38.2: Direct Display Interaction via Simultaneous Pen + Multi-touch Input

Ken Hinckley, Michel Pahud, and Bill Buxton
Microsoft Research, Redmond, WA 98052

Abstract
Current developments hint at a rapidly approaching future where simultaneous pen + multi-touch input becomes the gold standard for direct interaction on displays. We are motivated by a desire to extend pen and multi-touch input modalities, including their use in concert, to enable users to take better advantage of each.

1. Introduction
We are witnessing a shift towards displays that unify *input* and *output* on surfaces that sense as well as emit. In such systems the user interacts through *direct manual input*, that is, directly on the display with his hands. By contrast, traditional graphical interfaces employ *indirect manual input* [5] using a relative pointing device (mouse) and a cursor. This shift has led to renewed interest in both touch and pen input. When integrated with a display, both pen and touch are *direct input modalities*, albeit through a physical intermediary in the case of the pen. In what shall become a theme here, this is both a strength and a weakness for the pen— as is the lack of an intermediary for touch. Having two opposing sides makes a coin no less valuable.

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 [1,13]. As such one device cannot suit all tasks equally well. Your finger is no more suited for signing a contract, or drawing a sketch on a napkin, than is a pen for turning the page in a book, or holding your place in a manuscript. With the impetus to do *everything* with touch, we must underscore this point. The pen has a role to play as well.

But why the pen? Can't one type faster than one can handwrite? Yes, but only if our metric for creative output is in the cold calculus of words-per-minute. What is it that you wish to write? Are you making high-level comments on a manuscript? If so, composing your thoughts is likely to devolve into minutiae with a keyboard, whereas with a pen, brief annotations in context implicitly emphasize the important points. Likewise, if a pen is a poor choice to compose a business memo, then a keyboard is an equally poor choice to generate a breadth of design sketches [4,16]. That one form of work output is often valued more than the other in professional life is a deeper reflection on our society than it is on the effectiveness of the pen as an input device.

The transition to direct input is manifest in form factors ranging from hand-helds, slates, desktops, table-tops, and wall-mounted devices. The iPhone, Tablet PC, Wacom Cintiq, Microsoft Surface, and Smartboard are, respectively, examples of each. These examples include both pen and touch devices, but seldom does the same system support both. Even more seldom can one use both together [2,18,19], as with a stylus in the preferred hand and touch with the nonpreferred hand *(Fig. 1)*. Here, we pursue pen and touch as complementary rather than competitive inputs.

Our research is based on the premise that pen+touch systems present new challenges and opportunities for the designer. Our hypothesis is that the combination of pen and touch yields a richer design space of natural gestures than multi-touch input alone. When a system does not have to provide coverage of all possible functions with a single input modality, implicitly this leads one to ask where each modality should be used to best advantage, where a particular modality should not be used, and in what cases modalities should perhaps be treated interchangeably. To explore these issues, we prototyped a digital drafting table on the Microsoft Surface, using multi-touch and an IR-emitting pen. We developed an application for note-taking and mark-up that supports the key functions of writing, annotation, selection, copying, arrangement, and aggregation of objects [9,12,18].

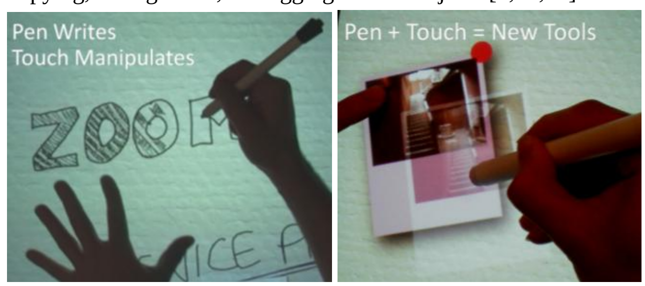

Figure 1. The roles of pen, touch, and multimodal pen+touch.

