Pen + Mid-Air: an exploration of mid-air gestures to complement pen input on tablets

Ilhan Aslan, Ida Buchwald, Philipp Koytek, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Aslan, Ilhan, Ida Buchwald, Philipp Koytek, and Elisabeth André. 2016. "Pen + Mid-Air: an exploration of mid-air gestures to complement pen input on tablets." In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction - NordiCHI '16, Gothenburg, Sweden, October 23 - 27, 2016*, edited by Staffan Björk, Eva Eriksson, Morten Fjeld, Susanne Bødker, Wolmet Barendregt, and Mohammad Obaid, 1. New York, NY: ACM Press. https://doi.org/10.1145/2971485.2971511.

Pen + Mid-Air: An Exploration of Mid-Air Gestures to Complement Pen Input on Tablets

Ilhan Aslan[1,2], **Ida Buchwald**[1], **Philipp Koytek**[1], **and Elisabeth André**[1]
[1]Human-Centered Multimedia Lab, University of Augsburg, Germany
[2] Center for HCI, University of Salzburg, Austria
lastname@hcm-lab.de

ABSTRACT
In this paper, we report on a series of studies, exploring the potential of pen and mid-air input on tablets. We describe a field study with an early prototype of a drawing application and follow-up inquiries, such as the development and comparison of gesture sets for pen and mid-air input and pen and multi-touch input with users in a lab environment. Overall, our results suggest that pen and mid-air input should be offered to complement traditional pen and multi-touch input on tablets. We illustrate the final user-defined gestures set for pen and mid-air input and discuss how user preferences of mid-air gestures over touch gestures seem to depend on the complexity of operations and the need of additional menus on the screen.

Author Keywords
Bi-manual; mid-air gestures; multi-touch; participatory design.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
Technologies, such as Apple's iPad Pro or Microsoft's Surface have increased general awareness and interest for pen input on touch screens. Even at a time when touch screens did not allow as precise pen input as it is possible on today's high-end technologies, pen input has always been a consideration for touch screen technology. Thus, a rich body of related work exists on the study of pen and (multi-) touch input, including combined usage of pen and touch (e.g., [17, 24, 34]). Recent work on improving touch input highlights additional benefits of combining touch input with mid-air gestures (e.g., [2, 3, 7, 10]) to address constraints, such as performance issues with touch-screen interaction in cars and touch target acquisition on small-sized screens.

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in:
NordiCHI '16, October 23-27, 2016, Gothenburg, Sweden

ISBN 978-1-4503-4763-1
DOI: http://dx.doi.org/10.1145/2971485.2971511

We aim to contribute to the existing body of work in bi-manual pen and gesture input on tablets by exploring the potential of mid-air gestures that are performed with the non-preferred hand in addition to pen usage with the preferred hand.

Many of today's touch screens already support pen and multi-touch input. Whilst on large touch surfaces the benefits of bi-manual input seem obvious, on smaller touch screens, such as tablet computers benefits of gestures performed with the non-preferred hand in addition to pen input are less clear. We argue that potential issues with tablet screens, such as screen occlusion during performing touch gestures with the non-preferred hand in parallel to pen input or undesired ergonomic constraints could be compensated if mid-air gestures were a sensible option. In our best of knowledge pen and mid-air bi-manual input on tablets have not been studied yet.

In order to fill this gap in research, first, we aim to gain initial feedback on usability and user experience of pen and mid-air input on tablets, assuming that usability and user experience of the interaction technique are promising. Second, we aim to explore for which operations on tablets real users might prefer pen and mid-air input and finally what specific gestures users would want to perform.

To address our aims we have conducted a field study with more than 200 users with an early prototype of a drawing application based on available off-the-shelf technology (i.e., an iPad and a leap motion controller). In follow-up inquiries we put aside initial technological constraints associated with the prototype and inquired in more detail operations and gestures for pen and mid-air input with users. We identified a set of basic operations inspired by four contextual inquiries(CIs) on how (digital) pens are used in practice for drawing and photo editing tasks. Then, we conducted a lab study with ten users, defining for each operation a gesture for both pen and multi-touch input, and pen and mid-air input.

The final user-defined gesture set for pen and mid-air input is based on user agreement scores, video observations, and a "think aloud" procedure [8]. Further insights were gained about when and why users prefer pen and multi-touch input over pen and mid-air input, and vice versa. In the following section, we summarize the background for pen and mid-air input and afterwards describe each step of our exploration in detail in terms of methodology applied, insights gained and conclusions made.

BACKGROUND

Most relevant is previous work in bi-manual input on touch screens via (i) pen and multi-touch, and (ii) touch and mid-air.

Pen and Multi-touch Input

For bi-manual actions Guiard [9] has proposed the *kinematic chain* model and hypothesized that left and right hands are connected links in a kinematic chain. In this connection the non-preferred hand (NP) sets the frame of reference within which the preferred (P) hand moves. NP is used for less frequent, gross, and slower actions whilst P conducts frequent, small, and quick actions. Guiard's model has served as inspiration and analytical tool for many related work in pen and multi-touch input (e.g., [17, 24, 33, 34]).

Pen and multi-touch input has been extensively studied by Ken Hinckley et al. (e.g., [17, 16, 14]). In addition to referring to the *kinematic chain* model they took inspirations from observations of how related tools, such as pens, physical paper, tablets or notebooks are used in practice. They have proposed new (digital) tools based on combinations of pen and touch input [17] and investigated the division of labor between pen and touch [17, 16]. Furthermore, sensing techniques for pen and tablet usage (e.g., [14]) have been explored to improve recognizing the context of interaction and consequently setting a foundation for future application scenarios for pen and touch input on tablets.

Matulic et al. [23] have evaluated uni- and bimodal pen and touch on tabletops, as well as how pen and touch-operated tabletops support active reading operations [24], highlighting the benefits of bi-manual interaction on performance. They have compared operations, such as annotation and rapid search on physical paper, a desktop computer application for PDF documents, and a tabletop solution utilizing bi-manual touch and pen input. For annotation and cut/copy and past actions on tabletops Wu et al. [31] present two gestures for pen and touch input. On a similar topic, Yoon et al. [34] have studied how pen and touch can be used in combination to compensate insufficient space on a tablet device for making annotations by tearing the digital paper and creating additional space. Brandl et al. [5] addressed occlusion issues on tabletops and proposed occlusion-aware menu design, which assigns pen to P and touch inputing to NP and position menus adaptively.

