Stylus-based Gestures for Text Editing on Tablet Devices Revisited

1st Author Name	2nd Author Name	3rd Author Name
Affiliation	Affiliation	Affiliation
Address	Address	Address
e-mail address	e-mail address	e-mail address
Optional phone number	Optional phone number	Optional phone number

ABSTRACT

Recent advances in stylus technology and handwriting recognition have made handwriting a viable text-entry option on touchscreen devices. Initial ideas for building intuitive and usable handwriting applications originated nearly 30 years ago. In this paper, we present a study intended to replicate studies from the 80s involving the elicitation of hand-drawn gestures from users for common text-editing tasks in order to design a 'guessable gesture-set and to determine 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, that users prefer using the stylus to finger touch for text editing, and that manipulating 'white space' is complex. We present our findings and a stylus-based, user-defined gesture-set for text editing that can be incorporated in the handwriting applications on today's tablet devices.

Author Keywords

Handwriting interfaces; stylus interaction; text editing; tablet devices;

ACM Classification Keywords

H.5.2. User Interfaces: Input devices and strategies (e.g., mouse, touchscreen)

INTRODUCTION

Tablet devices are widely used but text entry and editing using the virtual keyboard on these devices can be inconvenient and inefficient [11, 5]. 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 [15, 18, 12, 14], 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 [20].

Applications that support handwritten input on electronic devices typically leverage aspects of writing on paper, such as ease of use, with aspects of electronic documents, such as providing digital representations of the handwritten text, thus supporting features such as 'search'. Therefore, effective stylus-input and accurate handwriting recognition are two essential components 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 [1]. 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, for example, 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 research was done in the 80s and 90s on using handwriting for text entry and designing gestures for editing tasks on pen-based computers [8, 25, 22, 7, 9]. These interfaces were limited by the technology of the time and erroneous recognition systems. Nevertheless, the design insights they provide such as support for writing anywhere on the document and the editing gestures that 'mimic' writing on paper are very intuitive and largely applicable even today. These design aspects of handwriting interfaces have long since been forgotten and are essential elements 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 make design decisions for a handwriting interface on today's tablet devices. In this paper, we focus on devising stylus-based gestures for text editing which are a key component in the realization of usable handwriting applications. We performed an elicitation study almost identical to the ones described in the papers by Welbourn and Whitrow [22] and Wolf and Morrel-Samuels [25] to obtain user-defined gestures for common text editing tasks. Since 'touch' is the predominant form of interaction on today's tablet devices, the users could perform the gestures using either finger touch or a stylus. This paper mainly focuses on defining a gesture-set for text editing reflective of users' inputs, and other qualitative assessments which are outlined below and does not focus on the recognition aspects of the gestures.

In addition to eliciting intuitive gestures from users, the goals of our study included determining if the results of early research still apply given the much more technologically advanced and ubiquitous devices of today and exploring other issues in finger or stylus-based text editing. More specifically, we were interested in the following questions:

- What gestures do users choose for typical text editing tasks and to what degree do users 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 (for example, paper analogies, or touchscreen interaction) influence the users' gestures?
- How do participants manipulate whitespace with a stylus?

We provide design recommendations for stylus-based gestures for text editing tasks which are intuitive and independent of expertise and can be incorporated in handwriting interfaces on tablet devices. In the following sections, we discuss related work, describe our elicitation study, and discuss results before concluding with a discussion of our findings and potential future work directions and implications for design of stylus-based text entry and editing.

RELATED WORK

Notable work was done on handwritten text entry and editing on pen-based computers in the 80s and 90s [22, 25, 8, 7, 9]. Welbourn and Whitrow [22] and Wolf and Morrel-Samuels [25] describe studies wherein users were asked to perform gestures for various text-editing operations. The gestures presented in these papers mostly mimicked editing on paper. While the study was pen and paper-based in the paper by Wolf and Morrel-Samuels [25], a combination of a tablet (for stylus use) and a separate display (to view the writing) was used in the Welbourn and Whitrow paper [22]. 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. The specific procedure used in our study was based on the elicitation study described by Wobbrock et al. [24] where users were presented with the effects of editing operations (called 'referents') that showed the state of the interface before and after an operation and were asked to 'invent' the gesture that brings about the state change.

