Replicating User-defined Gestures for Text Editing

Poorna Talkad Sukumar
ptalkads@nd.edu

Anqing Liu
liu1@alumni.nd.edu

Ronald Metoyer
rmetoyer@nd.edu

Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, USA

ABSTRACT

Although initial ideas for building intuitive and usable handwriting applications originated nearly 30 years ago, recent advances in stylus technology and handwriting recognition are now making handwriting a viable text-entry option on touchscreen devices. In this paper, we use modern methods to replicate studies from the 1980's to elicit hand-drawn gestures from users for common text-editing tasks in order to determine a *guessable* gesture set and to ascertain if the early results still apply given the ubiquity of touchscreen devices today. We analyzed 360 gestures, performed with either the finger or stylus, from 20 participants for 18 tasks on a modern tablet device. Our findings indicate that the mental model of "writing on paper" found in past literature largely holds even today, although today's users' mental model also appears to support manipulating the paper elements as opposed to annotating. In addition, users prefer using the stylus to finger touch for text editing, and we found that manipulating "white space" is complex. We present our findings as well as a stylus-based, user-defined gesture set for text editing.

Author Keywords

Tablets, handwriting, stylus, gesture, guessability, elicitation study, text editing, replication

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: User Interfaces – Interaction styles, user-centered design.

INTRODUCTION

Tablet devices are widely used but text entry and editing using the virtual keyboard on these devices can be inconvenient and inefficient [6, 19]. However, considering the recent incorporation of highly effective stylus input in devices such as the Microsoft Surface, Apple iPad Pro, and Samsung Galaxy Pro, handwriting using a pen or stylus can be a viable text-input option. Handwriting also has proven cognitive benefits over typing including better memory retention and learning [20, 22, 24, 27], which promotes the case for designing and developing effective handwriting applications for use in place of typing for any generic text entry and editing tasks such as note-taking or form-filling [30].

ISS '18, November 25–28, 2018, Tokyo, Japan

DOI: https://doi.org/10.1145/3279778.3279793

Handwriting applications on electronic devices typically benefit from aspects of writing on paper, such as the ease of use, with aspects of electronic documents, such as providing digital representations of the handwritten text that support features such as "search". Therefore, effective stylus input and accurate handwriting recognition are key to the usability of these applications. The state-of-the-art handwriting applications on tablet devices, however, are far from being effective or intuitive despite advanced stylus-input technology and high recognition rates [10]. Many applications, for example, treat the stylus as a literal replacement for the virtual keyboard, requiring that users write in the designated keyboard area at the bottom of the screen. Or, to edit text on touchscreen devices, users may have to position the cursor with a touch gesture and then use the *backspace* key on the virtual keyboard to delete the text. These implementations simply do not take advantage of the potential of stylus-based input.

Seminal handwritten text-entry and editing research from the 80s and 90s [8, 9, 13, 32, 35] was limited by the technology of the time and error-prone recognition systems. Nevertheless, the design insights they provide, such as support for writing anywhere on the document and editing gestures that "mimic" writing on paper, are very intuitive and largely applicable even today but are often missing from the interfaces of today making them ineffective as a result.

We draw inspiration from and revisit the aforementioned papers from decades past to explore the design space of text entry and editing with a stylus. We focus on devising stylus-based gestures for text editing, with an emphasis on whitespace use and specification, which are key components in the realization of usable handwriting-as-text input applications. We performed an elicitation study almost identical to the ones described in the papers by Welbourn and Whitrow [32] and Wolf and Morrel-Samuels [35] to obtain intuitive gestures from participants. who have much more experience with touchscreens and digital markup than those thirty years ago. By replicating the elicitation studies, we also wanted to see if the present-day users' profound familiarity with these devices leads them to propose different gestures or if they echo the mental models of the participants in the earlier studies.

We were interested in the following questions:

- What gestures do participants choose for typical text-editing tasks and to what degree do participants agree on those gestures?

- Do participants prefer to use finger or the stylus for text editing?
- Where do participants prefer to write?
- What mental models (e.g., paper analogies, or touchscreen interaction) influence the participants' gestures?
- How do participants manipulate whitespace with a stylus?

We mainly focus on defining a gesture-set for text editing and do not consider *recognition* of the gestures. In the following sections, we discuss related work, describe our study, and present results followed by a discussion of our findings and implications for design of stylus-based text entry and editing.

RELATED WORK

Early work on handwriting interfaces

Notable work was done on handwritten text entry and editing on pen-based computers in the 80s and 90s [8, 9, 13, 32, 35]. Welbourn and Whitrow [32] and Wolf and Morrel-Samuels [35] describe studies to elicit gestures for various text-editing operations. The gestures mostly mimicked editing on paper. While the Wolf and Morrel-Samuels [35] study was pen- and paper-based, the Welbourn and Whitrow study consisted of a combination of a tablet for stylus use and a separate display to view the writing [32]. The text-editing tasks considered in these papers included deleting character/word/phrase, inserting character/word/phrase/space, and moving word/phrase. We borrowed largely from these two papers in terms of the general goal of obtaining user-defined gestures for handwriting interfaces and the text-editing tasks considered but we broadened the ideas and adapted them for today's devices.

Prior work also suggests potentially useful and intuitive features to incorporate in handwriting interfaces. Goodisman and Goldberg [8, 9] introduced the concept of "markup editing" which refers to the recognition of editing gestures but not applying those edits immediately to the text. It is especially useful when more than one person is editing the text enabling others to see the edits one intends to perform. Similarly, Kato and Nakagawa [13] proposed the concept of "lazy recognition" wherein the recognition of both the writing and editing gestures are delayed until a later time so that the creativity of the writer is not affected.