In sum, related work in pen and multi-touch input has studied and designed pen and touch input techniques for a variety of tasks, focusing on selection and manipulation tasks, enabling compound input tasks (e.g., marking items while at the same time navigating through the list of items). In order to inform their interaction designs researchers have often taken inspirations from existing practices with similar tools (e.g., pen and paper) and technologies (e.g., tablet usage and multi-touch gestures).

With regard to the design of pen and multi-touch input Wu et al. [31] have argued early on that the design of gestures was mostly ad-hoc in related work. Drawing on existing literature and their own observations, they proposed a set of design principles to define (i) how the beginning of gesture recognition should be defined, (ii) how gestures should be treated once they have began, and (iii) how gestures should be reused to accomplish different tasks. Whilst similar guidelines were already explored for bi-manual gestures [4, 19] they were mainly inspired by Guiard's *kinematic chain model* [9] and addressed issues in settings with a displacement between input devices (e.g., two mice) and output device (e.g., screen).

Touch and Mid-Air Input

Interactions with tabletops and interactive surfaces have been described as “natural” forms of human-computer interaction, allowing users to directly manipulate digital content bi-manual, through touch gestures (e.g., [28, 30]) or tangibles (e.g., [27]) on a large screen.

However, early on researchers in tabletop interaction design (e.g., [12, 28]) have recognized and addressed the constraints of planar interaction spaces, such as issues with grasping and manipulating 3D content on planar screens. Thus, the space above the tabletop and how it could be used to add depth and continuity to interaction with interactive tabletops has been studied. For example, Marquardt et al. [22] explored combining touch on tabletops and mid-air gestures above tabletop screens towards a continuous interaction space for tabletop interaction. They illustrate a range of new interaction techniques merging touch-screen and above screen input, but focusing on one-handed input, such as *lifting gestures* to reveal objects or adjust scale precision.

Proximity information has already been used to contextualize tabletop interaction [1], such as distinguishing and adapting to left and right hand usage. Hand or finger proximity information have also been used to adapt touch screen content, such as touch targets on notebooks and desktop settings [32], or on tablets under contextual constraints [2, 3]. Hereby, researchers have studied the effects of touch targets expanding or moving towards the approaching hand of the user. Above screen input on mobile devices and one-handed interaction has also been studied, considering finger movement before, after and between touch events (e.g., [7, 10]). As far as we know, research has not yet explored pen and mid-air input on small sized touch screens, such as tablets.

Designing Gestures

We believe that pen and mid-air input on tablets has the potential to combine benefits of bi-manual input (e.g., enabling compound selection/manipulation tasks and thus improving performance) with possibilities to use mid-air input to address contextual constraints (e.g., small sized screens) and provide flexibility (e.g., allow users to perform the same gesture on the screen or mid-air).

In comparison to multi-touch gestures, the design of mid-air gestures is still less explored. However, related work in bi-manual pen and multi-touch input provide valuable insights in terms of limitations of and potentials in bi-manual input, such as increase in performance, issues of hand occlusion, and difficulties in designing gestures. Furthermore, related work provides relevant operations (e.g., zoom, annotate) for bi-manual input on touch screens, which could also be performed through pen and mid-air input, assuming the interaction possibilities are well designed.

While in the past designers and developers designed gestures for gesture-based interfaces themselves based on their valuable expertise, today there are other options available. For example, inspired by Wobbrock's methodology [29], many researchers have in recent years designed gestures for novel interaction techniques by applying *participatory design* [26] and eliciting input directly from users. For example, gesture sets for motion gestures with mobiles [25] or bend gestures with flexible electronic displays [21] have been studied. Wobbrock's proposed methodology includes observing users during performing gestures (e.g., through video material and "think aloud" protocols) and computing user agreement scores. These scores are then used to inform and define a final gesture set with high probabilities for good usability and user acceptance.

FIELD STUDY

In this section we describe the first step in a series of inquiries we conducted to explore the potential of pen and mid-air input on tablets. The aim of this first step was to explore how users perceive pen and mid-air input in terms of usability and user experience. An opportunity was offered when the department [blinded] was invited to demonstrate their research in a public large shopping mall for three subsequent days. Everyone could attend this "science fair" and engage with demonstrators and prototypes. For this occasion, we designed a prototype system, which would allow us to gain initial insights from a large number of participants on our research goal, namely perceived usability and user experience of pen and mid-air input.

Prototype system

The hardware setup consisted of a second generation iPad, a leap motion 3D controller and a laptop computer on which we had a web server running. The leap motion controller was attached to the laptop via an USB cable and the iPad was connected to the laptop via WLAN. In addition, we used a capacitive pen, which could be used to interact with the touch screen. The 3D controller could be positioned on the desk next to the iPad, depending on handedness of the user.

A Javascript/HMTL5 application was developed that allowed combined input based on the capacitive pen and mid-air gestures performed above the 3D controller. Figure 1 presents screenshots of the prototype application. The upper part of the application screen consisted of a table with data in tabular form. Cells of the table were interactive and could be edited (i.e., highlighted or blacked out) via a combination of touch and mid-air gestures. A large part of the screen was a canvas for drawing with a pen. We included a digital knob in the lower left corner, which could be used to clear the canvas.

Interaction Technique

The main principle of the interaction design that we implemented is based on distributing touch-based input and mid-air input to two separate hands. Figure 2 presents the interaction setup, including the "connection" between left and right hand in a *kinematic chain*. In this setting the user "navigates" through possible options for the (selected) target by performing mid-air gestures with the NP and (simultaneously) localizes and confirms with P holding the pen by touching the target. Hereby, the frame set with NP has only an effect on the display, if the user confirms by selecting/touching the specific area on the screen, which might require quick, detailed, and small movements. The frame set (e.g., the color chosen with NP) was also visible on the top of the screen without P needing to touch the screen.

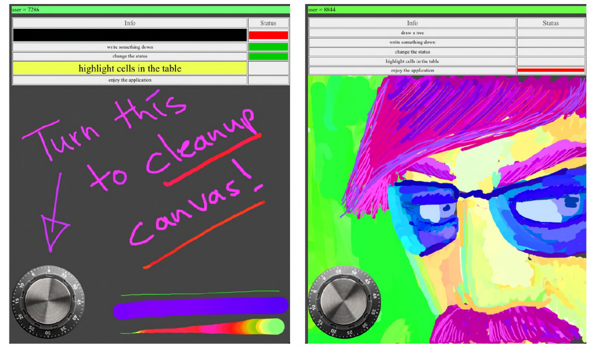

Figure 1. Exemplary screen shots of the tablet screen.