Goodisman and Goldberg [8, 7] of Xerox PARC provide valuable insights on various aspects of handwriting interfaces including their advantages and drawbacks, methods to counteract the erroneous recognition, and 'markup editing'. 'Markup editing' 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. Another advantage is that 'undo' operations can be performed in any order. A PARC-built *scratchpad* was used which had a transparent tablet atop a display and users wrote on the tablet. Similar to 'markup editing', Kato and Nakagawa [9] 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.

Although we are primarily interested in handheld tablet devices, it is interesting to note that stylus-based note-taking applications were also designed for large interactive whiteboards at Xerox PARC in the 90s [16, 13]. While these applications do not incorporate recognition of written text, Moran et al. [13] describe anything drawn or written on the whiteboard as a free graphical object (*freeGO*). These *freeGOs* can behave in various ways as perceived by the users and their behaviors are implied by the actions performed on them ('freeform' interaction). For example, there can be character *freeGOs* and by performing text-editing operations on them, they will be treated as text by the underlying system. The papers describe gestures as meta-strokes to distinguish them from the written material and provide gestures for selecting and grouping *freeGOs*.

Other work on pen-based applications in general list their characteristics, existing and potential applications, perceptions of users towards them and compare the use of stylus over the finger or mouse [3, 20, 4, 10]. The survey conducted by Long et al. [10] compared, in addition to other metrics, the use of text-editing gestures on the Apple Newton and PalmPilot. It was found that Newton users used gestures more often than Pilot users and that the difficulty in remembering Pilot gestures was stated as the main reason participants used them less frequently. This could be because the Newton gestures appear more intuitive than the Pilot gestures. In comparing finger and stylus interactions for pointing tasks, Cockburn et al. [4] found that the stylus was slower than finger pointing for tapping tasks but more accurate, especially with small targets and stylus was faster than finger touch for dragging tasks, possibly due to the higher friction between finger touch and the surface. Since we are interested in comparing finger and stylus interactions in the context of text editing, the above findings can be useful in interpreting our results.

Most recently, Gu et al. [1] present *In-Place-Ink* which is 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* [26] is a recently implemented handwritten 'spreadsheet' application affording general 'spreadsheet' functionality such as automatically filling cells, and applying formulas to cells. The functionality is invoked by the use of stylus-based gestures and while all the handwritten content is recognized, the content is retained and edited in digital ink format.

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

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

on (touch-based) gestures for text editing on these devices. Fuccella et al. [5] address the imprecision, unwieldiness, and twofold actions (point and pick from menu) involved in the conventional 'widget-based' text editing on touchscreen devices. They present one- and two-finger gestures for cursor-positioning, selecting text and clipboard tasks which are drawn indirectly and on the soft keyboard without hindering its use. This paper summarizes the typical user behaviors involved in text editing on today's touchscreen devices. *RefactorPad* [17] focuses on enabling programming on mobile devices and in addition to eliciting pen and touch gestures for refactoring tasks from users, gestures are also obtained for common text-editing tasks. We will also examine the use of touch vs. stylus based gestures in the context of general text editing.

GESTURE-ELICITATION STUDY

Our first goal in this paper 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 an elicitation study similar to that described by Wobbrock et al. [24] to elicit the most natural gestures for participants for common text editing tasks.

Study Design

We conducted a within-subjects elicitation study where participants were given referents, or examples of the desired output, and asked to perform the gesture that should 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. Users were asked to provide gestures for 20 such referents that were grouped 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

Although 'Move' is synonymous with 'Cut and Paste', the two operations are often perceived differently in the context of text editing. While 'Cut/Copy/Paste' are considered clipboard operations with two disjoint tasks where 'Paste' needn't occur immediately after 'Cut/Copy', 'Move' is often perceived as an uninterrupted 'draggable' operation and this is confirmed by our results.

The apparatus included 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. Each stroke was captured (using different colored ink for stylus and finger inputs) and tagged with the current task number for later analysis.

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 [redacted for anonymous review]. 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). Participants also reported that they have 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 frequently.

Procedure

Each session began with completion of an informed consent document. Participants then completed a pre-study questionnaire that consisted of questions 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. The participant was given the opportunity to ask any questions during and upon completion of the practice task.