In the early 2000s, efforts were made to combine the advantages afforded by paper and digital documents wherein machine-readable gestures were performed on printed documents which were executed in their corresponding digital forms [2, 11, 17].

Pen-based interactions

Other work on general pen-based applications list their characteristics, existing and potential applications, perceptions of users towards them and compare the use of stylus over the finger or mouse [3, 4, 18, 30]. Long et al. [18] compared, in addition to other metrics, the use of text-editing gestures on the Apple Newton and PalmPilot. They found that Newton users gestured more often than Pilot users and that the difficulty in remembering Pilot gestures was the main reason participants used them less frequently. Cockburn et al. [4] found that the stylus was slower than finger pointing for tapping tasks but more accurate, especially with small targets, and the stylus was faster than finger touch for dragging tasks.

Hinckley et al. [12] studied how people use physical paper and notebooks and based on their observations, presented a design for a Microsoft Surface application where pen is used for writing, touch for manipulating, and combinations of pen and touch inputs enable new functionality. These interactions also alleviate the need for physical buttons or widgets typically used on touchscreen devices. Riche et al. [28] compared the uses of analog and digital pens through elaborate diary studies and a survey. They present promising affordances of digital pens not provided by analog pens including "dynamic editing" of digital text and efficient storage and retrieval of (recognized) handwritten notes.

Since we are interested in comparing finger and stylus interactions in the context of handwriting applications, the above findings can be useful in interpreting our results.

In situ handwriting interfaces

Recently, Gu et al. [10] presented *In-Place-Ink* driven by motivations similar to ours - building more "direct" handwriting applications where users can write anywhere in contrast to the current applications where users write in separate boxes. *Tableur* [37], a recent handwriting "spreadsheet" application, affords general spreadsheet functionality that is invoked through stylus-based gestures. All the handwritten content is recognized, retained, and edited in digital ink format.

Touch- and stylus-based gestures for text editing

Although there are many papers on text entry on mobile devices, such as augmenting the virtual keyboard or providing alternative keyboard designs, very little work is found on touch-based gestures for text editing on these devices. Fuccella et al. [6] summarize the typical user behaviors involved in text editing on today's touchscreen devices and they address the imprecision, unwieldiness, and twofold actions, i.e. point and pick from menu, involved in the conventional "widget-based" text editing on touchscreen devices. *RefactorPad* [26] focuses on enabling programming on mobile devices and in addition to eliciting pen and touch gestures for refactoring tasks, gestures are also elicited for common text-editing tasks. We will also examine the use of touch vs. stylus-based gestures in the context of general text editing.

In *TextTearing* [36], touch- and pen-based gestures are presented for dynamically creating rows of whitespace anywhere in digital documents for the purpose of entering annotations. We also discuss dynamically creating whitespace, but for insert-text tasks, that adjusts to user needs.

Our work is similar to concurrent work done by Costagliola et al. [5] who also asked users to devise stylus-based gestures for text-editing tasks. However, we present a more elaborate gesture-elicitation study that includes measures of *guessability* and is designed to gauge if the observed behaviors of users are influenced by their use of today's mobile devices. Furthermore, we provide detailed discussions of the observed behaviors,

Instruction
Line 4 reads: 4. it was the season of Light, it was the of Darkness, Please insert the word ***season*** so that it reads: 4. it was the season of Light, it was the ***season*** of Darkness, Please indicate what gesture(s) you would use and how would the application respond?

Figure 1: Instructions for a sample ('Insert word' task) referent

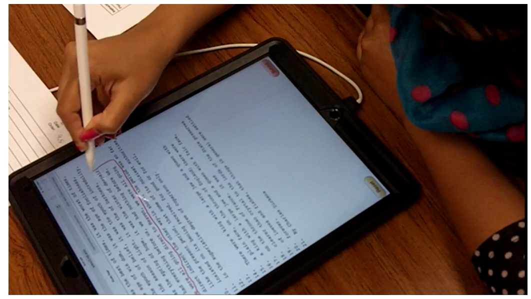

Figure 2: A participant performing the 'Select multiple lines' task during one of the study sessions.

such as user mental models and legacy bias. While Costagliola et al. use the gestures in combination with text entered through the soft keyboard [5], we present a vision for an "in situ" handwriting interface where we intend for the gestures to be used seamlessly with handwriting text. We further mention the differences in our respective studies and elicited gestures in the discussion section.

GESTURE-ELICITATION STUDY

Our first goal was to determine if the findings from nearly 30 years ago still hold with respect to the preferred gestures for text entry and editing on today's touchscreen devices. We thus conducted a study using a modern approach similar to that described by Wobbrock et al. [34] to elicit the most natural gestures for participants for common text-editing tasks.

Study Design

Our within-subjects elicitation study provided participants with referents, or examples of the desired output, and asked them to perform the gesture that would result in that referent. Each referent consisted of a passage of text before and after a particular text-editing task was carried out. For example, the referent for the "Insert Word" task is shown in Figure 1. We asked users to provide gestures for 20 such referents that were organized into five groups based on the type of operation:

- Deletion: delete character, word, phrase, space, long phrase, blank line
- Insertion: insert character, word, space, phrase, blank line, tab/indent
- Move: select and move word
- Cut/Copy/Paste: cut and paste phrase, copy and paste phrase
- Selection: select character, word, phrase, long phrase, multiple lines

The apparatus consisted of an Apple iPad Pro and an Apple Pencil. We developed a low-fidelity prototype application to present digital typed text to the user and capture user gestures intended to edit that text. The prototype presented the text on the touchscreen and displayed user strokes as they made gestures on the screen with either the Apple Pencil or a finger. We instrumented the prototype to capture each stroke (rendered using different colored ink for stylus and finger inputs) and tag them with the current task number for later analysis.