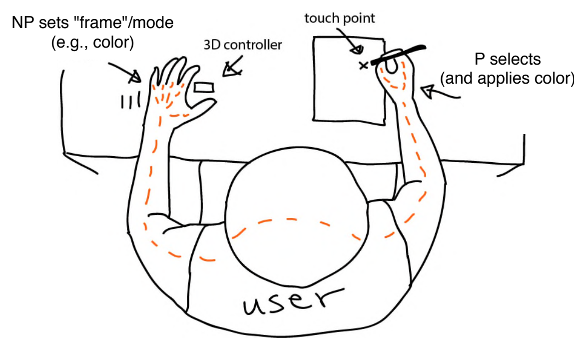

Figure 2. The interaction setup and the role of the user's hands is presented.

In order to explore a range of possibilities, the interaction technique was reused for different functions. For example, for the cells on the first column of the table NP set the brightness of the color. The color itself was fixed to yellow, allowing the user to black out cells or highlight them by setting their background to a bright yellow. For the second column NP was used to set the mode for color to either red or green, allowing users to check or un-check cells.

Users could navigate through colors rotating the palm. By changing the distance of the hand to the 3D controller users could change the size properties, such as the size of cells or thickness of ink when drawing. On the drawing canvas users could set ink color and thickness of ink with NP and draw or write with P.

Procedure of the Field Study

On a monitor that was positioned in front of the users, the types of interactions that were possible with the system were presented as part of a promotional video for the booth. Most visitors did not need further explanation on how to interact with the system. More than 200 visitors (mostly teenagers and young adults) tried out the application. Feedback was provided through questionnaires (i.e., the system usability scale (SUS) questionnaire [6], and a user experience (UX) questionnaire whose items were a subset of the AttrakDiff [11]). Our aim was to get initial insights on how participants perceived interacting based on pen and mid-air input.

The UX questionnaire was filled out by 195 visitors. The SUS questionnaire was filled out completely by 66 visitors (children below the age of 15 were not asked to fill in the SUS questionnaire, assuming that the questions were difficult to comprehend for children). In addition to the questionnaires, we observed how users interacted and occasionally chatted with visitors who showed particular interest in the system.

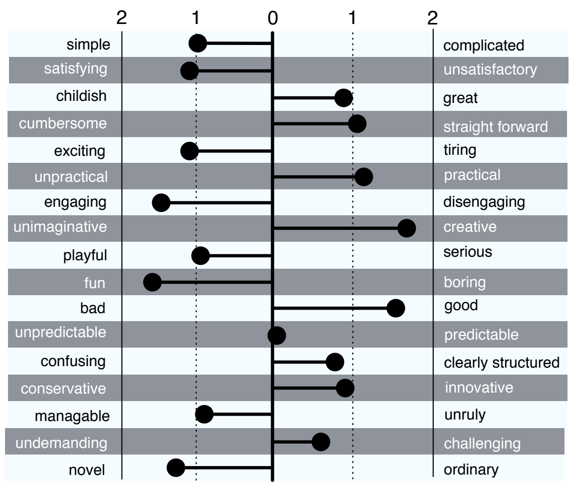

Figure 3. An overview of self-reported items is presented, which participants were asked to provide in order to describe how they perceived the interaction modality overall.

Results on Usability and UX

Figure 3 summarizes how users perceived pen and mid-air input, considering their experience when using the prototype system. The interaction technique was perceived even better than we anticipated, visitors came up with ideas for how they could use the system for themselves. Many of the visitors were teenagers, who argued that the drawing application felt like a game, and the game-like character of the application might explain why users reported the interaction technique to be playful, exiting and fun. Less surprising is that the interaction technique was perceived as novel and innovative, due to the novelty of the technology itself. Users also stated that the interaction technique was simple, straight forward, manageable, and overall good.

Although critical comments were made regarding difficulties to choose the exact color (when the range was continuous) based on mid-air gestures, the average SUS score was 70.1 on a scale from 0 to 100. The score provides a subjective assessment of overall usability. A score of 70.1 indicates reasonable usability of the interaction style with some room for improvements.

Discussion of Field Study Results

The purpose of the UX questionnaire was to get an impression of how users perceived interacting through pen and mid-air input. Many visitors who tried out the prototype spend more time than we expected with the system. Some older adults stated that it was cognitively challenging to coordinate both hands. They argued that they could definitely imagine to use the interaction technique with tabular data, however they would use it only if it was integrated in productivity applications, such as office applications.

We could observe that visitors also used the pen and mid-air input for compound actions (i.e., changing ink color and ink size with NP while at the same time drawing with P). Visitors who used the application for longer periods mainly used the mid-air input for compound actions during drawing and argued that the interaction technique made them more creative.

The positive results of the field study ensured our already existing assumption that users would like pen and mid-air as an interaction technique. However, there are some limitations of the field study, such as the simplicity of the prototype application and interaction technique that we used, as well as the choice of functions implemented in the prototype. Moreover our results are influenced by characteristics and constraints of the technologies used, such as imprecise input with pens on the second generation iPad or the recognition accuracy of the leap motion sensor (e.g., which recognized NP better if users spread their fingers).

DESIGNING GESTURES WITH USERS

In order to complement the results of the field study and to address its limitations we performed a series of follow-up inquiries. First, we conducted contextual inquiries (CIs) on pen usage, which made us understand a need to integrate users into the design process of pen and mid-air input. Then we designed gesture sets for pen and mid-air input, and pen and multi-touch input with users.

Contextual Inquiries of Pen Usage

We decided to observe real (digital) pen usage in the field, in order to inform our decisions about the final set of operations, which we would then design gestures for. To this end, we conducted four CIs and two workshops exploring the results of the observations. Whilst related work already had taken inspirations from observations, such as (physical) pen and paper usage, the purpose of the CIs was adding on the insights from related research and observing ourselves contextual constraints in (digital) pen usage.