The user was then presented with a randomized set of 20 tasks, one at a time. For each task, users were presented with the referent on paper, asked to read it, and instructed 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. They were instructed to demonstrate the most natural gesture for the given referent. The participants were asked to assume that the implemented interface will afford writing (and recognition) anywhere on the document and did not require the use of the virtual keyboard. We encouraged the participants to think aloud as they completed each task. In addition,

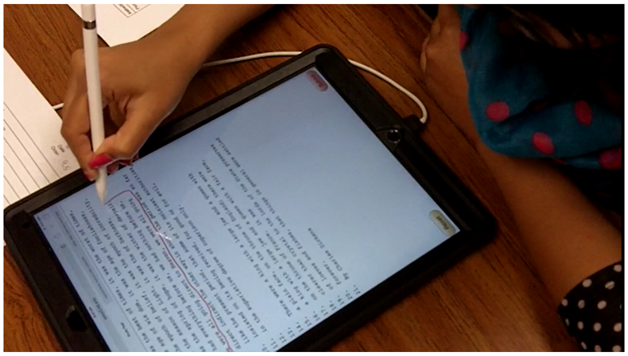


Figure 2: A participant performing the 'Select multiple lines' task during one of the study sessions.

for each task, we asked the participants how they would expect the system to respond to their gesture if it was capable of responding. After each task, the participant answered the following 3 questions by giving a rating from 1 (Entirely Disagree) to 7 (Entirely Agree):

1. The gesture I picked is a good match for its intended purpose.
2. The gesture I picked is easy to perform.
3. The gesture I picked is easy for me to come up with.

Participants were allowed to utilize the same gesture for any tasks within the same group, but not between groups. If a gesture was re-used between groups, the experimenter brought it to their attention and asked for a clarification. All sessions were videotaped for later reference and analysis. A participant performing a sample task during the study is shown in Figure 2.

ANALYSIS AND RESULTS

Since we aim to present a gesture-set for common text editing tasks as a result of this study, we use the 'guessability' methodology formulated by Wobbrock et al. [23] to both devise a 'guessable' gesture-set and to assess its guessability. Guessability can be defined as that quality of gestures (or any symbolic input) that makes their behavior (the referents for which they are designed) predictable and memorable to users with no previous experience [23]. We begin by grouping the collected participants' gestures and then apply the guessability methodology to the groups created 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. For marking the beginning and end, one participant may have used brackets ' [ ] ' while another may have used parentheses ' ( ) '. These two options, while not identical, were semantically equivalent (See Figure 3).

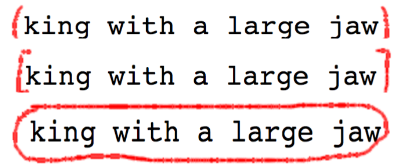


Figure 3: Three sample gestures used for the 'Select phrase' task. The first two gestures, although different in form, are considered semantically equivalent since they both involve marking the beginning and end of the phrase. The third gesture, involving circling or bounding the phrase, is considered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

To determine appropriate classification of gestures, two members of the research team conducted an open-coding [19] 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 codes were entered 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, the high-level categorization was based 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. Two members of the research team then used the codebook to label each participant gesture with a single code (or two codes in the case of a hierarchical gesture).

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 left gesture formulation for these tasks for future work.

To validate our grouping methodology, we recruited a volunteer external to the research team to independently code the participant gestures. We provided him with the entire codebook and the actual gesture set obtained from the participants. We then asked him to assign a code to each gesture based on the descriptions in the codebook. We used Cohen's kappa [21] to compute the inter-rater coding agreement between the two members of the research team and the independent coder.

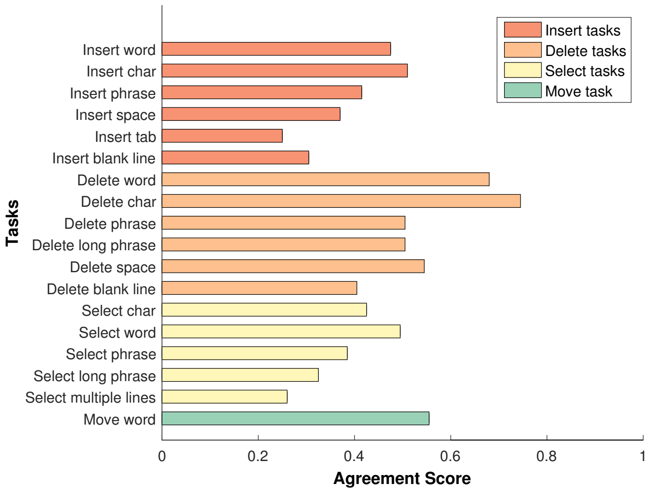


Figure 4: Agreement score for each referent.

