Selection-Based Mid-Air Text Entry on Large Displays

Anders Markussen, Mikkel R. Jakobsen, and Kasper Hornbæk

Department of Computer Science, University of Copenhagen
Njalsgade 128, Building 24, 5. Floor, 2300 Copenhagen, Denmark
{amark,mikkelrj,kash}@diku.dk

Abstract. Most text entry methods require users to have physical devices within reach. In many contexts of use, such as around large displays where users need to move freely, device-dependent methods are ill suited. We explore how selection-based text entry methods may be adapted for use in mid-air. Initially, we analyze the design space for text entry in mid-air, focusing on single-character input with one hand. We propose three text entry methods: H4 Mid-Air (an adaptation of a game controller-based method by MacKenzie et al. [21]), MultiTap (a mid-air variant of a mobile phone text entry method), and Projected QWERTY (a mid-air variant of the QWERTY keyboard). After six sessions, participants reached an average of 13.2 words per minute (WPM) with the most successful method, Projected QWERTY. Users rated this method highest on satisfaction and it resulted in the least physical movement.

Keywords: Text entry, mid-air interaction techniques, large high-resolution displays, Huffman coding, multitap.

1 Introduction

Devices and interaction techniques for text entry are much researched [24], and it is clear that the effectiveness of text entry is shaped by the context of use. For instance, mobile text entry is different from desktop text entry [22,30], and typing on a tactile keyboard requires little or no visual attention, whereas text entry on a touch surface requires visual attention. Thus, text entry in non-desktop settings presents new challenges and requires new methods [39].

The present paper is motivated by a need to support text entry in one such setting, users working with a large high-resolution display. Large high-resolution displays have been shown to improve productivity [11] and, in contrast to desktop displays, they promote physical movement [3]. Around large displays, users can move in order to navigate, explore, and make sense of data on the display. We seek to design text entry methods that allow users to move in front of the display, without having to hold a device or move to a fixed location to be able to enter text.

Recent research has helped users interact with large displays by supporting object selection and manipulation (e.g., [5,14,19,35]). Mid-air interaction [16], based on tracking of users' hands, may work well for interaction in the context where users move in front of a large display. Vogel and Balakrishnan [35], for instance, used

P. Kotzé et al. (Eds.): INTERACT 2013, Part I, LNCS 8117, pp. 401–418, 2013.

Vicon-tracking to let users point to a large display from a distance and manipulate the cursor; Nancel et al. [27] showed how mid-air gestures can be used to navigate a large display.

Whereas mid-air interactions have been explored for selection and manipulation, they are rarely used for text entry. Prior work approximates mid-air interaction by using devices such as the Nintendo Wiimote [9,33]. Other mid-air text entry techniques include AirStroke [28], a glove- and vision-based method using the Graffiti unistoke alphabet [10]. AirStroke provided a text entry rate 6.5 words per minute (WPM) without word completion. Kristensson et al. [20] demonstrated continuous recognition of mid-air gestures for writing Graffiti letters using a Kinect sensor to detect gestures within a predefined input zone.

We adapt existing selection-based text entry methods to mid-air interaction with large displays. Selection-based methods rely on series of movements and activations of UI components to facilitate text entry. We do so for several reasons: (1) Leveraging familiarity with existing techniques help users learn the techniques faster, which is preferable for walk-up-and-use contexts of large displays. (2) Although mid-air text entry can potentially benefit from the increased expressiveness and additional degrees of freedom of spatial 3D input, simple and effortless techniques is recommended when the user's goal is simple [7]. (3) Despite the potential of more expressive input, the most successful mid-air text entry method to date has to our knowledge been the ray-casting selection-based QWERTY method of Shoemaker et al. [33]. More studies of adaptations of text entry methods from other contexts, such as desktop or mobile computing, are needed in order to establish a base line for mid-air text entry. In order to simplify comparison, we have chosen to focus on single-character input (rather than predictive input) and on one-handed input.