An earlier generation of devices, such as the Palm Pilot (1996), supported both pen and touch. Users could punch the on-screen calculator with their fingers, or enter Graffiti script with the stylus. Clear lessons were that the best input modality depended on the task, and that this made a significant difference to the user. However, these devices were not pen AND touch, but rather pen exclusive-or (XOR) touch. They sensed only a single point of contact, and could not distinguish touch from pen inputs. Hence we lost an opportunity for meaningful exploration of multimodal interface approaches that combine pen and touch. But a new generation of digitizers is now emerging [7] that can sense multi-touch inputs while simultaneously distinguishing pen from touch.

Why is any of this important and not just a technological quibble? The answer lies not in technology, but in the human mechanism itself, how we are wired, and how our motor, sensory, and cognitive skills have evolved. These are the underpinnings of a natural user experience, not any particular technology or device.

We have multiple fingers for a reason. We do not just point, but we also grasp and manipulate. Furthermore, our nonpreferred hand is not a poor approximation of our preferred hand; rather, it is as skilled at the specialized role that it performs as the preferred hand is at its own role [8]. For a wide class of everyday actions, our hands have evolved to complement one another. People are also predisposed to manipulate physical objects and employ manual tools. Once again, handedness plays an important role. As a simple example, when writing, we hold the pencil in our preferred hand and manipulate the paper with our nonpreferred hand. If we translate this example to a computer screen, we might write on a tablet, electronic whiteboard, or desk with a stylus, and directly manipulate the underlying virtual document, map, or photo with our nonpreferred hand using touch input.

The leap of faith we ask, and believe is justified, is to assume that the richness of such examples that exist in the physical world is matched by analogous transactions in the digital domain. By building on human behaviors and perceptual mechanisms, a foundation of physically-grounded interactions enables natural, engaging, and novel non-physical interactions to be designed. It is the implications of this leap that motivate our research, and the purpose of this paper is to share the insights that we have gained.

2. Asymmetric Division of Labor

Let's proceed by pushing a bit harder on our pencil and paper example, by asking you to consider the following question: *Which hand do you write with, your right or your left?* Now, whether you answer "right" or "left," you are wrong. The answer is "Yes!" This is not a trick question. Rather the question is ill-posed. People write with both hands, as demonstrated by Guiard:

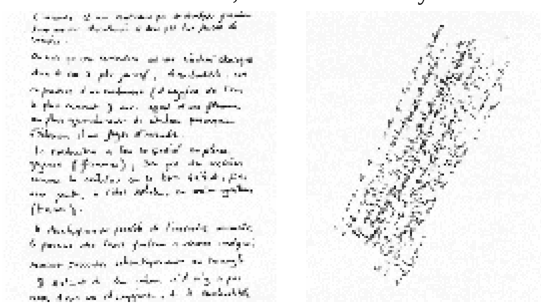

Figure 2. Guiard – transfer paper experiment [8]

What the above figure shows is the result of taking dictation on a sheet of paper. But on the right, we see the impressions left by the pen on a sheet of transfer paper surreptitiously left underneath. That is, it records the movements of the pen relative to the desk. This reveals that the nonpreferred hand sets the *frame of reference* for the action of the preferred hand; the nonpreferred hand repeatedly repositions and reorients the page so as to optimize the working space of the preferred hand [8]. This further implies that the *nonpreferred hand precedes the preferred hand* in its action.

Guiard's key insight was to turn the classic question asked in the study of handedness upside-down. Rather than asking which hand was best for a task– right or left– Guiard observed that most, if not all, manual interactions fundamentally involve both hands, with a differentiation of the roles between the hands. The correct question to ask then becomes: "What is the logic of the division of labor between the preferred and nonpreferred hands?"

Likewise, if in interface design we find ourselves asking which is best– touch or pen– then once again we must recognize an ill-posed question. The question is not which is best, but rather, *What should be the division of labor between pen and touch in interface design?* To begin to answer this question, we must consider the design properties of pen and touch as input modalities.

3. Properties Shared by Pen and Touch

We stated above that every input modality is best for something and worst for something else. Ultimately it is the designer's job to know what to use when, for whom, for what, and why. From a technology standpoint much of this turns on a nuanced understanding of the properties of an input modality. To offer insight into the main issues, the following tableau summarizes interaction properties shared by pen and touch. We do not characterize these properties as "pros" and "cons," as has been attempted elsewhere [2], to accentuate our belief that almost any property of a device can be advantageous in interaction design.