We first conducted a pilot study with two participants and based on the responses obtained, we further refined the instructions and study design.

Participants

A total of 20 participants volunteered for the study over a two-week span. There were 9 female participants and only one participant reported being left-handed. All participants were students from the University of Notre Dame. The average participant age was 24 and 15 participants were computer science majors. On a Likert scale from 1 (never) to 5 (always), participants rated themselves as frequent users of touchscreen devices (mean = 4.75). However, most of the participants rarely used a stylus on a touchscreen device (mean = 1.8). This distinction was desirable in order to understand potential effects of mobile touchscreen experience without the influence of current, often impoverished, stylus interaction on such devices. Participants also reported that they often used touchscreens for text entry and editing before (mean = 3.85), using the virtual keyboard as the method of input. 65% of the participants said they still write on paper with a pen or pencil regularly.

Procedure

Each session began with completion of an informed consent document. Participants then completed a pre-study questionnaire to collect demographic and individual difference information with respect to experience using and writing on touchscreen devices. The experimenter then described the study process and administered a practice task to the participant. We gave participants the opportunity to ask any questions during and upon completion of the practice task. For all tasks, participants were seated at a desk and the mobile device was placed on the desk.

We presented participants with a randomized set of 20 tasks, one at a time. For each task, we presented the referent on paper, asked the participant to read it, and instructed them to begin the task immediately after reading the referent by pressing a "start" button on the iPad. This allowed us to capture the time spent thinking about the most appropriate gesture, i.e. planning time, in addition to the time spent in performing the gesture, i.e. articulation time. Planning time was computed as the time between seeing the referent and starting the gesture.

The participants were instructed to use *either* their finger or the pen to draw the most natural gesture for the given referent. Given that our test interface was static and didn't respond to their input, we told the participants to assume that the implemented interface would be capable of responding to their gestures as they intended. Since previous work suggests that participants may factor in the recognition capabilities of the device when creating gestures [23], we told the participants to assume that the implemented interface would afford writing and recognition anywhere on the document and also that it would not require the use of the virtual keyboard. We encouraged the participants to think aloud as they completed each task. In addition, for gestures that were ambiguous or included multi-part interactions, we asked the participants how they would expect the system to respond to their gesture. The participant responses helped us understand their mental models which we elaborate on in the discussion section.

Following Wobbrock et al.'s approach [34], we obtained satisfaction responses from the participants on 7-point Likert scales after each task on how easy the gesture was to devise and to perform and whether it was a good match for the referent. Upon completion of the study, we compensated each participant with a $10-gift card.

Participants could use identical gestures for referents within the same group but were not allowed to reuse gestures among the different groups. If a participant were to re-use a gesture for referents belonging to different groups, the experimenter would bring it to their attention and ask them to disambiguate the two gestures. We videotaped all sessions for later reference and analysis. The same experimenter ran all the study sessions to maintain consistency and prevent potential experimenter biases [15]. A participant performing a sample task during the study is shown in Figure 2.

ANALYSIS AND RESULTS

Since we aim to determine a gesture set for common text-editing tasks, we use the *guessability* methodology formulated by Wobbrock et al. [33] to both devise a *guessable* gesture-set and to assess its guessability. They define guessability in symbolic input as:

> *"That quality of symbols which allows a user to access intended referents via those symbols despite a lack of knowledge of those symbols."*

We begin by grouping the participants' gestures and then apply the guessability methodology to the groups to define the final gesture set.

Gesture Grouping

We collected a total of 400 gestures (20 referents from 20 participants). For some referents, the gestures were quite unanimous, having only a few candidates. However, some referents generated more than ten different candidate gestures. In some cases, the set of gestures was unique in appearance, but similar in semantics. For example, for the task of selecting a phrase, some participants chose to mark the beginning and end of the phrase while others circled the entire phrase as shown in Figure 3. While one variant of the gesture marking

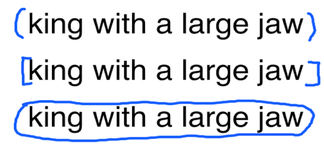

Figure 3: Three sample gestures used for the 'Select phrase' task. The first two gestures, although different in form, are considered semantically equivalent since they both involve marking the beginning and end of the phrase. The third gesture, involving circling or bounding the phrase, is considered semantically different from the first two.

the beginning and end of the phrase used brackets " [ ] ", another variant used parentheses " ( ) ". These two options, while not identical, were semantically equivalent.

To classify the gestures, two members of the research team conducted an open-coding [29] process on all 400 gestures and developed a set of codes to describe all possible semantic classifications of participant gestures for each referent. For each referent, the researchers entered the codes into a codebook with an example and a description of the gesture. In some cases, the gestures were hierarchical in nature and were coded using a high-level and low-level code. For example, for the "Delete word" referent, we based the high-level categorization on whether the word was deleted directly or selected first and then deleted. Following this categorization, the next level of grouping involved the specific symbol or gesture that was used to mark the word for deletion; these symbols included strikethrough using single or multiple lines and a cross.

We were unable to come up with coherent classifications of the participant gestures for two referents, namely, "Cut and paste a phrase" and "Copy and paste a phrase". This was due to the large number of semantically-unique gestures proposed by the participants, possibly as a result of these tasks being relatively more complex than the others. We did not include these two tasks in the final analysis and reserved gesture formulation for these tasks for future work.

Based on the codebook, two members of the research team then jointly coded all participant gestures. To validate our coding, we recruited a volunteer external to the research team to independently code the participant gestures using the defined codebook. We computed the inter-rater agreement of 85.8% using Cohen's kappa [31]. This indicates a very good agreement level and thus an appropriate set of codes[1].