On average, there was an 85.8% agreement rate between the research team's ratings and those of the independent rater, indicating a very good agreement level and thus an appropriate set of codes[2].

User-defined gesture-set

Now that we have successfully classified the participants' gestures into groups for each referent, we can now compute the agreement score among the participants using the formula presented by Wobbrock et al. [23]. The agreement 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 [23]. 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) for a referent, this too ensures a high agreement score. The formula for the agreement score [23] is given by

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

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 (or deemed from the same group) from P_r. Figure 4 shows the agreement scores for all the 18 referents.

According to Wobbrock et al. [23], guessability can be maximized by selecting those gestures for referents on which most participants agreed (or in other words, belonging to the largest group).

We designed our 'guessable' gesture-set by assigning the gestures belonging to the largest group to every referent. There were no conflicting gestures (assigning the same gesture to more than one referent) during the gesture-assignment and the final gestures obtained for all the referents were stylus-based. The gesture-set developed for common text editing tasks using the participants' gestures is presented in Figures 5 and 6.

The 'guessability' of the final gesture-set was calculated using the formula provided by Wobbrock et al. [23] 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 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.

We incorporate 'aliasing' [6, 24, 23] by assigning a group of (semantically-equivalent) gestures to a referent rather than a single specific gesture. 'Aliasing' [6, 24, 23] 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 [22], and increase the guessability of the gesture-set. For each referent, we present the set of synonymous gestures and we highlight the gesture that we would recommend with a * (see Figures 5 and 6).

The median time (median calculated as opposed to mean due to the presence of a couple of outliers) taken by the participants in both thinking of an appropriate gesture and performing the gesture for each referent is shown in Figure 8. This means that for referents which have a greater amount of time attributed to them, the participants could have either taken longer to perform or to think of an appropriate gesture or both and vice versa.

Suggested extensions to the gesture-set

While we think that the final gesture set (Figures 5 and 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 to the gesture-set. While the gestures for these tasks in the final set (Figures 5 and 6) appear appropriate for the length of text provided in the referents, they may be inappropriate for significantly longer lengths of text possibly appearing in practical scenarios and hence we suggest the respective extensions to simplify 'selection' of the text.

Qualitative Observations

Then and Now

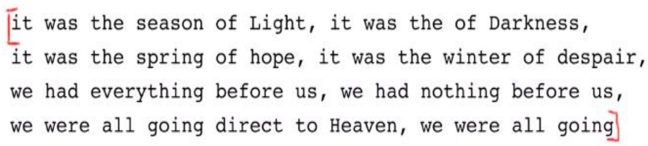
The text-editing tasks considered in the early elicitation studies [22, 25] consist mainly of 'Insert' (char, word, and

Tasks	Gestures
Insert word	the season of Darkness * the season of Darkness
Insert character	coparison only m * coparison only m coparison only m
Insert phrase	In such countries it was * it was ~~the lords of the~~ In both countries
Insert space	forever * forever forever forever forever forever forever
Insert Tab	~~There were a k~~ing with * There were a king with There were a king with a large jaw and a queen with
Insert blank line	forever. By Charles Dickens * forever. By Charles Dickens forever. By Charles Dickens forever. By Charles Dickens
Delete word	~~indirect~~ * ~~indirect~~ ~~indirect~~
Delete character	beding * beding beding
Delete phrase	~~we were all going direct to Heaven,~~ *
Delete long phrase	~~There were a king with a large jaw and a queen with a plain face, on the throne of England; there were a king with a large jaw and a queen with a fair face, on the throne of France.~~ it was ~~There were a king with a large jaw and a queen with a plain face, on the throne of England; there were a king with a large jaw and a queen with a fair face, on the throne of France.~~ it was (See Figure 7 for gesture extension) *
Delete space	wis dom * wis dom wis dom wis dom wis dom
Delete an empty line	There were a king with a large jaw and a queen with There were a king with a large jaw and a queen with

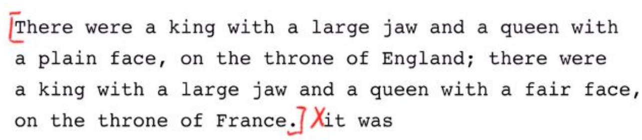
Figure 5: User-defined stylus-based gesture-set for common text editing tasks. The dominant gesture recommended for every referent is denoted by a * on its top-right corner.