This limited survey shows that pen and touch, while sharing common ground as direct input modalities, also exhibit many important differences, and these again differ substantially from the properties of indirect pointing devices such as the mouse. Indeed, this calls into serious question the commonplace strategy of operating systems to treat all pointing devices as "mice"– that is, interchangeable "virtual devices" [5]. Consider yourself, armed with this tableau, as licensed to fire on the spot anyone in your organization who refers to pen and touch inputs as "the mouse"– or at least to deliver a well-deserved tongue-lashing.

PROPERTY	**PEN**	**TOUCH**
Contacts	**1 point** *A single well-defined point.*	**1-10+ contact regions** *with shape information [6].*
Occlusion	**Small (pen tip)** *But hand still occludes screen.*	**Moderate** *("fat finger" [17]) -* **Large** *(pinch, palm, whole hand)*
Precision	**High** *Tripod grip / lever arm affords precision, writing, sketching.*	**Moderate** *Nominal target width for rapid pointing is ~ 15 mm [17]).*
Hand	**Preferred hand**	**Either hand / Both hands**
Elementary Inputs	**Tap, Drag, Draw Path**	**Tap, Hold, Drag Finger, Pinch**
Inter-mediary	**Mechanical Intermediary** *Takes time to unsheathe the pen. Pen can be forgotten.*	**None: Bare-Handed Input** *Nothing to unsheathe, nothing to lose. No lever arm.*
Acquisition Time	**High** *(first use: unsheathe pen)* **Moderate** *on subsequent uses: pen tucked between fingers.*	**Low** *No mechanical intermediary to acquire.*
Buttons	**Barrel Button** *(some pens)*	**None**
Activation Force	**Non-Zero** *Tip switch/ minimum pressure.*	**Zero (capacitive touch)** *Resistive touch requires force.*
False Positive Inputs	**Palm Rejection** (while writing) *Palm triggers accidental inputs, fingers drag on screen, etc.*	**"Midas Touch Problem"** *Fingers brush screen, finger on screen while holding device, etc.*

Figure 3. Tableau of design properties for pen and touch.

4. Graceful Degradation

We now consider *stationary* versus *mobile* usage contexts. Desktop, table, and wall displays are necessarily stationary, but form-factors such as slates transition between mobile and stationary use. To design a consistent user experience spanning all of these form factors, we seek a conceptual model that supports *graceful degradation* between stationary and mobile usage. For the latter the nonpreferred hand is largely occupied by holding the device itself, whereas for the former we wish to support efficient bimanual interactions that leverage the full potential of human hands, as well as simultaneous pen + touch input.

For example, with physical notebooks we have observed that people deftly *tuck the pen between the fingers of the preferred hand* while flipping pages or grasping scraps of paper [11]. Hence, users can effectively perform multi-touch gestures, such as pinching, while holding the pen tucked between the fingers, and thereby derive significant value even from unimanual interactions that interleave pen and touch inputs as needed. It is important to observe here that a mobile usage model, which *assigns core operations to unimanual touch with the preferred hand*, also serves a stationary usage model that instead assigns these tasks to *touch with the nonpreferred hand.* Bimanual pen + touch gestures can then be articulated in cooperation with the preferred hand to support more efficient interaction as well as advanced gestures.

5. Recognition and Modes

The next distinction we draw is that of *ink* vs. *command* input. The specter of recognition arises as soon as one contemplates marking a virtual sheet of paper. Does drawing a mark leave an ink stroke, is it immediately converted to text, or is it perhaps recognized as a command, such as a gesture to make copies, move objects, or turn the page? Ascribing intent to the motions of an input device is a fundamental problem. People often seem to assume that recognition can overcome this problem. In our view, it does not and will not. But let's back up a moment. Who is it that must do the recognition, and why? Rarely does a user say *"I wish this sheet of paper could understand what is written on it."* Notes in a notebook are for oneself. Annotations on a document are offered as feedback to another person. Significant value arises from experiences where it is a human who recognizes the marks.