User-defined gesture-set

After successfully classifying the participants' gestures into groups for each referent, we then computed the agreement score among the participants as in Wobbrock et al. [33]. This score is a measure pertaining to the concept of guessability which tells us how many of the proposed gestures for each referent and overall were shared among the participants [33]. If a majority of the participants agreed on a particular gesture for a referent, then the agreement score for that referent will be high. Additionally, if there are fewer groups of gestures

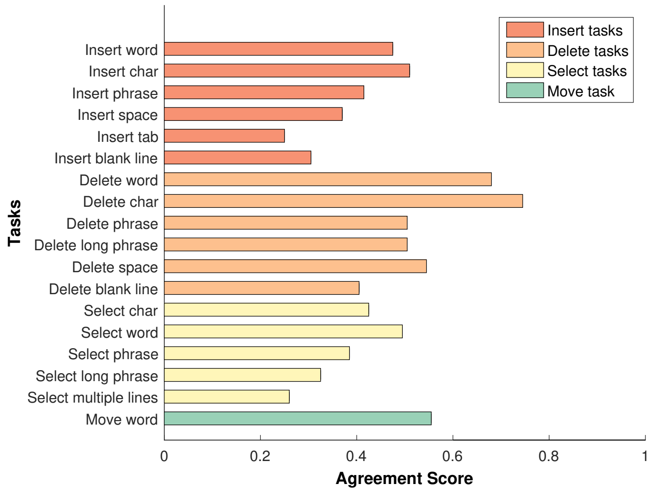

Figure 4: Agreement score for each referent.

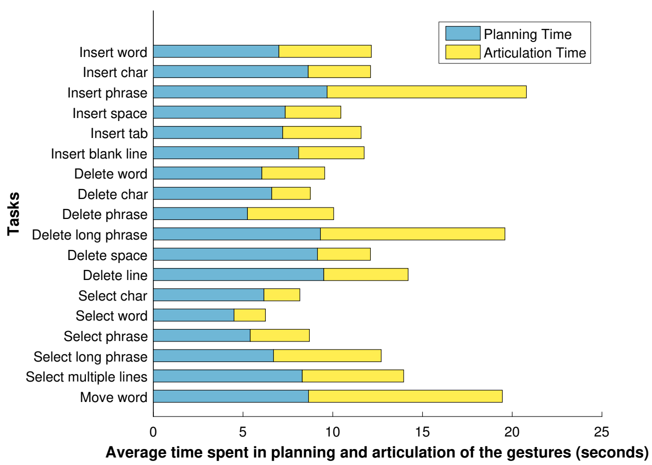

Figure 5: Average time spent by the participants in planning and articulating the gestures for each of the tasks.

for a referent, this too ensures a high agreement score. The formula for the agreement score [33] is given by

$$A = \frac{\sum_{r \in R} \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|}\right)^2}{|R|} \cdot 100\% \tag{1}$$

where r is a referent in the set of referents R, P_r is the set of proposals for referent r and P_i is the subset of identical gestures, i.e. deemed from the same group from P_r. Figure 4 shows the agreement scores for all the 18 referents. The overall agreement was 0.45 which means that the participants agreed on about 45% of the proposed symbols for a given referent on average, and this is comparable to agreement scores in [34].

According to Wobbrock et al. [33], guessability can be maximized by selecting those gestures for referents on which most participants agreed, or in other words, referents belonging to the largest group.

We designed our guessable gesture set by assigning the gestures belonging to the largest group for each referent. There were no conflicting gestures, i.e. the same gesture assigned to more than one referent, during the gesture assignment and the final gestures obtained for all the referents were stylus-based. The final gesture set is presented in Figure 6. For tasks containing multi-part gestures, the order in which they are performed is generally intuitive for most of the tasks, such as entering "(" before ")" for the "Select long phrase" task. For the "Insert" tasks, the sub-gesture denoting "where" is entered first followed by "what". For example, where was often denoted with a "∧" followed by the word, character or phrase to be inserted. The exact opposite order was used for for the *Select a word and move* task.

We incorporate *aliasing* [7, 33, 34] by assigning a group of semantically-equivalent gestures to a referent rather than a single specific gesture. Aliasing [7, 33, 34] refers to the practice of including multiple synonymous gestures for a referent in order to increase the recognition rates, make the interface more user-friendly and less stringent [32], and increase the guessability of the gesture-set. We present the user-defined gesture set with the dominantly-used gestures in Figure 6. We present the full set of gestures along with the aliases in the accompanying *supplementary material*.

The *guessability* of the final gesture-set was calculated using the formula provided by Wobbrock et al. [33] given by,

$$G = \frac{\sum_{s \in S} |P_s|}{|P|} \cdot 100\% \tag{2}$$

where P is the set of proposed gestures for all referents, and P_s is the set of proposed gestures using symbol s, which is a member of the resultant symbol set S.

We obtained a guessability score of 56.67% for our final gesture-set which is comparable to the guessability obtained by Wobbrock et al. [34]. This means that 56.67% of the proposed gestures of the participants are contained in the final set. Therefore, 204 gestures (non-unique) of the total 360 gestures collected are covered in the final set.

The average combined time for planning and articulation for each referent is shown in Figure 5. We did not find any significant correlations between the time (planning or articulation) and the post-task satisfaction responses (see Study Procedure).

Suggested extensions to the gesture set

While we believe that the final gesture set (Figure 6) contains intuitive gestures, we also suggest some extensions to the gesture set for a couple of tasks. These extensions include gestures which, although devised by a smaller group of participants, are no less intuitive and despite differing semantically, they can still be incorporated as *aliases* to the existing gestures.