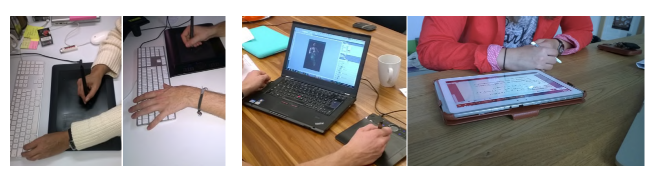

Figure 4. Exemplary material collected in during the CIs, showing the context of interaction and how professionals (see two images on the left side) position their non-preferred hands in order to use short cuts on the keyboard.

Participants and Procedure

The CIs were conducted within two weeks. Each week two CIs were conducted by two researchers followed by a brainstorming workshop with two additional researchers, analyzing the observations based on affinity diagrams. As participants we chose two software developers (1f, 1m) and two designers (1f, 1m). We visited the designers at their work places. We met the software developers who have recently completed their university studies at the university. In all cases we asked them to demonstrate how they use pen-based input and share their expertise with us.

Findings

Developer A frequently used a tablet (i.e., Samsung Galaxy Note) with a pen for three years during her university studies to make notes during lectures. She demonstrated how she used the tablet during lectures, explaining best practices and some weaknesses she has encountered over the years. We were especially interested in how she used her NP hand during making notes. She used NP sometimes to switch between functions. P was dedicated for writing and drawing, but during her demonstrations she used P also to select functions, such as the eraser function with the pen. When zooming in she used P while still holding the pen in P.

Developer B owns multiple Wacom tablets, which he mainly uses during his hobby of photo editing tasks. He demonstrated on his Wacom Bamboo how to use the tablet instead of a mouse. Zooming in and out was a frequent operation that he conducted and for which he used NP for keyboard buttons. Other shortcuts were also available on the keyboard, such as Alt+Shift to change the ink size of the pen. He used the pen button to move the canvas. He has been using Wacom tablets for many years and argued that one needs to get used to the indirect use of a pen, since pen and display are displaced.

Designer X is a graphic artist with many years of experience, who we visited in her office. She uses many applications, but has been using in particular Photoshop since 1996. Today she works mainly with a Wacom tablet, but has deactivated all buttons on the pen because of errors due to unintended use. She keeps her NP on the keyboard during her work. For some keyboard short cuts she even uses both hands on the keyboard. Five years ago she used to work with a Wacom Cintiq, which has a screen and allows direct pen input. Although one could rotate the Cintiq's display, the device itself was over all too bulky. She argued that today's tablets would have a comfortable size for many relevant tasks but that they still lack the precise pen input, which is offered by Wacom tablets without in-build screens. She told us that she prefers to work up to 20 hours apiece and during night hours in order to stay focused on the task. Her projects usually require between 20 and 40 hours of work.

Designer Y has been working for a design agency for four years and uses a Wacom tablet in several applications, including Corel Painter, Autodesk's Sketchbook Pro, and in particular Photoshop. He argued that after about two weeks, using a pen and tablet the interaction will feel as natural as using a mouse. He said that every expert sets their working station individually. He spends two to three hours on retouching tasks on one image. He uses the wacom tablet with two monitors, the monitor on the left presents the image and the monitor on the right presents the tools. He argued that horizontal hand motions (e.g., from left to right) are more precise than vertical hand motions. Thus, rotating the image is important. Sometimes he needed both hands to input keyboard shortcuts. He does not use the eraser function, since he considers turning the pen too cumbersome. Zooming in and out is an important task for which he uses keyboard shortcuts.

In sum, the contextual inquiries ensured that selection and manipulation tasks, such as zooming in and out in the canvas or rotating the canvas were relevant operations when working with a pen. Furthermore, that NP is often used for shortcuts. However and more importantly, we came to understand that requirements on pen and mid-air input needed to be flexible, allowing users to adjust pen and midair input and how they are mapped to operations to their individual preferences and needs. Once we understood that it would be more beneficial to study which kinds of pen and mid-air gestures users would want to perform themselves we focused our following efforts towards this next step of developing a gesture set for pen and mid-air input with users.

Developing a User-Defined Gesture Set

Wobrock et al. [29] have successfully demonstrated how to elicit input from users in the lab and how to design gesture sets with users. For our purpose of exploring pen and mid-air input, we have adapted their methodology of developing a gesture set with users. Instead of designing one gesture set for pen and mid-air input we asked users to design a set of gestures for pen and mid-air gesture, and another one for pen and multi-touch input. We did this to understand when users would prefer pen and mid-air input over pen and multi-touch input and vice versa.

Referent/Command	*Mean*	*SD*
Pan: Move canvas horizontal	2.0	1.73
Pan: Move canvas vertical	2.0	1.73
Increase ink size of pen	3.0	1.0
Decrease ink size of pen	3.0	1.0
Rotate canvas anti-clockwise	3.67	1.52
Rotate canvas clockwise	4.0	1.73
Change ink color of pen	3.67	0.57
Open menu	1.67	1.15
Close menu	2.0	1.73
Change brush	2.67	0.57
Undo	1.34	1.52
Move in timeline	3.34	1.52
Zoom in with focus	4.0	1.0
Zoom out	2.0	0.0
Zoom in	2.67	1.15

Table 1. The 15 commands for which participants received referents and designed gestures. The complexity of each command was rated by 3 experts on a (5 item) Likert scale. Each command was represented with a recorded video and textual description (1-2 sentences).

Referents

For 15 commands we created referents (see Table 1). Referents are effects of commands, which we described textually and represented with animations (i.e., video recordings). For example, the referent for a zoom in command shows a video recording (e.g., screen recording) of a tablet app while someone performs a pinch gesture to zoom in. Thus, the referent shows only the effect of zooming in without showing how (e.g., pinch gesture) the command was executed. In order to keep the look and feel of each referent similar we used an existing drawing application for tablets and utilized the application to create referents for all commands.

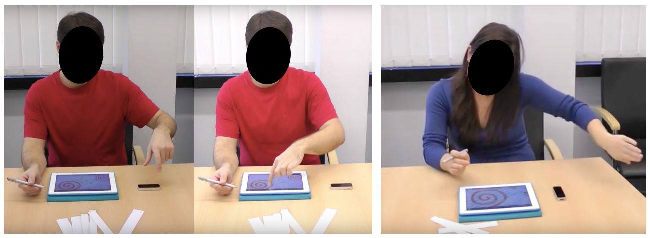

Figure 5. Screenshots of the video material showing the study setup and examplary participants exploring mid-air and multi-touch gestures.