Let's say that we do wish to recognize some strokes as gestures. Implicit in this statement is the need to distinguish a *command mode* for gestures as distinct from *ink mode* for leaving marks on the digital paper. Holding a button on the pen, or tapping on a lasso-selection icon, for example, are classic ways of *mode switching* between ink and commands in pen interfaces [15]. One often hears that "modes are bad," but modes are necessary to provide rich interfaces [10] that don't depend on the success of brittle recognition techniques. The key is to rapidly switch modes in a manner that is minimally demanding of the user's cognitive resources. Here, pen+touch has much to offer.

If we assign pen to *ink mode* and touch to *command mode*, the design then *puts the mode switch in the user's hands*. For example, in our prototype the user can jot notes with the pen, but then pinch with two fingers to zoom, swipe across the margin to flip pages, or use a single finger to drag objects such as photos. That is, when considered as unimodal inputs, the logic of the division of labor between pen and touch is that *the pen writes, and touch manipulates*. The mode switch occurs implicitly depending on whether the user interacts with pen or touch. As a desirable side-benefit, this strategy also can dispense with many ancillary interface widgets, such as toolbars stuffed with icons. This leaves more display space for the user's work, while reducing the distraction of secondary controls.

Drawing on all that has preceded, we now see how our approach falls into place along the dimensions that we have identified:

- Pen vs. touch modalities have differentiated effects in the interface. Ink mode is assigned to the pen, while multi-touch articulates commands: the pen writes, and touch manipulates.
- 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;
- These benefits are derived while leaving open the possibility of bimanual manipulations with simultaneous pen and touch.

It is in the consideration of this final point, where some of the most novel possibilities may lie, that we now turn our discussion.

6. From Elementary Inputs to Phrases

The preceding interactions that interleave pen and touch may suffice to justify further investment in pen+touch displays. However, we now consider creative ways for interaction designs to leverage *simultaneous pen and multi-touch* interactions to support new capabilities for multimodal bimanual interaction. Let's consider a typical direct-manipulation pen interface for copying an object such as a photo on a digital notebook page [12]. To copy the photo and place it at a desired position, the user must:

1. *Switch the pen from ink mode to command mode;*
2. *Select a photo by tapping or lassoing it with the pen;*
3. *Invoke* Copy *by selecting a command from a context menu associated with the selected photo;*
4. *Invoke the* Paste *command to place the copy onto the page;*
5. *Drag the copy to the desired location on the page;*
6. *Return the pen to ink mode.*

Now, let's contrast this with how our system implements a simultaneous pen+touch gesture for copying a photo. All the steps required by the canonical direct-manipulation approach can be phrased into a single pen+touch bimanual gesture as follows:

1. *Hold photo and drag off a copy with the pen (Fig. 1, right).*

Is this really just one step? Our observations of users suggest that this dedicated pen+touch gesture corresponds closely to the user's mental model of the common use case where one wants to create and place a copy of an object [14]. Hence, the gesture feels like a unitary action to the user, despite invocation of multiple input events on the devices. Consistent with Guiard [8], holding touch precedes the action of the pen, and frames the context of subsequent actions of the pen held in the preferred hand.

Not only does this approach have fewer steps, but by its very nature it encapsulates all the steps into a single gestural phrase. It is syntactically simpler and precludes many types of errors, including mode errors, that can occur with a traditional approach.

Where does the syntactical simplification come from? First, note that holding a finger on the photo *integrates object selection with the transition to gesture mode*. This combines two steps. Once the photo is held with a finger, dragging off a copy with the pen embeds three different pieces of information: the Copy command (verb), what is to be copied (direct object), and where it is to be copied to (indirect object) [14]. Finally, closure is inherent in the means used to introduce the phrase: simply releasing the nonpreferred hand from the screen returns the system to its default state (ink mode), where the pen once again writes. The muscular tension from maintaining touch on the photo is the glue that holds all of these steps together. The muscular tension also has the virtue that it provides continuous proprioceptive feedback to the user that the system is in a temporary state, or mode, where the action of the pen will be interpreted differently.