	There were a king with a large jaw and a queen with * There were a king with a large jaw and a queen with	
Select character	king *	king
Select word	king *	king
Select phrase	king with a large jaw *	king with a large jaw
Select long phrase	a plain face, on the throne of England; there were * a king with a large jaw and a queen with a fair face, on the throne of France. it was a plain face, on the throne of England; there were a king with a large jaw and a queen with a fair face, on the throne of France. it was a plain face, on the throne of England; there were a king with a large jaw and a queen with a fair face, on the throne of France. it was a plain face, on the throne of England; there were a king with a large jaw and a queen with a fair face, on the throne of France. it was	
Select * multiple lines	it was the spring of hope, it was the winter of despair, we had everything before us, we had nothing before us, we were all going direct to Heaven, we were all going (See Figure 7 for gesture extension)*	
Select a word and move	on its beoing received, for good or for evil, * perlative degree of coparison only. e a king with a large jaw and a queen with its beoing received, for good or for evil, :lative degree of coparison only. ı king with a large jaw and a queen with its beoing received, for good or for evil, Lative degree of coparison only. king with a large jaw and a queen with	

Figure 6: User-defined stylus-based gesture-set for common text editing tasks. The dominant gesture recommended for every referent is denoted by a * on its top-right corner.



(a) Select multiple lines



(b) Delete long phrase

Figure 7: Suggested gesture-extensions for the ‘Select multiple lines’ and the ‘Delete long phrase’ referents

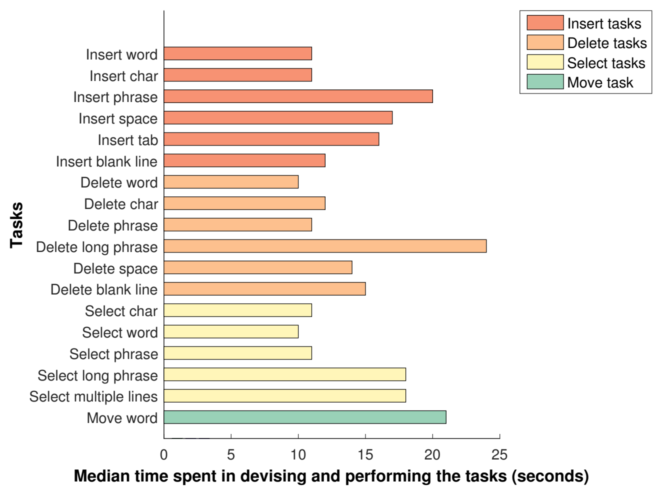


Figure 8: Median time spent by the participants on each of the tasks.

phrase), ‘Delete’ (char, word, and phrase), and ‘Move’ (word, phrase, and paragraph) tasks. Additionally, they also include ‘Join’, ‘Split’, ‘New paragraph’, ‘Add space’ and ‘Position phrase’ tasks [22, 25].

However, in the more recent paper by Fuccella et al. [5] which focuses on modern touchscreen device-usage, gestures are formulated for basic tasks - positioning the cursor, selecting, and clipboard operations (cut/copy/paste). The participants are then given ‘Insert’, ‘Delete’, and ‘Move’ tasks (similar to the ones in the early studies mentioned above) which are to be performed using the gestures formulated for the basic tasks.

For our study, we combined the tasks considered in the early elicitation studies [22, 25] as well as the ones in the paper by Fuccella et al. [5] so that our task-list was representative of the text-editing operations usually performed on paper and electronically.

In comparing our study results with those of Welbourn and Whitrow [22], and Wolf and Morrel-Samuels [25] for overlapping tasks, we find that the gestures presented for ‘Insert’ (char, word, and phrase), ‘Delete’ (char, word, and phrase), and ‘Move’ (word) tasks are largely similar. 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 gestures obtained in our study (Figures 5 and 6) as well as the earlier studies and these gestures can be said to reflect the mental model of writing on paper.