Participants and Procedure

Ten participants (5f, 5m) took part in the study. All participants were recruited at the university campus and were either students or university staff. Two participants were left handed. All participants were accustomed to using multi-touch input on touch screens. The study was conducted in a meeting room with two researchers. One researcher provided instructions while the other researcher setup the camera and took notes. The second researchers task was to specifically note any comments made considering mental models of the participants while they explored gesture options.

After welcoming the participants, they were provided a short introduction to the study, explaining that the researchers needed their help in designing gestures for set of commands/referents. Figure 5 presents screen-shots from recordings, including the setup of the lab study.

As apparatus for the study, we used an iPad and a Leap Motion sensor. On the iPad, referents (i.e., video recordings of effects) were presented and participants could demonstrate how they envisioned multi-touch gestures and pen interaction using the tablet as a reference. The leap motion sensor was introduced as an example device which can recognize any mid-air gesture performed approximately above it but no further information was provided on real constraints of the range of the 3D sensor. We asked participants to position the leap motion sensors where ever they wanted and move it around if they wanted to. The purpose of including the leap motion device was to provide a physical reference point for mid-air gestures and our curiosity about how they would want to position such a sensor. Nevertheless, we informed participants that they could freely choose any gestures they felt fitting to the command.

The order of the referents was randomized to minimize any bias. We did this by asking participants to first pick a card from a set of cards lying on the desk in front of them, turn the card around, and read out loud the referent's description on the card. Consequently the video recording, which matches the textual description of the referent was presented on the tablet. Participants could watch the recording multiple times if needed. Afterwards they were asked to explore pen and mid-air gestures, and pen and multi-touch gestures to create the effect which was presented on the video. They were also asked to "think aloud".

Last but not least, we asked them to provide which gestures they prefer most as pen and mid-air input, and which gestures they prefer as pen and multi-touch input. Then, we asked them to rate goodness and ease of use for both gestures on a (5 item) Likert scale and to decide which modality they preferred and briefly explain their choice of preference.

All sessions were video recorded and post-hoc annotated by three researchers. Overall we collected 300 gestures (i.e., 150 pen and midair gestures and 150 pen and multi-touch gestures).

Results of Lab Study and Discussion

Effects on User Preferences

Each participant was asked to rate both input gestures (pen and multi-touch, and pen and mid-air) they proposed for a referent on two Likert scales, considering their proposition's goodness of match for the intended purpose and its ease to perform. In addition, we asked participants to rate which modality they preferred for each referent.

Analyzing the subjective ratings, we found that pen and multi-touch input was perceived as easier to use than pen and mid-air input ($\chi^2 = 14.80$, p¡0.001). Pen and multi-touch input was also perceived as a better match for the intended use than pen and mid-air input ($\chi^2 = 16.53$, p¡0.001).

	User Preferences	
Gesture Set (#)	*Pen & Mid-air*	*Pen & Multi-Touch*
All (300)	40%	60%
One-handed (240)	31%	69%
Two-handed (60)	57%	43%

Table 2. Overview of user preferences on input modality and number of hands used.

However, users preferred in only 60% over all cases the pen and multi-touch option over the pen and mid-air alternative (see Table 2). That is, in 40% of all cases the pen and mid-air alternative was preferred over the pen and multi-touch option. Of all gestures 80% were performed one-handed (either with P or NP) and 20% two-handed. For two-handed gestures, pen and mid-air input was preferred in 57% and pen and multi-touch was preferred in 43%. Users indeed preferred pen and mid-air input over pen and multi-touch when they chose to use two hands for input ($\chi^2 = 8.19$, p¡0.005).

Users proposed two-hand input for changing the ink size of the pen (#12), the brush (#12), and the ink color of the pen (#11), moving in the time-line (#7), zooming in with focus(#7), rotating the canvas (#4), zooming in/out(#4), moving the canvas horizontally/vertically(#2), and for undo(#1).

Figure 7 presents how complexity of referents (rated by experts) relate to preferences made by users, considering input

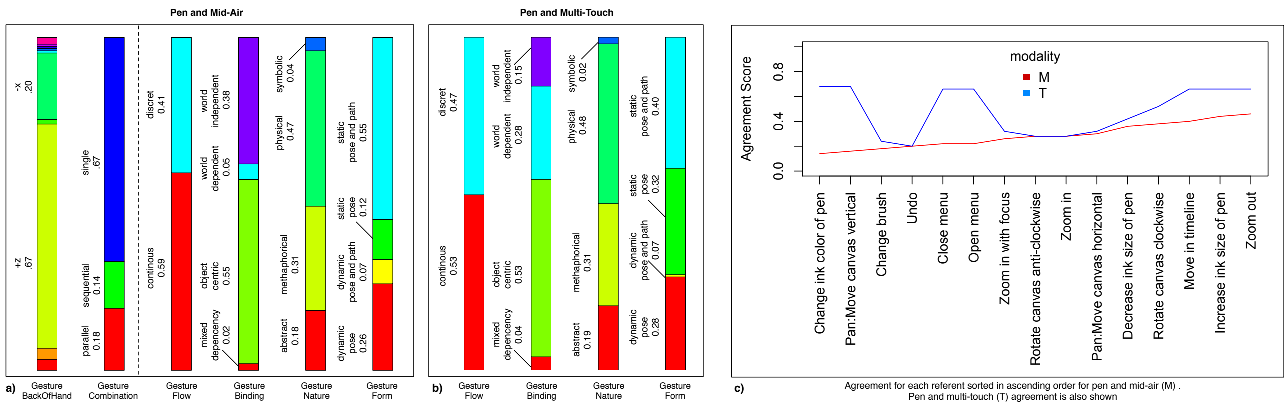

Figure 6. a) and b) present the percentage of gestures in each taxonomy for pen and mid-air gestures and pen and multi-touch gestures c) presents the overview of agreement scores for all referents and both input modalities (i.e., Pen + Mid-Air and Pen+Multi-Touch).

modality and number of hands used. Users seem to prefer pen and multi-touch input for referents with higher complexity ($\chi^2 = 32.78$, p¡0.001). On the right of Figure 7 it looks as if complexer referents are mapped to two-hands, however the statistical analysis shows no significant effect ($\chi^2 = 13.51$, p=0.06).

In sum, when users prefer two hands for input they seem to prefer pen and mid-air over pen and multi-touch. Pen and mid-air seems to be preferred over pen and multi-touch to complete simple operations.

