ruler tool; rather, we are arguing that using an object as its own straightedge has interaction design properties that a dedicated ruler tool does not.

Composition of the Straightedge with Cutting

To illustrate how our interaction design allows multiple interactions to be composed together into phrases, we implemented X-acto cutting along straightedges (Fig. 5). The user finger-tap-selects the photo to cut, and then uses the two-finger thumb and index finger grasp on an overlapping item to establish a straightedge. The user then strokes the pen across the selected photo along the straightedge to cut it. The user can even cut around the corner of the overlapping item. Composition of techniques is possible because tapping

provides a way to define a collective scope [25] of more than one object, and because holding phrases together multiple pen and touch inputs into a single chunk that the system interprets as a command [9].

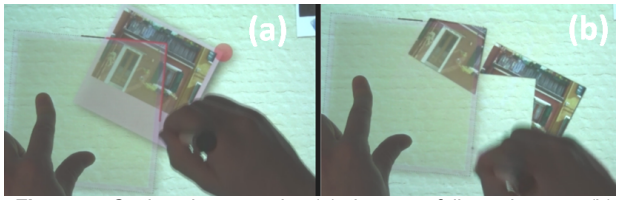

Fig. 5. Cutting along an edge (a); the scrap falls on the page (b).

Brushing: Using a Photo as Its Own Tool

Our system uses pen+touch to enable the user to employ content on the page as a brushing tool. If the user holds a photo and strokes the pen on the canvas, the pen stroke uses the photo as a brush effect (Fig. 6). While the user continues to hold, each subsequent pen stroke is mapped to a common viewport within the photo so that the user may "rub out" different portions of the photograph (see video). This technique is similar to the *Art History Tool* in Adobe Photoshop, but our technique can produce some creative effects without the complexity of modes, parameters, and layers required to use the Photoshop Art History Tool.

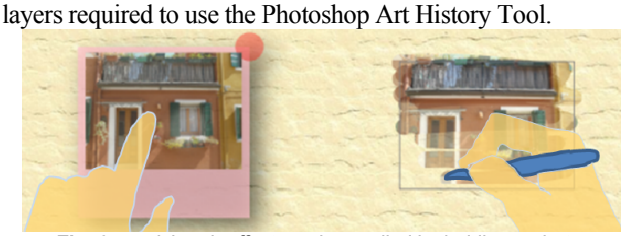

Fig. 6. A brush effect can be applied by holding a photo.

This technique provides another example of the facility with which the user can hold an item to produce a mode that is specific to that object, or that uses the object as a parameter. It also illustrates how the application of that mode can include one or more strokes (applications) of the tool while the nonpreferred hand holds the state.

Tape Curve: Holding a Pen Stroke as a Drawing Tool

Tape drawing is a familiar technique for illustrators: any craft store carries "artist tape for curves." To produce a straight line the artist pins the end of the tape with the left hand, unrolls a length of unfastened tape with the right hand, and then slides the left hand along the tape to adhere it. Alternatively, to draw curves, the artist pivots the unfastened length of tape in the right hand while the left hand follows the tape. Balakrishnan [3] observes that corresponding *digital tape* exhibits an intriguing interaction property: it supports both straight lines and variable radius curves within the same tool, without switching modes.

Balakrishnan's implementation is in "tape drawing mode" by default, and requires physical buttons [3]. But in our system, by default the pen draws freeform ink strokes. While the user is drawing a pen stroke, he just holds a finger on it to turn the remaining portion of the stroke into a tape curve. Lifting the finger from the stroke immediately reverts to the default freehand ink state. While the technique seems intriguing, we have not yet user-tested it.

Fig. 7. Tape: Users can fluidly mix ink, straight lines, & curves.