Although the ‘Join’, ‘Split’, and ‘New paragraph’ tasks considered by Welbourn and Whitrow [22] 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 (‘aliases’) [22]. We suspect that this could be due to the wording of the referents and we worded our referents to be consistent with electronic text-editing terminology. The ‘Add space’ task considered by Wolf and Morrel-Samuels [25] sounds ambiguous and going by the referent presented in the paper, it could be considered equivalent to the ‘Insert blank line’ referent in our study. They chose writing ‘space’ in double quotes and drawing an arrow pointing to the position of insertion as the gesture for this task while our gestures (Figure 5) mostly include a vertical downward-pointing arrow (e.g. dragging the current line down to make space) for this task.

All the ‘Select’ tasks, delete long phrase, delete blank line, and insert tab are additional gestures that were included in our study which were not considered in the early studies [22, 25]. The gestures obtained for all the ‘Select’ tasks (figure 5) were also reflective of paper analogies.Furthermore, ‘selecting’ text is often considered as 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 (Figures 5 and 6), all the ‘deletes’ were performed directly and did not include the selection step. 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. Of course, there is no one right answer. Designers should consider providing aliases for gestures, specifically for *tasks of varying sizes* (e.g. select a single word vs. select multiple paragraphs).

Mental Models

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. However, we also observed other mental models at work in the behaviors exhibited by participants.

Modern touchscreen device-usage and the WIMP paradigm also influenced some of the participant-gestures. Although the interface afforded direct access of any position, a few participants first positioned a 'cursor' at the required position before performing an editing task. Tapping using the stylus, tapping and holding the stylus and drawing a vertical line at the designated position were ways in which the cursor was positioned. Additionally, for certain 'Insert' tasks, subsequent to positioning the cursor, one participant assumed that a box would pop-up in which the word or phrase could be written and a few others wrote the text in the space below where the virtual keyboard usually appears.

Touchscreen interaction-influenced gestures were also performed by a couple of participants for certain tasks not normally performed when writing on paper. One participant used the two-finger 'pinch and zoom' gesture for inserting space and another participant assumed a menu would pop-up post-cursor-positioning from which 'space' or 'tab' could be selected to be inserted at that position. For 'Selecting a long phrase', a couple of participants performed gestures similar to the widget-based technique [5] normally used on touchscreen devices by tapping and holding the stylus/finger at the starting-point and dragging it to the ending-point. These behaviors are clearly influenced by current touchscreen usage and interaction design and are more appropriate when using the finger. 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, 'I' was used to denote 'Insert', 'S' for 'Insert space', and 'Del' for 'Delete'. We suspect that these gestures may have been influenced by the mental model of keyboard shortcuts used on desktops. Since a percentage of the participants were Computer Science majors, we also came across a few programming language-influenced commands such as '\tab' and '\n' for 'Insert tab' and 'Insert line' tasks, respectively. We did not consider such idiosyncrasies found in participant-gestures for inclusion in our final gesture-set.

Design Implications: In order to design a 'guessable' gesture set which 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 are incompatible and non-intuitive when it comes to editing text using a stylus on a handwriting application. Therefore, designers should consider gesture-sets that are in accordance with the mental model of 'writing on paper'. 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 incorporating the paper metaphor necessitates incorporating a 'dynamic in-situ' interface which allows 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 through 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 [5, 17, 4] also hold true in our study. Generally speaking, stylus affords greater precision and ease of use with respect to tapping and dragging tasks, and handwriting. Touch-interaction can be used with relatively less ease for tapping and dragging tasks but is highly unsuitable for handwriting. The drawbacks of touch-interaction for text-editing discussed by Fuccella et al. [5], such as the difficulty in selecting from the beginning of the line due to the position coinciding with the edge of the display and increasing difficulty in using finger touches as the 'font size' decreases, are also valid and perhaps explain why few participants used finger touches for our tasks.

Although our study permitted the use of stylus and touch interchangeably, different functionality is generally associated with the two interaction types and this was observed during our study. While stylus interaction is almost always associated with creating digital ink on a handwriting application, typical touch-interaction on today's touchscreen devices does not create digital ink and is typically used for navigation purposes. While a few participants did use the stylus and touch interchangeably, we adhere to these functional distinctions going by the behavior of the majority of the participants.