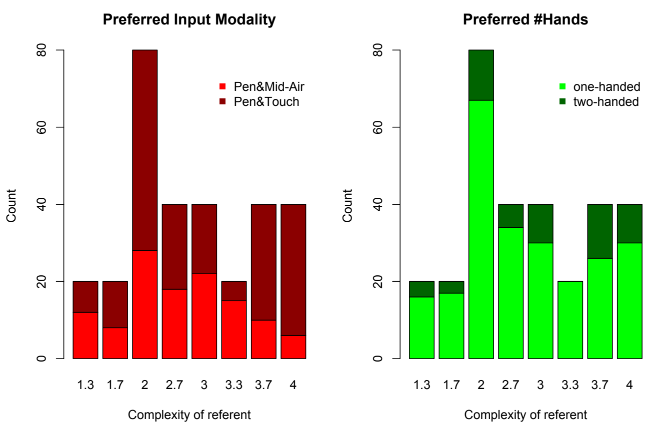

Figure 7. User preferences and the complexity ratings for referents.

Taxonomy of Pen and Mid-air Gestures

At first, we used Wobbrock et al's [29] taxonomy for surface gestures along the dimensions Form (which we applied to NP in case of two hand gestures and P if only P was used), Nature, Binding, and Flow to annotate all 300 gestures we collected. In a second iteration we added two additional dimensions, aiming to improve the categorization of mid-air hand movements during pen and mid-air input (see Figure 6a and Figure 6b).

For the second iteration we took some inspiration from Ruiz et al.'s [25] taxonomy for motion gestures for mobile interaction. First, we introduced a dimension *(Gesture Combination)* which holds the categories simple (i.e., single gestures),and serial and parallel gestures, which is a splitting of the opposite of simple gestures (i.e., compound gestures that can be decomposed into simple gestures). Furthermore, we added a second dimension *Gesture BackOfHand* to describe NP position over time during pen and mid-air input. As categories for *Gesture BackOfHand*, we used letter combinations of x, y, and z denoting the axis and with a "+" or "-" sign to denote directional information. In our setting x and y axis represented the plane spanned by the desk and the z axis the space above the desk. For example "+z" means that the back of the hand was facing upwards and away from the desk surface (and screen surface) at all times during the gesture."-x" means the back of the hand was facing sidewards away from the tablet.

Pen and Multi-Touch vs. Pen and Mid-Air Input

Figure 6a and b provide a general categorization of both input modalities (i.e., pen and mid-air, and pen and multi-touch) and a comparison between them on dimensions explaining how gestures were performed, considering physical characteristics (e.g., continuous vs discrete gestures) but also conceptual ones ((e.g., metaphoric vs. symbolic gestures)).

The data highlights a similarity between mid-air gestures and multi-touch gestures, especially along the dimensions Flow (i.e., the percentage of continuous and discrete gestures is similar for both input modalities) and Nature (i.e., both input modalities have been categorized mainly as physical and metaphorical). However, Figure 6a and b also shows, while both modalities are highly object-centric (e.g., requiring some information about the objects they produce or affect), considering Gesture Binding that multi-touch gestures are more world dependent (e.g., they need to be performed directly on the touch screen). Mid-air gestures were categorized different than multi-touch gestures in terms of how not only the gesture is often static but also the path that existed in mid-air gestures.

Figure 6c shows agreement scores for each referent and presents an overview of the consensus between participants. The agreement scores are higher when more users agree on the same gesture for the same referent. Agreement scores are computed applying the mathematical formula proposed by Wobbrock et al. [29]. The scores we report are similar to scores reported by Wobbrock for surface gestures. For example, for the referent "Zoom in" they report for surface gestures agreement scores of about 0.25 for one-hand gestures and about 0.27 for two-hand gestures.

As expected, there was more agreement for pen and multi-touch input than for pen and mid-air input. However, the high agreement scores computed for pen and multi-touch input, such as for the referents "change ink color of pen", "close/open menu", and "increase ink color of pen" result of participants using often simple tap gestures with an imaginary button on the touch screen, which they would tap to open a tools menu, for example. That is, one could argue that by telling and showing us that they would produce the effect with a tap on a non-existing button they failed to produce the effect without changing the graphical interface.

Defining a Gesture Set for Pen usage and Mid-air Input

Based on the agreement scores we have computed gesture sets for pen and multi-touch input and pen and mid-air input. That is, we assigned the largest group of identical gestures for each referent to the referent. We observed diverse pen and multi-touch gestures, such as various forms of pinch gestures (e.g., pinch with P, pinch with NP, pinch with two or more fingers, pinch with NP in combinations with setting focus on the screen with P). However, the largest group of agreement were standard pinch gestures with NP on the screen for zooming in and out, and for zooming in with focus at the touch position of the pinch gesture on the screen. Same results were achieved for twist gestures for rotating the canvas, and pan gestures to move the canvas, and tap gestures to open/close all kinds of menus. Consequently, the kinds of multi-touch gestures that users would want to perform do already exist and will be recognized by today's high-end touch devices.

Considering mid-air gestures we observed more conflicts, meaning that the same gesture was preferred for multiple operations. However, users agreed overall only a small set of gestures, such as making a fist, a flat hand or a pinch gesture. In comparison to touch gestures, which are performed directly on the target when performing mid-air gestures users applied proxemic information, such as orientation of NP towards the screen or distance of NP to the screen or an imaginary object in 3D. For zoom or rotation operations the imaginary object was often referred to as the "camera", assuming that NP represents the canvas on the tablet, which displays the camera image. Thus, users suggested zooming and rotating the canvas by moving their hand in relation to the position of the imaginary camera (which was often imagined at head height).

The set of mid-air gestures computed based on user agreements and considering any conflicts is presented on the left hand side of Figure 8. Similar to related work we observed reversible gestures for dichotomous referents and thus merged whenever possible to the proposed gestures to optimized the final set of gestures. A concern that was voiced by some users and is well known in the design of mid-air gestures is recognizing when gesture recognition starts and ends. We considered this issue when solving conflicts; that is, when we had to choose between two gestures that users equally agreed on. As mentioned above some users applied a mental model, using an anchoring mechanism either some imaginary object but also the tablet, the pen, or finger touch. Thus, some gestures are conceptually similar to what has already been proposed for bi-manual input with 3D content or for 3D modeling on and above tabletops (e.g., [18]).