In this paper, we contribute an analysis of the design space for mid-air text entry using a structured approach that enables researchers to relate future analyses to ours. Further, we contribute an evaluation of three mid-air text entry methods that match the context of using large high-resolution displays. The methods we propose are adapted versions of previously successful methods from three different domains; game controller text entry, mobile phone text entry, and a previously successful mid-air text entry method. The methods provide a solid baseline for comparison of future mid-air text entry methods.

2 Design Space for Mid-Air Text Entry

Many considerations in designing for mid-air text entry are similar to those encountered when designing text entry in other contexts; previous work describes them thoroughly (e.g., [18]). Below we therefore focus on design considerations specific to mid-air text entry, aiming to sum up earlier mid-air text entry work in the process; Fig. 1 shows some initial design ideas that we also discuss.

A guiding context of use for the present analysis is work around large high-resolution displays. The scope of our analysis is text entry methods that support input of single characters through hand movement. Although predictive methods can

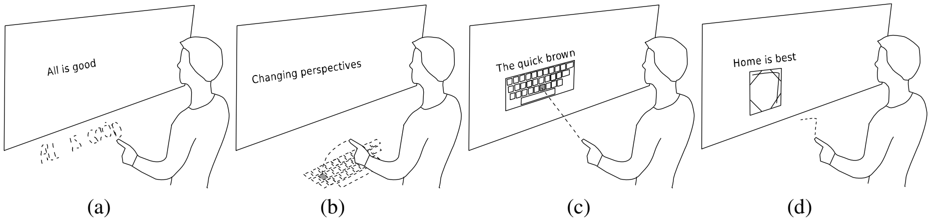

Fig. 1. Some initial ideas for mid-air text entry. (a) shows handwriting in mid-air; (b) shows typing on an imaginary keyboard directly in front of the user; (c) shows ray-casting to a QWERTY keyboard; (d) shows EdgeWrite gestures in mid-air.

perform significantly better, we consider single-character entry a baseline that supports a variety of text entry needs (e.g., entering a code or acronym). Even though writing with coarse body movements is certainly possible (e.g., the photographer Howard Schatz's Body Type), we follow earlier work and focus on movement of the fingers and hands (though only one hand at a time).

We structure the discussion using the design space analysis methods of MacLean et al. [25]. They distinguished questions (about what a design should do), options (answers to questions), and criteria (ways of assessing designs) as three key components for mapping a design space.

2.1 Questions and Options

Q: What Type of Movement? Earlier work has two uses of user's hand movement in mid-air. In gesture-based techniques, users write either freely or using a set of gestures. This is the idea in Fig. 1d and in many other studies [9,18,20]. GesText [18], for instance, uses accelerometer data for text entry. The most successful version used single-depth vertical and horizontal gestures to achieve 5.4 WPM.

In selection-based techniques, users point at symbols laid out in either 2D (e.g., using a QWERTY layout, see Fig. 1c) or 3D (e.g., as in [33]). Shoemaker et al. found better performance and satisfaction with techniques using 2D layouts (QWERTY and circular) compared with a 3D technique where symbols were laid out in a cube [33].

Q: 2D or 3D? Whereas many text entry methods use some form of interaction with a 2D surface (i.e., work on touch screen devices), mid-air interaction provides pitch, yaw and roll [1,9] in addition to position in 3D.

2D-approaches can mimic typing on a surface. Fig. 1b has the users imagine a QWERTY keyboard floating in front of them and use that plane for input; handwriting in mid-air (Fig. 1a) also creates an imaginary surface on which the user writes. Such designs are simple; Benko [6] suggested that we try to achieve the simplicity of touch-enabled devices when designing mid-air interaction techniques. In Kristensson's work [20], Grafitti is used in free-air, but depth (z-distance) does not appear to be used in classifying gestures: effectively, users are writing on a plane.

3D-approaches can use all of the six degrees-of-freedom. However, there seems to be a trade-off between the richness of 6-DoF gestures and an increase in complexity. For instance, the distance to the screen (or away from ones body) could be used for making selections. Research on GesText [18], however, suggested that for accelerometer input, using depth was not efficient.