Color Palette: the Finger Shadow

We have also experimented with the Finger Shadow (Fig. 8), which uses a finger-tap on the canvas to bring up in-place commands. This accesses additional commands, including color pots (Fig. 8). In early demonstrations of our system we observed that most people expect touching the color pots to enable smearing of colors with the finger. To probe this issue further, in this context we *intentionally* break our design rule for unimodal inputs that *the pen writes and touch manipulates*. To some degree, this echoes our earlier observation that common controls should allow use of either pen or touch. Yet here, pen and touch are not completely interchangeable either: the pen produces a crisp ink stroke, while the finger smears colors onto the page (Fig. 8, right). To prevent finger painting from becoming a heavyweight mode, each successive finger paint stroke appears fainter until "all the ink has been rubbed off" the user's finger (Fig. 8, right). This echoes how artists work with physical media such as charcoal sketching.

Fig. 8. **Left**: Dragging from the Finger Shadow (here with the pen) opens the menu. **Right**: Finger painting using the color pots.

With finger painting, we have to face a genuine design dilemma: when ink remains on one's finger, should it be possible to finger paint on top of objects? In our system, the answer is no. Touching an object to select or move it (or touching other controls, such as flipping pages) always takes precedence over finger painting. Otherwise, a more explicit means to "stop" finger painting and return to the default behavior that *touch manipulates* would have to be introduced, and we did not wish to do so at this point.

NEW "TOUCHES" ENHANCE PEN+TOUCH WORKFLOW

Finally, we explore some techniques that extend the interaction states that can be supported with multi-touch interactions. This enhances the workflow of interleaved touch-pen-touch interactions, whether articulated with both hands, or with one hand by alternating pen and touch.

Creating New Objects via the Bezel Menu

Earlier we observed that users bring in new materials from above (B6). Creating a new object implies making something out of nothing: how does one "directly manipulate" objects that are not on the screen? Furthermore, most of the time one wants to create and place an object [25]. Standard answers to these questions consume screen real estate (e.g. tool palettes),

or input device state transitions (e.g. buttons, tap-and-hold) that either aren't available for direct input, or could be better used for other purposes. To address this challenge we devised a novel finger-activated *bezel menu* that extends the concept of bezel-crossing gestures [30], while adapting from marking menus the concept a graceful novice-to-expert transition with self-revealing gestures [25].

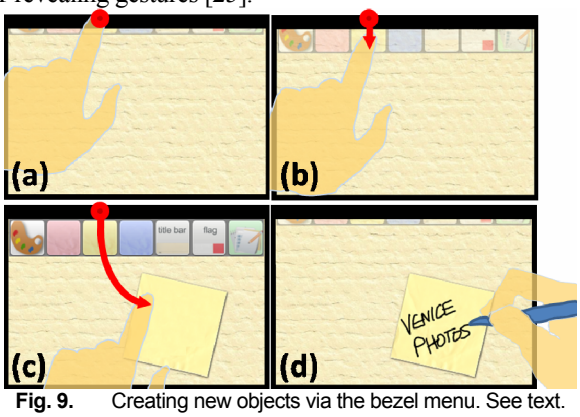

Fig. 9. Creating new objects via the bezel menu. See text.

The bezel menu is visible only as a semi-transparent strip along the top edge of the screen (Fig. 9a). To access it, the user slides a finger across the bezel, which starts to pull out the bezel menu (Fig. 9b). With the bezel menu open, the user may then tap on an icon to create a new object such as a post-it note. This is the self-revealing, "novice" mode of making a selection. The bezel menu then hides itself again. In the expert gestural mode, once a user is familiar with the location of commonly used items in the bezel menu, he can perform a continuous finger-drag that crosses through the item and onto the canvas (Fig. 9c) to *create and place an object* (such as a digital post-it) in a single transaction. This aspect of the technique particularly resonated with test users. The user can then immediately annotate the post-it, thus supporting a tight create-place-annotate workflow.

Navigation to Pages, Books, & the Extended Workspace

To flip the page, we again employ a bezel gesture: the user swipes across the top right corner to go to the next page, or across the top left corner to go to the previous page. The user can flip through *multiple* pages by starting at the corners, and then dragging downward to scrub through the notebook. A page curl in these corners provides a subtle visual affordance for page flipping. Page flipping responds only to touch; the pen can still write in this area.