For the "Delete long phrase" and "Select multiple lines" referents, we suggest adding the gestures shown in Figure 7. While the dominant gestures and aliases for these tasks appear appropriate for the length of text provided in the referents, they may be inappropriate for significantly longer lengths of text

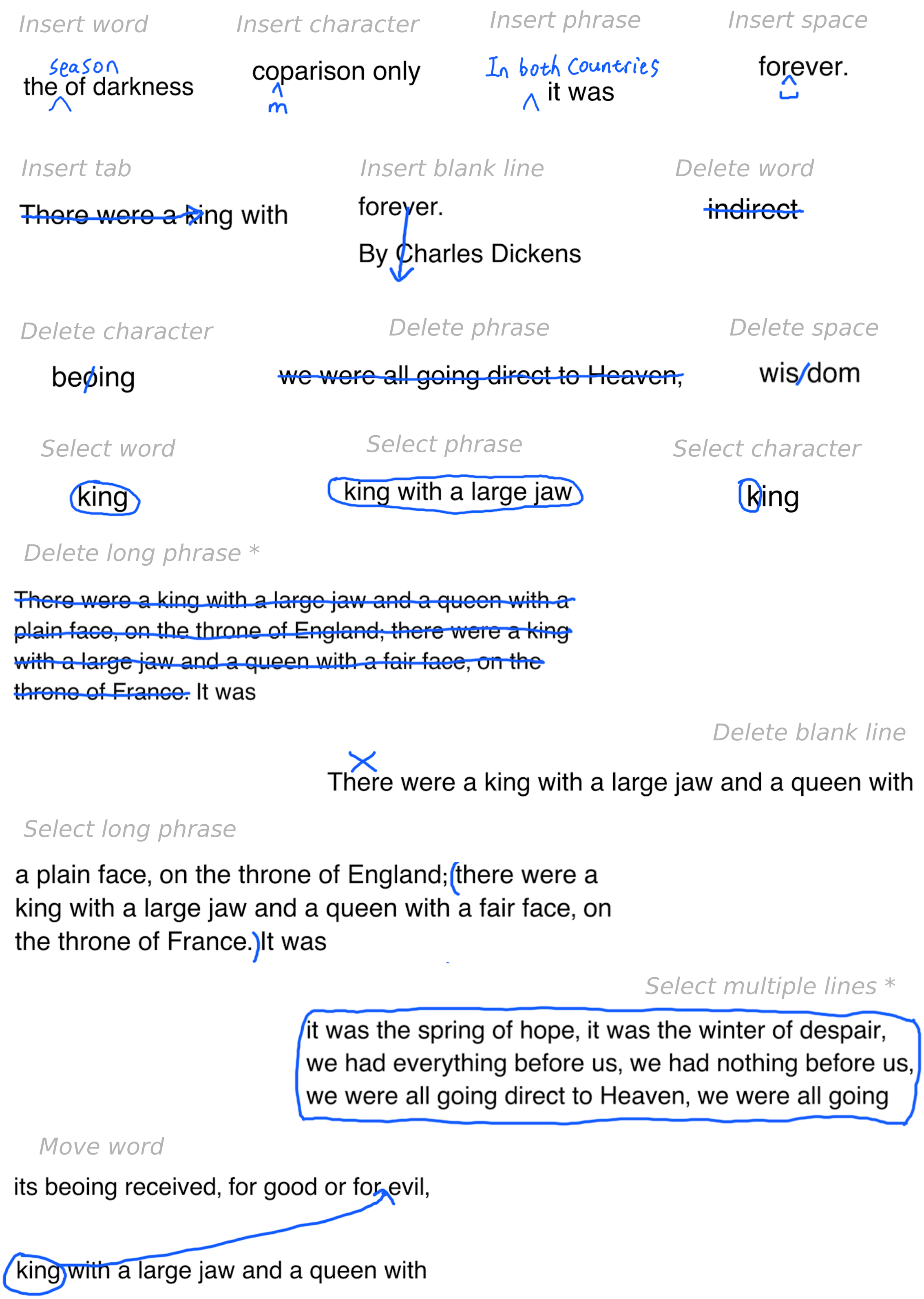

Figure 6: User-defined stylus-based gesture-set for common text editing tasks. The dominant gesture is represented for each referent above and the complete set of user-defined gestures with the aliases are included as supplementary material. We also propose alternative gestures for the "Delete long phrase" and "Select Multiple Lines" referents (denoted by a *) in Fig. 7.

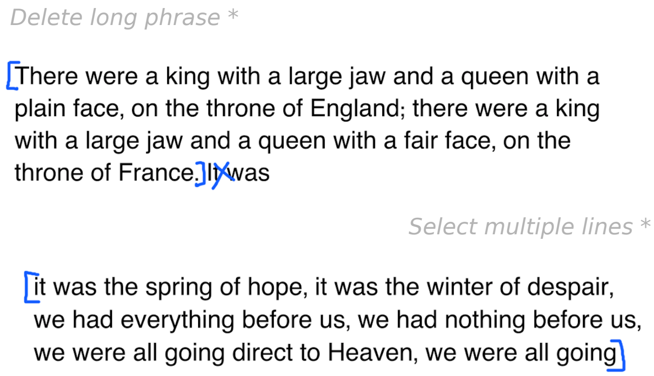

Figure 7: Suggested gesture-extensions for the "Delete long phrase" and "Select Multiple Lines" referents.

possibly appearing in practical scenarios and hence we suggest the respective extensions to simplify "selection" of the text.