Q: Typing in Relation to What? Another question is whether to use an explicit point of reference for making gestures or selections. Touch and mid-air interaction differ in that the touch surface can implicitly maintain a point of reference for the user, whereas this is not the case for mid-air interaction. Several options exist:

- Absolute point of reference, such as the display surface (Fig. 1c). Many mid-air input techniques use this approach [27].
- Relative point of reference, which could include the other hand (as in imaginary interfaces, [15]) or the location of ones feet.
- Kinesthetic point of reference, that is, a remembered hand position. For selection-based input using a QWERTY layout (see Fig. 1b), the user might initiate text entry by placing both hands on an imaginary plane; the position of left and right index fingers map to f and j on a virtual keyboard that is transformed to fit the finger placement. While this is attractive, it is well know that human hands drift [26].

Q: Visual Feedback or Not? Given the lack of tactile feedback, typing on a touch surface is primarily supported by visual feedback. In mid-air, visual support is even more challenging to provide, as mid-air text entry at large displays uses indirect input, that is, the input space is separated from the output space [17]. Users may need feedback on tracking of their movements, feedback on movements in relation to recognized gestures or characters, and feedback on production of characters.

Q: How to Initiate and Finish Writing? A well-known challenge in gesture-based input is to identify when gestures start and stop [4]. Specific gestures, pinches, input zones, and so forth has been used to delimit gestures (see for instance [35,36]). A similar challenge for selection-based input methods is to determine when a symbol is activated. In Kristensson's work [20], Grafitti input was delimited to an input zone and gestures were ignored outside this zone.

2.2 Criteria

Intuitiveness, Efficiency and Learnability. For some use contexts, the method for entering text must be easy to learn. For instance, a goal for "walk-up-and-use" systems might be that novices can enter text with minimal introduction, and perform acceptably without practice. Wobbrock et al. [37] and North et al. [29] evaluated intuitive gestures for multi-touch surfaces. The design of mid-air interaction techniques could benefit from similar studies. One approach is to draw on users' experience with widespread text entry methods. For instance, Fig. 1b and Fig. 1c benefit from users' knowledge of the QWERTY layout.

Multi-user Support. With multiple users around a large high-resolution display, text entry methods must satisfy additional criteria. First, users may physically interfere with each other's use of the display (e.g., by blocking the view of the display). Second, physically-based interactions must be social acceptable, else users might avoid physical movement because of fear of looking "silly" [31].

Distance- and Visibility-Dependence. Shoemaker et al. [33] argued that mid-air text entry methods differ in how they are affected by the distance and visibility of the display used for entering text. For instance, Fig. 1b is not distance-dependent, but Fig. 1c is.

Tracking Sensitivity. Many tracking technologies have been used for mid-air interaction, including optical tracking, gyroscopic sensors, and magnetic sensors. Some design options require accurate tracking (e.g., handwriting recognition), whereas others can do with very low tracking precision (e.g., 2D gesture-based input like Fig. 1d).

Effort and Fatigue. The motor effort needed to perform mid-air text entry (e.g., due to imprecise tracking) can be relatively large compared to typing on a keyboard. Extended periods of large movements in mid-air can cause fatigue. One approach to dealing with fatigue is to extend methods for movement minimization [22] to include the full range of body motions involved in mid-air text entry. For instance, text entry methods could be compared a priori on the effort they induce on hands, elbows, and shoulders.

3 Three Candidate Methods

The design space for mid-air text entry methods just outlined is huge. In order to identify candidate text entry methods in the space that are relevant to large high-resolution display interaction, we made two overall decisions. First, we have chosen to focus on selection-based input. Although gesture-based text input may potentially be intuitive and efficient for entering text mid-air, it is difficult to develop competitive text entry performance using existing gesture-based techniques. For instance, [28] reported a mean entry speed of 6.5 WPM for AirStroke without word completion. Second, we limit body movement to reduce fatigue, focusing on movement of hands and fingers.