Design Implications: We recommend that stylus-based gestures for text-editing be incorporated in handwriting 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 makes for seamless transitions between handwriting and editing and easy distinction between editing and navigation.

White-Space Use and Manipulation

One of the differences between our study results and those of prior studies was the treatment of whitespace. In several of the tasks, users were required to enter text via handwriting. 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 would 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 physical paper - that they must find white space in which to write. 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 acceptable method for supporting "in-situ" text entry. This is also consistent with

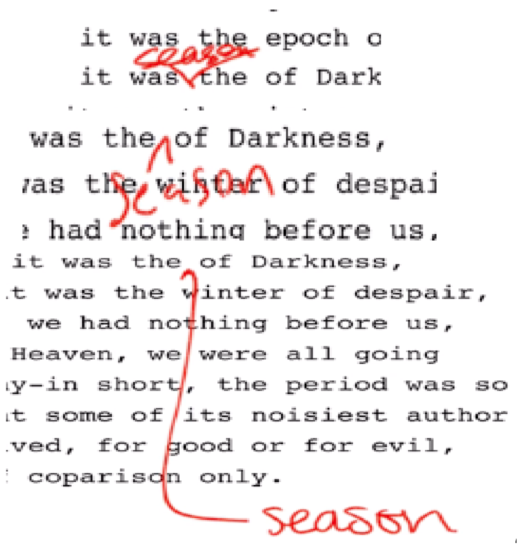


Figure 9: Three examples of whitespace usage.

some participants' use of the "pinch to zoom" gesture for creating whitespace.

For inserting a small space (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 put or on the text immediately following where that space should be located (See Figure 6).

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.

To further explore participants' whitespace manipulation, we compared their answers to the post-task questionnaires (see Study Procedure) with regards to the whitespace manipulation tasks. We divided the average participant-responses for the three post-task questions into two groups - one group containing the 'Insert space', 'Insert tab' and 'Insert blank line' referents and the other group containing the remaining 15 referents. We then performed t-tests between these two groups to see if their means differed significantly for each of the three post-task questions. It was found that the difference in the means of participant-responses was statistically significant for the first ($p < 0.05$) and third ($p < 0.05$) questions (the first group having lower means) but not for the second question. This suggests that the participants, in general, found it harder to come up with gestures for the 'Insert' (space, tab, and blank line) tasks and also thought that the gestures they devised did not match those tasks very well.

Design Implications: Whitespace manipulation is complicated and deserves more attention. Our current recommendation to designers is to treat whitespace as a manipulable object, specifically for deletion operations. The whitespace exists in this case, and can be addressed with a gesture directly over it. Insertion, however, is more complex. It does not exist, so how does one specify something that does not have a physical representation. One option is to require that users specify an insertion point and a symbol to represent the space object. Alternatively, one can employ the model that users drag existing text around to make the whitespace. 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 insert. Designers should adopt an insertion interaction that creates space at the desired point of insertion to allow for 'in-situ' specification of text. This 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 created to provide the user with a place to write.

CONCLUSION AND FUTURE WORK

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. We compared our results with those of past literature, presented the similarities and differences, provided analyses of various aspects of observed user behavior, and provided implications for designers of applications that require text entry and editing. Overall, our findings indicate that the final gestures elicited from participants are largely reflective of the mental model of 'writing on paper' which conforms with the results of early studies, that users prefer using the stylus to finger touch for text-editing and that 'white-space' manipulation is complex.

Future work includes devising gestures for clipboard operations (cut/copy/paste), 'undo' operation, and incorporating touch gestures for document-traversal such as scroll up/down and zoom in/out. Presently, there is a lack of information available on the frequency of editing tasks [5] but given this information, we can further refine the gesture set by prioritizing more frequently occurring tasks. Another design consideration would be to include line numbers in the document inspired by the results obtained by Raab et al. [17] wherein selection of multiple complete lines and blocks was done using the line numbers by a majority of the participants. 'White-space' manipulation presents a somewhat interesting area of research and gestures for entering and editing white space need some attention in future work. Our study indicates that a successful text entry and editing interface must also allow users to write in white space that exists, or to make whitespace by creating 'writing windows' directly within the recognized text to insert new text.