Flipping between Distal Locations

To support observation B9, *Hold Page while Flipping* (Fig. 2f), we implemented a technique whereby the user can hold their place in the dual-screen view of the digital notebook by resting a finger of the nonpreferred hand near the margin of the page (Fig. 10a). This enables the user to keep their place in a document with very little effort or conscious attention– much as is the case for real document navigation. All the user has to do is hold the current page with the nonpreferred hand. As long as the user holds it, he can navigate to other pages (Fig. 10b) and later flip back by sliding the nonpreferred hand towards the spine of the book (Fig. 10c, d). Users liked flipping back and forth, particularly for comparing pages, but sometimes lifted by accident (thus losing the "held" page). This interaction was only implemented for the dual page view because the metaphor seemed to make the most sense there.

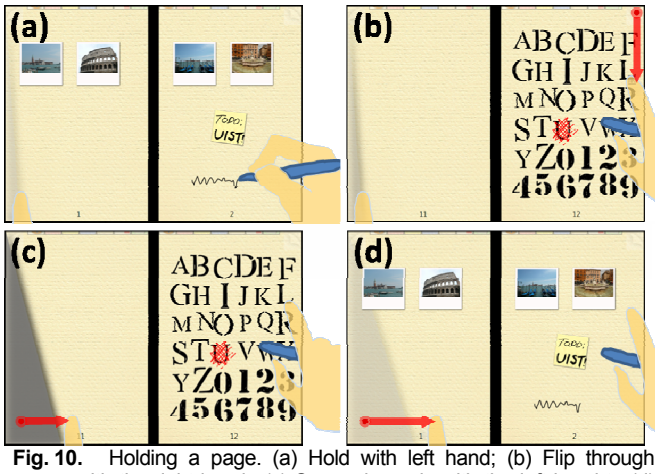

Fig. 10. Holding a page. (a) Hold with left hand; (b) Flip through pages with the right hand.; (c) Sweep inwards with the left hand to (d) flip back to the original "held" page.

Cross-Screen Pinch: Zoom to Extended Workspace

What about navigation to the "extended workspace" surrounding the notebook (B6)? In dual-screen view, our application supports *cross-screen pinch* to semantically zoom to a "desk view" that serves this role (Fig. 11). Here, the user can arrange books, documents, and photos. The user can leave post-it notes or other visuo-spatial cues as contextual reminders that implicitly indicate what to do next when resuming a task [4, 24]. This aspect of the extended workspace particularly resonated with users.

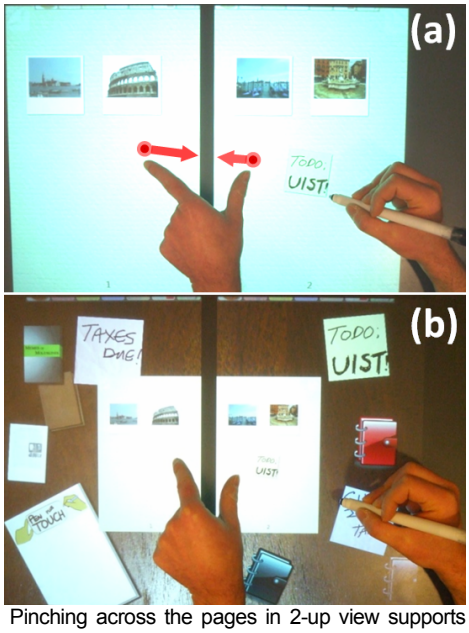

Fig. 11. Pinching across the pages in 2-up view supports semantic zooming to the "desk view" extended workspace.

Note that a pinch gesture that spans the two pages offers a gesture that is distinct from both zooming within a page (two-finger pinch on a page) as well as scaling of individual objects

on the page (via two-finger pinch of the object). Thus, like bezel gestures, *cross-screen gestures* offer another "escape hatch" in the interaction design language that enables simple and unambiguous expression of additional interactions with gestures that correspond well to the semantics of the desired operation– here, semantic zooming to an extended workspace.