For all methods, hand tracking is implemented using a glove with reflective markers attached to the back of the hand. A marker tracks the location of the index finger. Differences in angle between the hand's orientation and the vector connecting the location of the hand and the fingertip are used to detect taps. An increase in the angle of more than 5 degrees followed by a decrease of more than five degrees within 500 ms is interpreted as a tap. These thresholds were identified during pilot studies.

Based on the design space presented earlier, we can compare our methods to earlier work. A key design decision is that we use orthogonal projection instead of ray casting, as used for instance by Shoemaker et al [33]. With ray casting, users' movements are magnified as they move away from the display, which results in distance-dependent performance [33]. Given our aim of supporting users in moving in front of a large display, an important criterion was to make text entry performance distance-independent. Our methods work by projecting the hand's location orthogonally onto the display. Motor control is thus unaffected by the user's distance to the display.

3.1 H4 Mid-Air

H4 Mid-Air is an adaptation of H4 Writer [21], a text entry method operated with the thumb using four buttons on a physical game controller. H4 Writer allow users to type 20 WPM after sufficient training [21]. H4 Mid-Air uses the character selection mechanism of H4 Writer, which is based on Huffman coding: To produce a character the user produces the sequence that encodes the character. Visual feedback helps users learn the sequence for different characters.

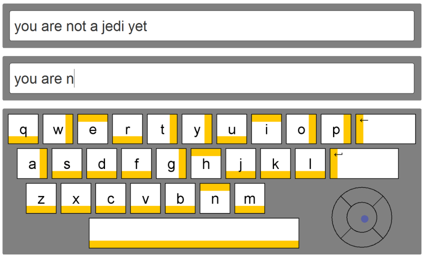

Fig. 2. The user interface for H4 Mid-Air

We adapted the method to mid-air text entry in two ways. First, without a physical controller, the user enters the sequence for a character by moving the hand, which is projected onto the display, to one of four zones and tapping (described in a later section). The motor space is divided into four slices, for each of the four zones, surrounding a "dead" center (4cm in diameter). The slices are open-ended in motor space to make targeting easier. The position of the user's hand relative to the center is shown in an on-screen radar. Second, H4 Mid-Air differs from H4 Writer by providing visual feedback on a QWERTY layout (see Fig. 2). In pilot studies, we tested the visual feedback used by MacKenzie et al. [21]. However, in this method the characters are relocated between the four zones for each step of the selection sequence and we found

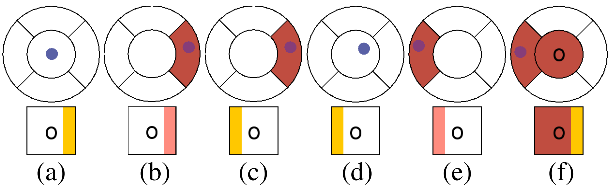

Fig. 3. Production of an 'o' with H4 Mid-Air. In order to produce an 'o', the two step Huffman code "right, left" needs to be completed. This is done through the following sequence of steps: (a) text entry has been initialized (b) hover in the indicated area. Indication is highlighted (c) a tap results in an update of the feedback and the indication changes to the next Huffman code (d) move to next action (e) hover in the indicated area (f) tap produces character. Key and center of radar is highlighted for 500 ms.

that much time was spent scanning for characters. We experimented with visual feedback that minimize scanning time. The resulting feedback is designed to improve learning and to minimize the time spent scanning for characters: (a) the user locates characters in the well-known QWERTY layout instead of the unfamiliar four-zone layout; (b) once a character is found, the user can maintain focus on the character to get visual feedback on the required sequence, without having to rescan after each input action in the sequence. The visual feedback consists of highlighting one of the sides of the key to indicate which zone that needs to be activated next to produce the character. The production of the character 'o' is shown in Fig. 3.