Discussion

Then and Now

The influence of ubiquitous mobile touchscreen devices in today's society appears to be quite subtle. In comparing our study results with those of Welbourn and Whitrow [32], and Wolf and Morrel-Samuels [35], we find that the gestures presented for "Insert" (char, word, and phrase), "Delete" (char, word, and phrase), and "Move" (word) tasks are largely similar as we expected. Drawing a caret symbol "∧" to denote "Insert", deleting a character by drawing a vertical or slanted line over it, deleting a word and phrase using strike-through, and finally, moving a word by first circling it and then drawing an arrow from it to the final position were consistent with earlier studies as well as the concurrent study by Costagliola et al. [5]. These approaches reflect the general mental model of annotating text on paper.

Although the "Join", "Split", and "New paragraph" tasks considered by Welbourn and Whitrow [32] are synonymous with the "Delete space", "Insert space" and "Insert blank line" tasks, respectively, in our study, our gestures did not match their dominant gestures for these tasks. However, our gestures are contained in their extended gesture set, i.e. aliases [32]. We suspect that this could be due to the phrasing of the referents since a similar difference is observed in the gestures presented by Costagliola et al. [5] who also use the terms "Join" and "Split". Since our goal was to elicit gestures for use on modern devices, we worded our referents to be consistent with modern electronic device usage.

The Select tasks, delete long phrase, delete blank line, and insert tab are additional gestures that were included in our study but were not considered in the early studies [32, 35]. The gestures obtained for all the select tasks (Figure 6) were also reflective of paper annotation analogies. Furthermore, "selecting" text is often considered a basic operation that is used as part of other composite tasks such as deletion and move. However, by including both "Select" tasks as well as the composite tasks in our study and allowing the participants to think about them independently, we found that this is not always the case. For example, in the final gesture set (see supplementary material), all the "deletes" were performed directly and did not include the selection step. We suspect this is due to the relatively short length of the passages to be deleted, where it makes sense to specify the deletion directly. A long passage, however, would quickly become tedious with this form of deletion. While we recommend supporting this approach, we also suggest an extension (Figure 7) to support deletion of significantly longer phrases more effectively.

Design Implications: While we see consistency in gestures for certain tasks, others vary based on the user's mental model as well as the "size" of the task to perform. Designers should consider providing aliases for gestures, specifically for *tasks of varying sizes* (e.g. select a single word vs. select multiple paragraphs). Our participants treated text/space as manipulable digital elements. While prior studies indicate that users employ a paper *annotation* metaphor [32, 35], our results indicate that current participants may use a similar metaphor, but also consider the digital text to be dynamic - something that can be not only *removed*, but *moved* to make space.

Mental Models

One of our goals in replicating the elicitation studies was to see if the participants' experiences with touchscreen devices influenced their gestures. In other words, we wanted to see if *legacy bias* [21] played a role in the gestures they proposed. Legacy bias is a debatable topic associated with gesture elicitation studies which causes participants to propose gestures influenced by their accustomed interactions with other interfaces and technologies [21, 23]. While it is considered an undesirable effect by some since it may lead the participants to disregard what may be better suited for the system at hand [21], we share the views of those who see it as a basis for elicitation studies and hence beneficial [14].

As seen from the final gesture set and explained in the *Then and Now* section above, a majority of the participant-gestures were influenced by paper analogies - both in our study and in the original studies. We also observed other mental models at work, however, that can be classified as legacy-inspired. These behaviors, described below, were exhibited by small groups of participants and hence did not qualify to be included in the final gesture-set.

Modern touchscreens and the WIMP paradigm appeared to influence some of the participant gestures. Although the interface afforded direct access of any position on the screen, 3 participants first positioned a "cursor" at the required position before performing an editing task. They did this by tapping using the stylus, tapping and holding the stylus, and drawing a vertical line at the designated position. Additionally, for certain "Insert" tasks, subsequent to positioning the cursor, one participant assumed that a dialog box would pop up in which the word or phrase could be written.

Unexpected touchscreen-interaction-influenced gestures were also observed in responses for the "Insert" task. One participant used the two-finger 'pinch and zoom' gesture for inserting space and another participant assumed that after positioning the cursor, a menu would pop-up from which "space" or "tab" could be selected. For "Selecting a long phrase", a couple of participants performed gestures similar to the widget-based technique [6] normally used on touchscreen devices, where the users taps and holds the stylus/finger at the starting-point and a manipulator appears for the user to drag to choose the ending-point. These behaviors are clearly influenced by current touchscreen usage and interaction design and are more appropriate when using the finger which does not support as precise a positioning as the stylus. For most participants, however, the paper-based mental model was strong enough to overcome these tendencies resulting in gestures that we believe are more appropriate for text interaction.

Although gestures are typically "actions" or comprised of symbolic input, we observed a behavior in some participants wherein handwritten "commands" were used as gestures. For example, one participant used "I" to denote "Insert" and "S" for "Insert space", and 2 participants used "Del" for "Delete". We suspect that these gestures may have been influenced by the mental model of keyboard shortcuts used on desktops. Since 15 out of the 20 participants in our study were computer science majors, we, unsurprisingly, also came across a few programming-language-influenced commands such as "\tab" and "\n" for "Insert tab" and "Insert line" tasks, respectively. Six participants provided such gestures for the two mentioned referents. While our participant sample, comprised mostly of computer science majors, is a limitation of our study design, we, nevertheless, did not consider such idiosyncrasies found in the participant gestures for inclusion in our final gesture set.

Design Implications: To design a *guessable* gesture set that is independent of expertise, it is important to disregard gestures reflective of esoteric user groups (such as the programmers mentioned above). Although users can generally be expected to be very familiar with the mental models of the WIMP paradigm and modern touchscreen-usage, we found that these metaphors were invoked less often when it comes to editing text using a stylus. Rather, the paper-annotation metaphor is quite strong and appropriate. Additionally, including *aliases* will make the interface more user-friendly and cater to the differences in users' gestures within the same mental model. It should be noted that truly incorporating the paper metaphor necessitates incorporating a *dynamic in-situ* interface which supports handwriting (and recognition) *anywhere* on the document. We discuss this approach in more detail later.