Summary Discussion of "New Touches"

One could argue that these are just touch techniques that have little to do with pen and touch input. Yet while writing, the secondary tasks of creating new objects, flipping back and forth between pages, and navigating to the extended workspace can all help retrieve, compare, and create task-relevant information. Because these fluidly support the primary task of writing and annotating, these touch techniques afford tight interleaving of secondary tasks (via the nonpreferred hand) with the primary task (writing and annotation with the pen). This helps to "keep the user in the flow" of concentrated work. It is also telling that our desire to support this workflow helped to spark the idea for techniques that extend direct touch interfaces, which comes full circle to our belief that the presence of both pen and touch modalities can alter our perspective on how to most effectively design for touch itself.

DISCUSSION & HIGH-LEVEL REACTIONS FROM USERS

Our multimodal pen+touch gestures use touch to hold items, in combination with one or more pen strokes that act in reference to an item that the user is holding. This approach enforces a strong notion of phrasing [9] in the resulting interface design: there are no persistent modes, but rather tool use is always tied to holding an object with the nonpreferred hand. There is no possibility of getting stuck in a tool mode with the pen, nor is there ever a question of how to return to the default action of drawing ink strokes on the page. Once the user releases objects, drawing on the page with the pen always leaves ink strokes.

This approach is akin to nonpreferred-hand mode switching [27], but requires no physical button and thus readily scales to a plurality of modes [32]. It also offers an additional advantage: object selection is integrated with the mode switch itself when the user's hand touches down on an object on the display. Hence a unique design property of pen+touch is the facility with which it can support modes and tools specific to particular objects on the screen. That is, we contribute the insight that object selection, mode switching, and phrasing together multiple inputs are all seamlessly integrated in this class of technique.

While our emphasis in this paper has not been on evaluation, we founded our approach on a design study of people's naturally occurring behaviors, and we have tested the techniques with 11 test users. Users may not have need for some specific tools, and although our current prototype includes all of these techniques, it is not our expectation that future systems would need to include all of these techniques, as some have been added for pedagogical reasons so as to draw out the nuances of how pen and touch complement one another.

Thus, our current gesture set is more complex than a "real" production system would support. The combination of pen + touch does not automatically extract us from the classical trap with gestural interfaces: each additional gesture that one adds to a system proportionally reduces the value of all the other gestures. However, the distinctness of using pen versus touch, coupled with the highly contextual nature of human memory, does seem to modestly increase the number of gestures that a system can include, versus a system based solely on touch, for example. Nonetheless this belief remains to be tested by formal experiments.

Users found our approach to combined pen and touch interaction appealing. They quickly formed habits around the general pattern (hold touch + gesture with pen) common to most of our gestures. As one user commented, "I didn't think about switching between pen and touch at all… it just kind of happened…" Another commented that "the way it works is just like the way I already work in my notebook."

But clearly, we have a *designed* gesture set that we would not expect users to "guess" without guidance. As one user commented, "I wouldn't have guessed the gestures work that way, but once I tried it, it felt pretty natural." This does bring us to one general weakness of the prototype: currently the multimodal pen+touch gestures are not self revealing, nor in most cases is there sufficient feedback of the current action of the tools until after the user finishes articulating a gesture. In some cases it would be difficult to add feedback at the initiation of a gesture because the resulting effect (e.g. stapling vs. drag-off-copy) cannot be known until the pen is lifted or starts to move. This issue is exacerbated in our prototype because our pen has no hover state, which precludes the possibility of showing the current function before the pen touches down. Nonetheless we believe existing self-revelation techniques, such as GestureBar [5] or Highlighter Hints [23], would be suitable for integration with our prototype.

As noted by a reviewer, the concept of an *object* underlies many of our gestures, but also suggests a more nuanced hierarchy of contextual actions on object types. The *canvas* is a manipulable element (it zooms and pans) as well as a container for *pieces of documents* (e.g. post-its, photos). These pieces serve as the background for smaller entities, such as notations upon a post-it. Indeed, users expect the marks upon such pieces to move with the piece; our implementation even goes so far as to split ink strokes at the boundaries of the pieces to maintain this metaphor. Furthermore, "system" objects such as *menus* and *tools* play by different rules that the pieces. Even this category is blurred (as it is in real life), for example by our use of photos as straightedges. This hints that perhaps a deeper theory or guiding principles for contextual actions, which act upon on a hierarchy of object types, may exist.