The motor space can be recentered by closing the hand. After 500ms the cursor is removed from the radar. Reopening the hand will define the current location of the hand as the center of motor space and put the cursor at the center of the radar. This allows clutching to redefine the motor space if a position becomes uncomfortable.

3.2 MultiTap

MultiTap shows a reduced keyboard with nine keys and a dot cursor that can be controlled by moving the hand (see Fig. 4). Three or four characters are mapped onto each key. To produce a character, the user taps the corresponding key once or multiple times—the number of taps corresponding to the character's index on that key. The character is produced when a different key is tapped or after a delay of 800ms. For example, tapping twice on "ABC", followed by tapping another key, produces 'B'. The location of the user's hand is projected onto the display plane. However, cursor movement is relative to the center key, which measures 4cm×4cm in motor space. This size was determined through pilot studies. The center is the only key with a fixed size in motor space; other keys are infinitely sized. Clutching results in resetting the motor space and positioning of the cursor at the center of keyboard. As visual feedback, a key is highlighted in orange when the cursor hovers over the key (as shown in Fig. 4). On activation of the key, the background of currently selected character is highlighted in red until the tap timeout expires or until the next activation of the key. A tap timeout of 800ms was found to be appropriate through pilot studies. This timeout is a bit lower than the usual tap timeout of 1-1.5 seconds [8] for similar methods on mobile devices.

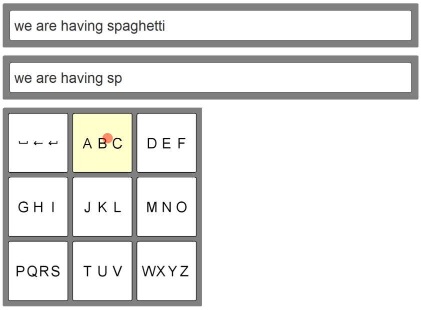

Fig. 4. The user interface for MultiTap

The method aims to provide a visibility-independent text entry method. We thus adjusted the size of the center key so that users are able to home in on the key without visual feedback during a pilot study, we tested that users were able to tap a button of this size with their eyes closed. Our intent is to leverage proprioception in homing in on the center key. Also, this method is inspired by a method that has been used on mobile phones and may thus be familiar to many users.

3.3 Projected QWERTY

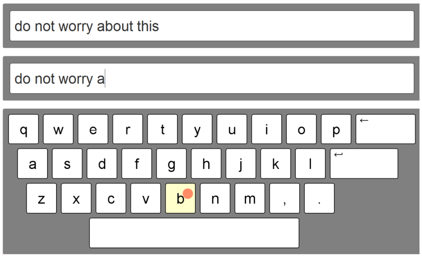

Fig. 5. The user interface for Projected QWERTY

Projected QWERTY shows a standard QWERTY keyboard layout on the display, with a dot cursor that can be controlled by moving the hand (see Fig. 5). A character is produced by moving the hand to a key and tapping. The location of the user's hand is projected onto the display plane. The motor space is 20cm×10cm, about the size of a physical QWERTY keyboard of a small to medium laptop. The keyboard layout was shown in about 50cm×25cm. Pilot studies found these dimensions appropriate. Tracking, tapping and clutching are implemented in the same way as for H4 Mid-Air and MultiTap. Clutching results in resetting the motor space and positioning of the cursor at the center of keyboard. Visual feedback is similar to MultiTap: hovering over a key highlights the key in orange (see Fig. 5); when activated, the key is highlighted for 500ms or until the cursor leaves the key; when producing a character, it is highlighted in red in the transcribed string for 500ms.

Projected QWERTY resembles the QWERTY keyboard technique of [33]. Instead of ray casting from the hand to the screen based on the hand's orientation, Projected QWERTY projects the hand's location onto the display plane. Thereby cursor movement relative to hand movement is independent of distance: in contrast, moving the cursor using ray-casting leads to magnified cursor movement as the user moves away from the display. Another benefit is that the motor space is absolute. Projected QWERTY is thus theoretically visibility-independent, since key locations in motor space are constant. However, taking the number of keys on a QWERTY keyboard into account, actual visibility-independence is not expected.