Stylus vs. Touch

Although participants responded in the pre-study questionnaire that they rarely (on average) use a stylus on tablet devices, 90% of the them chose to use the stylus throughout the entire study as opposed to a finger. The other 10% of the participants applied finger gestures to draw gestures for a select few tasks but adopted the stylus for the majority of the tasks. The characteristics of stylus and finger touch interactions described previously in literature [4, 6, 12, 26] also hold true in our study. Generally speaking, the stylus affords greater precision and ease of use with respect to tapping and dragging tasks, and handwriting. This was reflected by the majority who used the stylus.

Design Implications: We recommend that stylus-based gestures for text-editing be incorporated in text editing applications on today's tablet devices, reserving finger gestures for navigation (e.g., zoom in/out, swipe to next page). This is not only reflective of the users' preferences in our study but also facilitates transitions between handwriting and editing and easy distinction between editing and navigation. Combined touch and pen interactions explored in the literature, such as mode switching techniques [12, 16, 25], can also be implemented in the stylus-based handwriting applications.

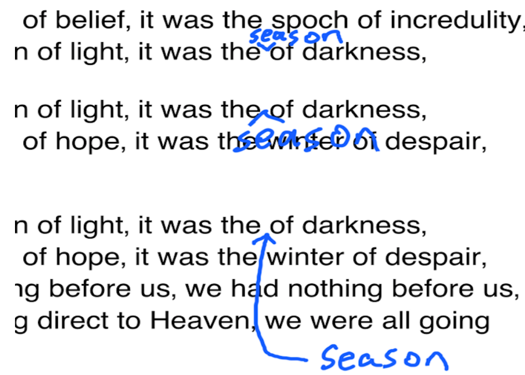

Figure 8: Three examples of whitespace usage.

White-Space Use and Manipulation

A clear difference between our results and those of prior studies was in the treatment of whitespace. For the various insert-text tasks (see Figure 8), 15 out of 20 participants chose to find and write in white space in either the gap between lines or the empty space at the bottom of the screen. Three people wrote the inserted words directly over the existing text and two participants said they preferred that a text-box or text-window "pop up" to let them write in. These observations suggest that users prefer to write in "blank spaces", however their mental model is consistent with that of annotating physical paper - that they must find *existing* white space in which to write.

For inserting a single space, participants generally selected an insertion point and drew a symbol to represent a "space" to be inserted. However, for larger space insertions (tab and whole line), participants tended to "move" existing text with an arrow gesture that started either where the space should be inserted or on the text immediately following where that space should be inserted (See Supplemental Material).

In the "delete an empty line" referent, participants mainly did one of the two following things: (1) deleted a blank line by crossing it out or (2) grabbed and moved the lines below it upward. Out of 20 participants, 14 of them treated the whitespace as an object that could be deleted directly as in (1). The other 6 participants chose to move the below paragraphs upward in order to get rid of the blank line.

The above participant behaviors can be considered as surprising observations resulting from our study. The participants were only instructed to assume that the interface would respond to their gestures as they intended in order to not restrict their inputs and our instructions did not mention white space or its manipulation or *how* the interface would respond dynamically.

Design Implications: Whitespace manipulation, in general, is complicated and deserves more attention. Our current recommendation to designers is to treat whitespace as a manipulable object, specifically for deletion operations, where gestures can be performed directly over the space. Insertion of whitespace, however, is more complex. Our current inclination is to be consistent with the whitespace as object metaphor and to design systems that utilize a symbol for the whitespace to be inserted (e.g. a tab arrow or a newline arrow). Further studies are needed to determine which option is most appropriate for insertion of whitespace for various sized tasks.

For insertion of text, it is clear that users prefer to write in open whitespace, ideally near the point of insertion. While no participants suggested that the system could "create" space *within* the typed text for them to write, this observation suggests that such a dynamic whitespace window may be an ideal method for supporting *in-situ* text entry. Indeed, Gu et al. found that in-place ink was a more effective and acceptable method than indirect writing [10]. This "text window" should dynamically resize as the user writes to create space as necessary, in much the same way that text moves out of the way when inserting new text using the mouse pointer and keyboard typing. The difference for handwriting is that the initial whitespace must be dynamically created to provide the user with the whitespace in which to begin writing.

CONCLUSION

We have presented a study replicating elicitation studies completed nearly 30 years ago to obtain user-defined gestures for text-editing tasks on handwriting interfaces. Overall, our findings indicate that the final gestures elicited from participants are largely reflective of the mental model of "annotating on paper," which conforms with the results of early studies, that users prefer the stylus to finger touch for text-editing and that "white-space" manipulation is complex. "White-space" manipulation presents a somewhat interesting area of future work and gestures for entering and editing white space need some attention.