CONCLUSION AND FUTURE WORK

Despite rapture with the iPhone (and now iPad), multi-touch is not the whole story. Every modality, including touch, is best for something and worst for something else. The tasks demanded of knowledge workers are rich and highly varied [1,4,24], and as such one device cannot suit all tasks equally well. One's finger is no more suited for signing a contract, or drawing a sketch on a napkin, than is a pen for turning the page of a book, or holding your place in a manuscript. With the addition of the pen, user interfaces afford creation of new ideas, rather than unbounded consumption of content produced by others.

We have advocated an approach where *the pen writes, touch manipulates, and the combination of pen+touch yields new tools*. This articulates how our system interprets unimodal pen, unimodal touch, and multimodal pen + touch inputs, respectively. The space of pen and touch, while potentially complex, is also sufficiently rich that it offers the opportunity to craft new user experiences that are uniquely well suited to how people naturally work with pen and paper—without being beholden to physical mimicry of paper at the same time. It remains to prove that our approach scales from a demo to a full-blown application, nor have we yet demonstrated how these rich techniques enhance the effectiveness and user experience of less glamorous applications, such as working with a spread-sheet. We believe that they can and will, but that is a long-shot from actually doing so. As well, despite our hopes and projections as to how well these techniques will work on other form factors, the fact remains that we have not yet done those tests. There is still work to do.

ACKNOWLEDGEMENTS

Special thanks to Georg Petschnigg, Andrew Allen, Jon Harris, Scott Erickson, Sabrina Boler, Tito Pagan, Joel Jacobs, and Yasser Asmi for ideas, insights, & discussions.