3.4 Differences among Methods

Next, we describe the main differences among the methods. First, the methods vary in the need for visual feedback on hand movements. Projected QWERTY has small activation areas in motor space, which makes it difficult to target a key without visual feedback. Both H4 Mid-Air and MultiTap use open-ended activation areas, located around the point of reference, which should allow users to point without visual feedback. Text entry without visual feedback assumes that the user has memorized the activation sequences for all characters ([21] suggested it possible for H4 Mid-Air).

Second, the number of buttons/activation areas is varied among methods, which may impact the visibility-dependence of the method. However, it also has an impact on the physical movement needed to operate the methods. H4 Mid-Air requires a sequence of movement and taps, MultiTap requires one movement followed by a sequence of taps, and Projected QWERTY requires only one movement and one tap in order to produce a character. We investigate how these differences affect the amount of hand movement required and the usability of the methods.

Third, the methods aim to ease adoption by novices. Using the well-known alphabet layouts for Projected QWERTY and MultiTap should help novices adopt the methods faster. Although H4 Mid-Air leverages the QWERTY layout, it does not benefit from previous user experience as well as the other methods. However, the H4 Huffman coding has proven fast in previous longitudinal studies [2,21].

4 Empirical Study

To evaluate the three mid-air text entry methods, we conducted a controlled experiment. The experiment spanned six sessions in which participants used the methods to transcribe sentences.

4.1 Participants

Six participants (one female) were recruited; ages ranged from 21 to 28. One participant performed the experiment left-handedly. None of the participants were native English speakers, but all participants rated their level of English between good (2) and fluent (6).

4.2 Apparatus

A 2.80m×1.20m display containing 7680×3240 pixels was used. The display is back-projected by 12 projectors that are arranged as tiles in a 4×3 layout. Participants stood 2 meters away from the display while transcribing sentences.

For tracking, we used the OptiTrack (http://www.naturalpoint.com/optitrack) motion capture system equipped with 24 V100:R2 cameras. The system provides tracking data at 100 fps. The tracking precision was ±4 mm over the entire tracking volume; participants were located in a part of the volume with higher precision. Although the OptiTrack system is expensive, we decided to use it for several reasons. First, affordable tracking systems available at the time of this study have low precision. High-precision tracking reduces noise in the data and thus gives confidence that we are measuring the performance of the techniques, and not effects of noise caused by current tracking equipment. Second, affordable systems (e.g., Microsoft Kinect) have limited fields of view and require the user to interact at certain, constrained distances. This limits users' ability to move freely, which is needed around large displays. Third, tracking technology is improving at a high rate. The present study can be replicated and the text entry methods practically applied with widespread equipment within a few years; use of high-precision tracking thus ensures a better baseline for future research.

4.3 Tasks

Users were asked to transcribe randomly selected phrases from the MacKenzie and Soukoreff corpus [23]. Sentences were transcribed as unconstrained text entry [38].

Consequently, users were allowed, but not forced, to delete previously entered text and correct any errors that they noticed. Participants were instructed to complete sentences as quickly and accurately as possible.

4.4 Design

The experiment was conducted as a within-subjects design with the three text entry methods (H4 Mid-Air, MultiTap, and Projected QWERTY) and text entry session as independent variables. Dependent measures were text entry speed, error rate, physical hand movement, and subjective satisfaction: they are detailed in the next subsection.

Participants completed 6 sessions. During each session, participants transcribed text with all text entry methods, completing 2 blocks of 5 sentences with each method. The order in which the methods were used was fully counterbalanced between participants and sessions. In all, 1080 phrases were transcribed (6 participants × 6 sessions × 3 text entry methods × 2 blocks × 5 phrases).