We focus on the *copy* gesture above, but our system implements many pen+touch gestures. For example, users can employ the pen to *slice* photos by holding a photo with a finger, and then crossing the photo with the pen to define a freeform cut path. Or one may draw a *straightedge* by holding a photo and stroking the pen along its edge. One may even combine these actions into compound phrases, such as by holding an object and then *slicing* along the *straightedge* thus defined (Fig. 4). This illustrates the richness of the vocabulary that users may articulate with our approach.

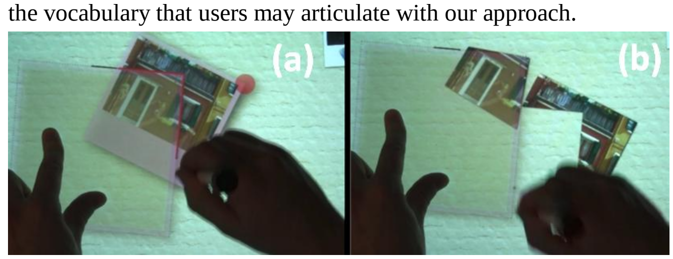

Figure 4. A pen+touch phrase: slice photo along a straightedge.

Earlier in the discussion above we stated a principle: the division of labor between pen and touch for unimodal inputs is that *the pen* *writes, and touch manipulates*. Now, we can articulate how our system interprets *multimodal* pen + touch inputs: *the combination of pen + touch yields new tools*, such as the aforementioned copy, slice, and straightedge tools. If pen+touch yields new tools, implicitly this means that in some contexts we must violate the original principle: the pen does not always write, nor does touch always manipulate. Our explorations convince us that if a system strictly limits itself so that the pen ONLY writes, and touch ONLY manipulates, this leads to a simple and consistent but artificially crippled system.

By treating multimodal pen + touch inputs differently, our system opens up a design space of new gestures that also have the virtue of leveraging how people naturally use their preferred and nonpreferred hands together. We emphasize the strengths of pen and touch as input modalities, while their use in conjunction allows us to simultaneously sidestep many of their weaknesses.

7. Incidental Contact (Palm Rejection)

Despite the advantages enumerated above, simultaneous pen and touch suffers a serious limitation in that if one rests the palm of the hand on the screen while writing, this represents a "touch" to the computer. The result may be false inputs such as accidentally panning or zooming the page. Our work partially addresses this problem, but to be clear, we do not claim to have solved it.

A simple form of palm rejection goes a long way: one just discards touches with a large contact area. However, large touches start small as the hand moves into contact with the display. Furthermore, the knuckles or side of the hand may precede the pen as it comes into contact with the display. Hence, deciding whether a touch is a true intentional manipulation is not an instantaneous binary decision, but rather is a real-time assessment that varies as the articulation of a combined pen and touch movement plays out over time.

Likewise, in reference to the tableau of Fig. 3, we must recognize that since many touch technologies require zero contact force to trigger an input, false positive inputs will remain an inherent property of multi-touch interaction, including its combination with pen input. As such, clever interaction technique designs that take advantage of this fact [3], as well as more sophisticated "accidental touch" filtering algorithms, will be integral to a rewarding pen+touch user experience. These are fundamental issues that urgently need further research.

8. Conclusion

People have multiple fingers, two hands, and highly developed skills for handling physical objects: we have shown how all of these are defining characteristics of natural pen and touch interaction. Likewise we have shown how our design carefully considers mobile vs. stationary use, ink vs. command input, and the phrasing of elementary actions into higher-level constructs that suit the user's mental model. The map of issues that we have laid out in this manuscript should help the reader to navigate through this thicket of interrelated issues and considerations.

We have advocated a division of labor between pen and touch 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. We have contributed novel pen + touch gestures, while also raising, by way of examples, design issues and questions for the reader to ponder. How should the roles of pen and touch be differentiated (or not) in your own user interface designs? The answers may differ for users of your system, but the design issues we have identified here will arise again and again.