REFERENCES

1. Adler, A., Gujar, A., Harrison, B., O'Hara, K. Sellen, A. *A diary study of work-related reading design implications for digital reading devices*. CHI'98. p. 241 - 48.
2. Agarawala, A. Balakrishnan, R. *Keepin' It Real: Pushing the Desktop Metaphor with Physics, Piles and the Pen*. CHI '06.
3. Balakrishnan, R., Fitzmaurice, G., Kurtenbach, G., Buxton, W. *Digital Tape Drawing*. UIST'99. p. 161-70.
4. Bondarenko, O., Janssen, R. *Documents at Hand: Learning from Paper to Improve Digital Technologies*. CHI 2005.
5. Bragdon, A., Zeleznik, R., Williamson, B., Miller, T., LaViola, J. *GestureBar: improving the approachability of gesture-based interfaces*. CHI'09. p. 2269-78.
6. Brandl, P., Forlines, C., Wigdor, D., Haller, M., Shen, C. *Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces*. AVI 2008, p. 154-61.
7. Brandl, P., Leitner, J., Seifried, T., Haller, M., Doray, B., To, P. *Occlusion-aware menu design for digital tabletops*. CHI 2009 Extended Abstracts. p. 3223-28.
8. Buxton, B., *Sketching User Experiences: Getting the Design Right and the Right Design*. 2007, San Francisco: Morgan Kaufman.
9. Buxton, W. *Chunking and Phrasing and the Design of Human-Computer Dialogues*. IFIP `86, p. 475-80.
10. Buxton, W., Myers, B. *A Study in Two-Handed Input*. CHI'86.
11. Chen, N., Guimbretiere, F., Dixon, M., Lewis, C., Agrawala, M. *Navigation Techniques for Dual-Display E-Book Readers*. CHI '08.
12. Cohen, P. *The role of natural language in a multimodal interface*. UIST'92. p. 143-49.
13. Cohen, P., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen,L., Clow,J. *QuickSet: Multimodal Interaction for Distributed Applications*. ACM Multimedial 97.
14. Davidson, P.L., Han, J.Y. *Extending 2D object arrangement with pressure-sensitive layering cues*. UIST '08.
15. Engelhard, L. *Native Dual Mode Digitizers: Supporting Pen, Touch and Multi-Touch Inputs in One Device on any LCD*. Society for Information Display SID 08 Digest. p. 1306-09.
16. Frisch, M., Heydekorn, J., Dachselt, R. *Investigating Multi-Touch and Pen Gestures for Diagram Editing on Interactive Surfaces* ITS '09 Conf. on Interactive Tabletops and Surfaces.
17. Greenberg, S., Buxton, B. *Usability evaluation considered harmful (some of the time)*. CHI '08. p. 111-20.
18. Guiard, Y., *Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model.* The Journal of Motor Behavior, 1987. **19**(4): p. 486-517.
19. Hartmann, B. *Pictionaire: Supporting Collaborative Design Work by Integrating Physical and Digital*. CSCW '10.
20. Hinckley, K., et al. *Codex: a dual screen tablet computer*. CHI '09.
21. Hinckley, K., Pahud, M., Buxton, B. *Direct Display Interaction via Simultaneous Pen+Multi-touch Input*. SID 2010 Digest, p. 537-40.
22. Hinckley, K., Pierce, J., Sinclair, M., Horvitz, E. *Sensing Techniques for Mobile Interaction*. UIST 2000, p. 91-100.
23. Hinckley, K., Zhao, S., Sarin, R., Baudisch, P., Cutrell, E., Shilman, M., Tan, D. *InkSeine: In Situ Search for Active Note Taking*. CHI'07.
24. Kidd, A. *The marks are on the knowledge worker*. CHI '94. 186-91.
25. Kurtenbach, G., Buxton, W. *Issues in Combining Marking and Direct Manipulation Techniques*. UIST'91.
26. Kurtenbach, G., Fitzmaurice, G., Baudel, T., Buxton, B. *The Design of a GUI Paradigm based on Tablets, Two-hands, and Transparency*. CHI'97, p. 35-42.
27. Li, Y., Hinckley, K., Guan, Z., Landay, J. *Experimental Analysis of Mode Switching Techniques in Pen-based User Interfaces*. CHI '05.
28. Marshall, C., Bly, S. *Saving and Using Encountered Information:Implications for Electronic Periodicals*. CHI '05.
29. http://en.wikipedia.org/wiki/ Microsoft_Courier. 2009.
30. Roth, V., Turner, T. *Bezel swipe: conflict-free scrolling and multiple selection on mobile touch screen devices*. CHI 2009.
31. Ruiz, J., Bunt, A., Lank, E. *A model of non-preferred hand mode switching*. *Graphics interface 2008* p. 49-56.
32. Ruiz, J., Lank, E. *A study on the scalability of non-preferred hand mode manipulation*. ICMI '07, p. 170-77.
33. Schilit, B., Golovchinsky, G., Price, M. *Beyond paper: supporting active reading with free form digital ink annotations*. CHI'98.
34. Sellen, A., Kurtenbach, G., Buxton, W., *The Prevention of Mode Errors through Sensory Feedback.* J. Human Computer Interaction, 1992. **7**(2): p. 141-64.
35. Terrenghi, L., Kirk, D., Sellen, A., Izadi, S. *Affordances for manipulation of physical versus digital media on interactive surfaces*. CHI '07. p. 1157-66.
36. Wilson, A.D., Izadi, S., Hilliges, O., Garcia-Mendoza, A., Kirk, D. *Bringing physics to the surface*. UIST 2008.
37. Wu, M., Balakrishnan, R. *Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays*. UIST 2003.
38. Wu, M., Shen, C., Ryall, K., Forlines, C., Balakrishnan, R., *Gesture Registration, Relaxation, and Reuse for Multi-Point Direct-Touch Surfaces*, *IEEE Intl. Workshop on Horizontal Interactive Human-Computer Systems*. 2006.
39. Yee, K.-P. *Two-Handed Interaction on a Tablet Display*. CHI 2004.