REFERENCES

1. In-Place-Ink: Toward More Direct Handwriting Interfaces, author=Gu, Jiseong and Lee, Geehyuk. In *Proceedings of the 2016 ACM on Interactive Surfaces and Spaces*, ACM (2016), 67–76.
2. Altman, D. G. *Practical statistics for medical research*. CRC press, 1990.
3. Briggs, R. O., Dennis, A. R., Beck, B. S., and Nunamaker, Jr., J. F. Whither the Pen-based Interface? *J. Manage. Inf. Syst. 9*, 3 (Dec. 1992), 71–90.
4. Cockburn, A., Ahlström, D., and Gutwin, C. Understanding Performance in Touch Selections: Tap, Drag and Radial Pointing Drag with Finger, Stylus and Mouse. *Int. J. Hum.-Comput. Stud. 70*, 3 (Mar. 2012), 218–233.
5. Fuccella, V., Isokoski, P., and Martin, B. 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 (2013), 2785–2794.
6. Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. The Vocabulary Problem in Human-system Communication. *Commun. ACM 30*, 11 (Nov. 1987), 964–971.
7. Goldberg, D., and Goodisman, A. Stylus user interfaces for manipulating text. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*, ACM (1991), 127–135.
8. Goodisman, A. A Stylus-Based User Interface for Text: Entry and Editing. Master's thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 1991.
9. Kato, N., and Nakagawa, M. The design of a pen-based interface shosaifor creative work. *Advances in Human Factors/Ergonomics 20* (1995), 549–554.
10. Long, A. C., Landay, J. A., and Rowe, L. A. PDA and Gesture Uses in Practice: Insights for Designers of Pen-Based. Tech. rep., Berkeley, CA, USA, 1998.
11. MacKenzie, I. S., Zhang, S. X., and Soukoreff, R. W. Text entry using soft keyboards. *Behaviour & information technology 18*, 4 (1999), 235–244.
12. Mangen, A., Anda, L. G., Oxborough, G. H., and Brønnick, K. Handwriting versus Keyboard Writing: Effect on Word Recall. *Journal of Writing Research 7*, 2 (2015).
13. Moran, T. P., Chiu, P., van Melle, W., and Kurtenbach, G. Implicit Structure for Pen-based Systems Within a Freeform Interaction Paradigm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, ACM Press/Addison-Wesley Publishing Co. (New York, NY, USA, 1995), 487–494.
14. Mueller, P. A., and Oppenheimer, D. M. The pen is mightier than the keyboard advantages of longhand over laptop note taking. *Psychological science* (2014), 0956797614524581.
15. Park, S., and Shin, D. 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 (2015), 30.
16. Pedersen, E. R., McCall, K., Moran, T. P., and Halasz, F. G. Tivoli: An Electronic Whiteboard for Informal Workgroup Meetings. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, ACM (New York, NY, USA, 1993), 391–398.
17. Raab, F., Wolff, C., and Echtler, F. RefactorPad: editing source code on touchscreens. In *Proceedings of the 5th ACM SIGCHI symposium on Engineering interactive computing systems*, ACM (2013), 223–228.
18. Read, J. C. A study of the usability of handwriting recognition for text entry by children. *Interacting with Computers 19*, 1 (2007), 57–69.
19. Saldaña, J. *The coding manual for qualitative researchers*. Sage, 2015.
20. Schomaker, L. From handwriting analysis to pen-computer applications. *Electronics & Communication Engineering Journal 10*, 3 (1998), 93–102.
21. Smeeton, N. C. Early history of the kappa statistic, 1985.
22. Welbourn, L., and Whitrow, R. J. A gesture based text editor. In *Proceedings of the Fourth Conference of the British Computer Society on People and computers IV*, Cambridge University Press (1988), 363–371.
23. Wobbrock, J. O., Aung, H. H., Rothrock, B., and Myers, B. A. Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems*, ACM (2005), 1869–1872.
24. Wobbrock, J. O., Morris, M. R., and Wilson, A. D. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2009), 1083–1092.
25. Wolf, C. G., and Morrel-Samuels, P. The use of hand-drawn gestures for text editing. *International Journal of Man-Machine Studies 27*, 1 (1987), 91–102.
26. Zgraggen, E., Zeleznik, R., and Eichmann, P. 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, 2016), 2362–2368.