Widespread enthusiasm for multi-touch interfaces belies an oft-overlooked truth: without careful design and a deep understanding of the strengths and weaknesses of touch as an interaction modality, a natural interface a touch-screen does not make. It has to be kept in mind that there is a difference between an *input technology* and either an *interaction technique* or a *conceptual model*– much less a natural *user experience*. Hence, touch and pen input technologies only lead to a natural experience when lots of hard work meets a thorough and nuanced understanding of these modalities, their strengths and weaknesses, when to use them, and when not to use them. Our goal here has been to impart a sense of these issues, as well as to provide example techniques that illuminate the design space. Our hope is that this can help to spur the further excitement and investment necessary for the emerging area of pen + touch input to flourish as the future of displays.

9. References

[1] Adler, A., Gujar, A., Harrison, B.L., O'Hara, K., Sellen, A. *A diary study of work-related reading design implications for digital reading devices*. CHI'98.

[2] Brandl, P., Forlines, C., Wigdor, D., Haller, M. and Shen, C. *Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces*. AVI '08 *Conf. on Advanced Visual Interfaces*. p. 154-61.

[3] Brandl, P., Leitner, J., Seifried, T., Haller, M., Doray, B., To, P. *Occlusion-aware menu design for digital tabletops*. CHI 2009 Extended Abstracts.

[4] Buxton, B., *Sketching User Experiences: Getting the Design Right and the Right Design*. 2007, San Francisco: Morgan Kaufman.

[5] Buxton, W., *Touch, Gesture, and Marking*, in *Readings in Human-Computer Interaction: Toward the Year 2000*, R. Baecker, et al., Editors. 1995, Morgan Kaufmann Publishers. p. 469-82.

[6] Cao, X., Wilson, A.D., Balakrishnan, R., Hinckley, K. and Hudson, S.E. *ShapeTouch: Leveraging contact shape on interactive surfaces*. IEEE *TABLETOP 2008* International Workshop on Horizontal Interactive Human Computer Systems p. 129-36.

[7] 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.

[8] 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.

[9] Hinckley, K., Dixon, M., Sarin, R., Guimbretiere, F. and Balakrishnan, R. *Codex: a dual screen tablet computer*. CHI '09. p. 1933-42.

[10] Hinckley, K., Guimbretiere, F., Baudisch, P., Sarin, R., Agrawala, M., Cutrell, E. *The Springboard: Multiple Modes in One Spring-Loaded Control*. CHI '06.

[11] Hinckley, K., Yatani, K., Pahud, M., Coddington, N., Rodenhouse, J., Wilson, A., Benko, H. and Buxton, B. *Manual Deskterity : An Exploration of Simultaneous Pen + Touch Direct Input*. alt.chi, CHI 2010 Extended Abstracts (non-archival publication).

[12] Hinckley, K., Zhao, S., Sarin, R., Baudisch, P., Cutrell, E., Shilman, M., Tan, D. *InkSeine: In Situ Search for Active Note Taking*. CHI 2007.

[13] Kidd, A. *The marks are on the knowledge worker*. CHI '94. p. 186-91.

[14] Kurtenbach, G. and Buxton, W. *Issues in Combining Marking and Direct Manipulation Techniques*. UIST'91. p. 137-44.

[15] Li, Y., Hinckley, K., Guan, Z., Landay, J. *Experimental Analysis of Mode Switching Techniques in Pen-based User Interfaces*. CHI 2005. p.461-70.

[16] Oviatt, S., Cohen, A.O., Mann, A.J., *Designing Interfaces that Stimulate Ideational Super-fluency in Science*, Communications of the ACM.

[17] Vogel, D. and Baudisch, P. *Shift: A Technique for Operating Pen-Based Interfaces Using Touch*. CHI 2007.

[18] Wu, M., Shen, C., Ryall, K., Forlines, C. and Balakrishnan, R., *Gesture Registration, Relaxation, and Reuse for Multi-Point Direct-Touch Surfaces*, in *IEEE International Workshop on Horizontal Interactive Human-Computer Systems*. 2006. p. 185-92.

[19] Yee, K.-P. *Two-Handed Interaction on a Tablet Display*. CHI 2004.