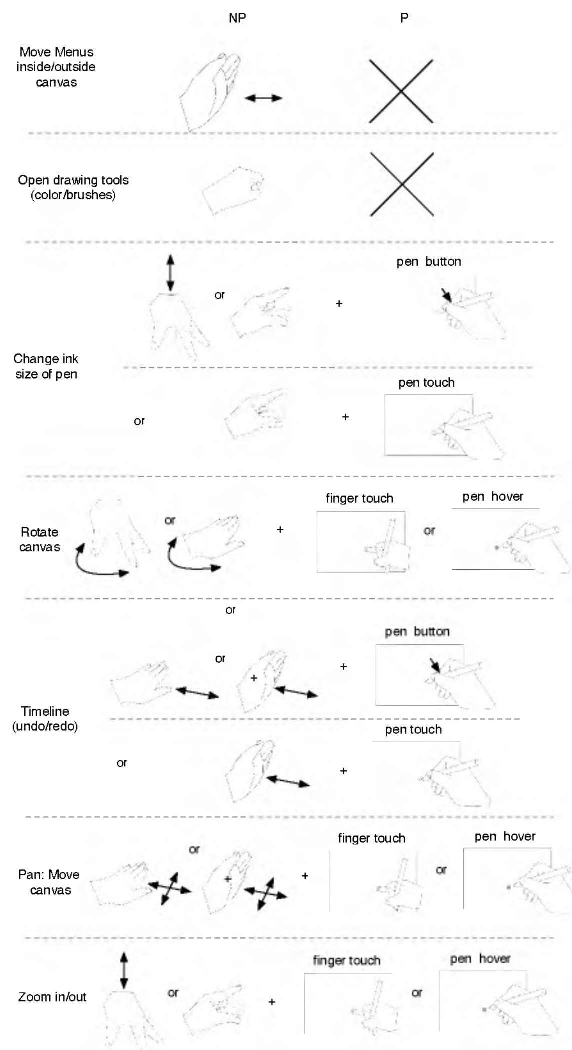

Figure 8. The mid-air gestures participants agreed on are presented on the left side. On the right are interactions with P based on pen usage or finger touch.

SUMMARY OF FINDINGS AND IMPLICATIONS

The insights we gained throughout our inquiries support our initial assumption that mid-air gestures performed with NP have potential to complement bi-manual input on tablets when P is assigned to pen usage. Our results highlight that users perceived pen and mid-air input as a usable interaction technique associated with mainly positive user experience descriptions, such as fun, exciting, playful, and engaging to use.

We believe that mid-air input offers a large space for expressive and bodily interactions, and thus when integrated in pen-based user interfaces has potential to render interaction more engaging.

In the lab study we found that users preferred mid-air gestures over multi-touch gestures to complement pen input for simple operations. Multi-touch was preferred to complete complex operations or when the operation could be completely with one hand (i.e., P). Thus, mid-air gestures with NP were preferred over multi-touch gestures with NP when users wanted to use two hands for operations. This was often the case when users argued that they needed to open a menu before making a selection (e.g., selecting a new brush). One user argued *"using a Mid-air gestures feels less disruptive ... and one can keep focusing on the task in front"*.

Related work has already argued for performance benefits of bi-manual input (e.g., [20]) and we ourselves have observed in the contextual inquiries that NP is often used to perform short cuts (e.g., use a keyboard to activate eraser tool instead of turning the pen around). It seems that mid-air gestures with NP are preferred over multi-touch gestures, because they seem to require less precision and are less world-dependent (e.g., they do not have be performed at a specific space). Pen based input often requires precision. Therefore, the benefit of quickly setting a mode, opening a menu, etc. with freedom of movement and without effort in precision seems clear.

In retrospective, the set of gestures that we presented for pen and mid-air input can be more precisely described as mid-air gestures with NP to complement pen usage with P. Hereby pen usage with P means that users may have a pen in their hand but are still capable and prefer to perform multi-touch input. For example, users may perform pinch gestures with their middle finger and thumb while still holding a pen.

Limitations and Next Steps

In our research we have not looked at pen usage behavior in detail, but an insight we gained is that during pen usage users adjust their grip on the pen to use finger touch or multi-touch while still holding the pen in their hands. Hinckley et al. [15] have studied grip behavior tablet users and proposed a taxonomy for pen grips. We believe that there is some potential to combine not only direct pen, finger touch or pen hover with mid-air input for bi-manual input, but use pen grip behavior. As proposed by Hinckley et al. [13], we also believe that there is potential to use pen motions to complement bi-manual input on touch screens; however, users in our lab study were not explicitly instructed to explore pen motions and they did not explore pen motions as an option by themselves.

Considering that users did not choose to perform complex mid-air gestures with NP and that people might already wear smartwatches or wristbands, one could, in a next step imagine to explore if the capabilities of these wearable technologies are already sufficient to realize the gesture set we have proposed.

Overall, we believe the insights we presented will benefit the design of bi-manual input for a range of devices, such as mobile phones, watches and devices with similar other displays, including large but graphically overloaded screens and screen areas. Moreover, we expect that the ongoing technological advancements in the area of touch screens and pen-based input will render the contribution of this paper even more valuable and ecologically valid.

CONCLUSION

We have reported on a series of inquiries studying the potential of pen and mid-air input as a bi-manual input modality for tablets. We started our investigations with a prototype system based on off-the-shelf technology, which we utilized in a field study for observing a large number of users and evaluating perceived usability and user experience of pen and mid-air input as an interaction technique. The positive feedback on usability and user experience confirmed our initial assumptions and motivated our follow-up inquiries. These inquiries included identifying relevant operations for pen usage on tablets and exploring what gestures users would want to perform and when they would choose pen and mid-air input over pen and multi-touch input and vice versa.

We found that when users prefer two hands for input on a tablet they seem to prefer pen and mid-air over pen and multi-touch. Furthermore, pen and mid-air seems to be preferred over pen and multi-touch to complete simple operations, providing users freedom of movement with their non-preferred hand. It also offers a way to quickly open a menu, change mode, etc. without loosing focus on the screen and actions performed with the pen and users' preferred hand. Our investigations lead to a set of gestures for pen and mid-air input, which we propose as an alternative that complements pen and multi-touch input.

ACKNOWLEDGMENTS

This research was partly funded by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development (Christian Doppler Laboratory for "Contextual Interfaces").