REFERENCES

1. Douglas G Altman. 1990. *Practical statistics for medical research*. CRC press.
2. Daniel Avrahami, Scott E Hudson, Thomas P Moran, and Brian D Williams. 2001. Guided gesture support in the paper PDA. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*. ACM, 197–198.
3. Robert O. Briggs, Alan R. Dennis, Brenda S. Beck, and Jay F. Nunamaker, Jr. 1992. Whither the Pen-based Interface? *J. Manage. Inf. Syst.* 9, 3 (Dec. 1992), 71–90. DOI: http://dx.doi.org/10.1080/07421222.1992.11517968
4. A. Cockburn, D. Ahlström, and C. Gutwin. 2012. Understanding Performance in Touch Selections: Tap, Drag and Radial Pointing Drag with Finger, Stylus and Mouse. *Int. J. Hum.-Comput. Stud.* 70, 3 (March 2012), 218–233. DOI: http://dx.doi.org/10.1016/j.ijhcs.2011.11.002
5. Gennaro Costagliola, Mattia De Rosa, and Vittorio Fuccella. 2018. A technique for improving text editing on touchscreen devices. *Journal of Visual Languages & Computing* 47 (2018), 1–8.
6. Vittorio Fuccella, Poika Isokoski, and Benoit Martin. 2013. Gestures and widgets: performance in text editing on multi-touch capable mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2785–2794.
7. G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. 1987. The Vocabulary Problem in Human-system Communication. *Commun. ACM* 30, 11 (Nov. 1987), 964–971. DOI: http://dx.doi.org/10.1145/32206.32212
8. David Goldberg and Aaron Goodisman. 1991. Stylus user interfaces for manipulating text. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*. ACM, 127–135.
9. Aaron Goodisman. 1991. *A Stylus-Based User Interface for Text: Entry and Editing*. Master's thesis. Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science.
10. Jiseong Gu and Geehyuk Lee. 2016. In-Place-Ink: Toward More Direct Handwriting Interfaces. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*. ACM, 67–76.
11. François Guimbretière. 2003. Paper augmented digital documents. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*. ACM, 51–60.
12. Ken Hinckley, Koji Yatani, Michel Pahud, Nicole Coddington, Jenny Rodenhouse, Andy Wilson, Hrvoje Benko, and Bill Buxton. 2010. Pen+ touch= new tools. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 27–36.
13. Naoki Kato and Masaki Nakagawa. 1995. The design of a pen-based interface âĂŸshosaiâĂŹfor creative work. *Advances in Human Factors/Ergonomics* 20 (1995), 549–554.
14. Anne Köpsel and Nikola Bubalo. 2015. Benefiting from legacy bias. *interactions* 22, 5 (2015), 44–47.
15. Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.
16. Yang Li, Ken Hinckley, Zhiwei Guan, and James A Landay. 2005. Experimental analysis of mode switching techniques in pen-based user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 461–470.

17. Chunyuan Liao, François Guimbretière, and Ken Hinckley. 2005. PapierCraft: a command system for interactive paper. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*. ACM, 241–244.
18. Allan C Long, James A. Landay, and Lawrence A. Rowe. 1998. *PDA and Gesture Uses in Practice: Insights for Designers of Pen-Based*. Technical Report. Berkeley, CA, USA.
19. I Scott MacKenzie, Shawn X Zhang, and R William Soukoreff. 1999. Text entry using soft keyboards. *Behaviour & information technology* 18, 4 (1999), 235–244.
20. Anne Mangen, Liss G Anda, Gunn H Oxborough, and Kolbjørn Brønnick. 2015. Handwriting versus Keyboard Writing: Effect on Word Recall. *Journal of Writing Research* 7, 2 (2015).
21. Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, Jacob O Wobbrock, and others. 2014. Reducing legacy bias in gesture elicitation studies. *Interactions* 21, 3 (2014), 40–45.
22. Pam A Mueller and Daniel M Oppenheimer. 2014. The pen is mightier than the keyboard advantages of longhand over laptop note taking. *Psychological science* (2014), 0956797614524581.
23. Uran Oh and Leah Findlater. 2013. The challenges and potential of end-user gesture customization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1129–1138.
24. Seyoung Park and Donghee Shin. 2015. Effects of text input system on learner's memory: handwriting versus typing on tablet PC. In *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*. ACM, 30.
25. Ken Pfeuffer, Ken Hinckley, Michel Pahud, and Bill Buxton. 2017. Thumb+ Pen Interaction on Tablets.. In *CHI*. 3254–3266.
26. Felix Raab, Christian Wolff, and Florian Echtler. 2013. RefactorPad: editing source code on touchscreens. In *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, 223–228.
27. Janet C Read. 2007. A study of the usability of handwriting recognition for text entry by children. *Interacting with Computers* 19, 1 (2007), 57–69.
28. Yann Riche, Nathalie Henry Riche, Ken Hinckley, Sheri Panabaker, Sarah Fuelling, and Sarah Williams. 2017. As We May Ink?: Learning from Everyday Analog Pen Use to Improve Digital Ink Experiences.. In *CHI*. 3241–3253.
29. Johnny Saldaña. 2015. *The coding manual for qualitative researchers*. Sage.
30. Lambert Schomaker. 1998. From handwriting analysis to pen-computer applications. *Electronics & Communication Engineering Journal* 10, 3 (1998), 93–102.
31. Nigel C Smeeton. 1985. Early history of the kappa statistic. (1985).
32. LK Welbourn and Robert J Whitrow. 1988. A gesture based text editor. In *Proceedings of the Fourth Conference of the British Computer Society on People and computers IV*. Cambridge University Press, 363–371.
33. Jacob O Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A Myers. 2005. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems*. ACM, 1869–1872.
34. Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1083–1092.
35. Catherine G Wolf and Palmer Morrel-Samuels. 1987. The use of hand-drawn gestures for text editing. *International Journal of Man-Machine Studies* 27, 1 (1987), 91–102.
36. Dongwook Yoon, Nicholas Chen, and François Guimbretière. 2013. TextTearing: opening white space for digital ink annotation. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, 107–112.
37. Emanuel Zgraggen, Robert Zeleznik, and Philipp Eichmann. 2016. Tableur: Handwritten Spreadsheets. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 2362–2368. DOI: http://dx.doi.org/10.1145/2851581.2892326