REFERENCES

1. Annett, M., Grossman, T., Wigdor, D., and Fitzmaurice, G. Medusa: A proximity-aware multi-touch tabletop. In *Proc. of*, UIST '11, ACM (2011), 337–346.

2. Aslan, I., Krischkowsky, A., Meschtscherjakov, A., Wuchse, M., and Tscheligi, M. A leap for touch: Proximity sensitive touch targets in cars. In *Proc. of*, AutomotiveUI '15, ACM (New York, NY, USA, 2015), 39–46.

3. Aslan, I., Meneweger, T., Fuchsberger, V., and Tscheligi, M. Sharing touch interfaces: Proximity-sensitive touch targets for tablet-mediated collaboration. In *Proc. of*, ICMI '15, ACM (2015), 279–286.

4. Balakrishnan, R., and Hinckley, K. The role of kinesthetic reference frames in two-handed input performance. In *Proc. of*, UIST '99, ACM (1999), 171–178.

5. Brandl, P., Leitner, J., Seifried, T., Haller, M., Doray, B., and To, P. Occlusion-aware menu design for digital tabletops. CHI EA '09, ACM (2009), 3223–3228.

6. Brooke, J. Sus-a quick and dirty usability scale. *Usability evaluation in industry 189*, 194 (1996), 4–7.

7. Chen, X. A., Schwarz, J., Harrison, C., Mankoff, J., and Hudson, S. E. Air+touch: Interweaving touch & in-air gestures. In *Proc. of*, UIST '14, ACM (2014), 519–525.

8. Ericsson, K. A., and Simon, H. A. Protocol analysis, 1993.

9. Guiard, Y. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior 19*, 4 (1987), 486–517.

10. Han, J., Ahn, S., and Lee, G. Transture: Continuing a touch gesture on a small screen into the air. CHI EA '15, ACM (2015), 1295–1300.

11. Hassenzahl, M., Burmester, M., and Koller, F. Attrakdiff: Ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In *Mensch & Computer 2003*. Springer, 2003.

12. Hilliges, O., Izadi, S., Wilson, A. D., Hodges, S., Garcia-Mendoza, A., and Butz, A. Interactions in the air: adding further depth to interactive tabletops. In *Proc. of*, UIST '09, ACM (2009), 139–148.

13. Hinckley, K., Chen, X. A., and Benko, H. Motion and context sensing techniques for pen computing. In *Proc. of*, GI '13, Canadian Information Processing Society (Toronto, Ont., Canada, Canada, 2013), 71–78.

14. Hinckley, K., Pahud, M., Benko, H., Irani, P., Guimbretière, F., Gavriliu, M., Chen, X. A., Matulic, F., Buxton, W., and Wilson, A. Sensing techniques for tablet+stylus interaction. In *Proc. of*, UIST '14, ACM (2014), 605–614.

15. Hinckley, K., Pahud, M., Benko, H., Irani, P., Guimbretière, F., Gavriliu, M., Chen, X. A., Matulic, F., Buxton, W., and Wilson, A. Sensing techniques for tablet+stylus interaction. In *Proc. of*, UIST '14, ACM (2014), 605–614.

16. 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. CHI EA '10, ACM (2010), 2793–2802.

17. Hinckley, K., Yatani, K., Pahud, M., Coddington, N., Rodenhouse, J., Wilson, A., Benko, H., and Buxton, B. Pen+ touch= new tools. In *Proc. of*, UIST'10, ACM (2010), 27–36.

18. Jackson, B., Schroeder, D., and Keefe, D. F. Nailing down multi-touch: anchored above the surface interaction for 3d modeling and navigation. In *Proc. of*, Canadian Information Processing Society (2012), 181–184.

19. Kabbash, P., Buxton, W., and Sellen, A. Two-handed input in a compound task. In *Proc. of*, CHI '94, ACM (1994), 417–423.

20. Klemmer, S. R., Hartmann, B., and Takayama, L. How bodies matter: Five themes for interaction design. In *Proc. of*, DIS'06 (2006).

21. Lahey, B., Girouard, A., Burleson, W., and Vertegaal, R. Paperphone: Understanding the use of bend gestures in mobile devices with flexible electronic paper displays. In *Proc. of*, CHI '11, ACM (2011), 1303–1312.

22. Marquardt, N., Jota, R., Greenberg, S., and Jorge, J. A. The continuous interaction space: interaction techniques unifying touch and gesture on and above a digital surface. In *INTERACT 2011*. Springer, 2011.

23. Matulic, F., and Norrie, M. Empirical evaluation of uni-and bimodal pen and touch interaction properties on digital tabletops. In *Proc. of*, ITS'12, ACM (2012), 143–152.

24. Matulic, F., and Norrie, M. C. Supporting active reading on pen and touch-operated tabletops. In *Proc. of*, AVI '12, ACM (2012), 612–619.

25. Ruiz, J., Li, Y., and Lank, E. User-defined motion gestures for mobile interaction. In *Proc. of*, CHI '11, ACM (2011), 197–206.

26. Schuler, D., and Namioka, A. *Participatory design: Principles and practices*. CRC Press, 1993.

27. Underkoffler, J., and Ishii, H. Urp: A luminous-tangible workbench for urban planning and design. In *Proc. of*, CHI '99, ACM (1999), 386–393.

28. Wilson, A. D., Izadi, S., Hilliges, O., Garcia-Mendoza, A., and Kirk, D. Bringing physics to the surface. In *Proc. of*, UIST '08, ACM (2008), 67–76.

29. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proc. of*, CHI '09, ACM (2009), 1083–1092.

30. Wu, M., and Balakrishnan, R. Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proc. of*, UIST '03, ACM (2003), 193–202.

31. Wu, M., Shen, C., Ryall, K., Forlines, C., and Balakrishnan, R. Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In *Horizontal Interactive Human-Computer Systems, 2006. TableTop 2006.*, IEEE (2006), 8–pp.

32. Yang, X.-D., Grossman, T., Irani, P., and Fitzmaurice, G. Touchcuts and touchzoom: Enhanced target selection for touch displays using finger proximity sensing. In *Proc. of*, CHI '11, ACM (2011), 2585–2594.

33. Yee, K.-P. Two-handed interaction on a tablet display. CHI EA '04, ACM (2004), 1493–1496.

34. Yoon, D., Chen, N., and Guimbretière, F. Texttearing: Opening white space for digital ink annotation. In *Proc. of*, UIST'13, ACM (2013), 107–112.