4.5 Data Collection and Analysis

We collected data describing participants' interaction with the methods, including data on the location of their hand. From these data we used StreamAnalyzer [38] to calculate text entry speed and error rate, and derived a measure of hand movement. Text entry speed was calculated using equation 1, where |T| is the length of the transcribed string and S is the time in seconds from the entry of the first character to the entry of the last character.

$$WPM = \frac{|T|-1}{S} \times 60 \times \frac{1}{5} \;, \tag{1}$$

Error rate was calculated using the methods described by Soukoreff and MacKenzie [34], as Minimum String Distance (MSD), Uncorrected Errors (ErrUC) and Corrected Errors (ErrC).

To provide a quantitative measure of the physical effort put into typing, we defined Hand Movement Per Word (HMPW). HMPW is calculated as the sum of the distances travelled by the hand between tracking frames. Calculating the sum of distances over data containing noise may result in erroneous values for HMPW. We therefore ran the Douglas-Peucker algorithm [12] on the movement data with a threshold of 2mm in order to minimize noise. HMPW is measured in meters per word in the transcribed string; one word is five characters including whitespaces. As with WPM, HMPW is measured from the entry of the first character to the entry of the last character and is calculated as follows:

$$HMPW = \frac{|HM|*5}{|T|-1} \tag{2}$$

|HM| is the sum of distances between consecutive tracking frames and |T| is the length of the transcribed string. We removed the first phrase with each text entry method from this calculation because it was typed slower and with more hand

movements than subsequent sentences, and because participants said they used the first phrase to get used to a method.

Subjective satisfaction was measured using three instruments at various stages throughout the experiment: (1) To get an estimate of how much effort participants had to put into the operation of a text entry method, SMEQ [32] was administered to participants each time they had finished using a text entry method. (2) We adapted thirteen questions from the ISO-9241-9 standard [13] to evaluate physical operation, fatigue and comfort, speed and accuracy, and overall usability. We administered these questions at each participant's first and last session to gauge their experience with the methods after little training and after some training. (3) After the first and last session, participants ranked the three text entry methods (1 being the one they liked the most, 3 being the one they liked the least).

4.6 Procedure

At the first session, participants were introduced to the concept of mid-air text entry and to the three methods being evaluated. Participants were then allowed to practice with each method. They were asked to practice until they felt confident with using the method, and felt they would be able to reproduce any randomly chosen character. No participant entered more than 4 sentences per text entry method during practice.

At the beginning of sessions 2 to 6, participants were asked to practice with each method until they felt confident that they would be able to reproduce any randomly chosen character. Typically, participants practiced only one sentence to reacquaint themselves with a method. Session lasted 45 minutes on average.

In each session, participants completed two blocks of five sentences with each text entry method. Participants were allowed a short break after each block. After having transcribed the two blocks of sentences with a method, participants were administered an electronic SMEQ; at the end of session 1 and 6 participants answered the ISO questionnaire and ranked interfaces by order of preference.

5 Results

A 3 (text entry method) × 6 (session) repeated measures analysis of variance was performed on the entry speeds, the error rates, and the measure of hand movement. Significant effects were examined using Bonferroni-corrected pairwise comparisons.

5.1 Text Entry Speed

Fig. 6 shows the text entry speed in words per minute (WPM) for the text entry methods across sessions. We found a main effect for text entry method, $F(2, 10) = 109.63$, $p < .01$. Pairwise comparisons showed that Projected QWERTY ($M = 11.63$, $SD = 2.29$) was faster than MultiTap ($M = 8.38$, $SD = 2.45$), $p < 0.01$, which again was faster than H4 Mid-Air ($M = 4.19$, $SD = 1.25$), $p < 0.01$. In the final session, Projected QWERTY achieved 13.2 WPM (*SD* = 1.55), MultiTap almost 9.5 WPM (*SD* = 2.19), and H4 Mid-Air 5.2 WPM (*SD* = 1.31).

A main effect was also found for session, $F(5, 25) = 176.22$, $p < .01$, showing that users improved over sessions. Speed improved from first to last session by 39% for Projected QWERTY, 47% for Multi-Tap, and 80% for H4 Mid-Air, all significant at the $p < .001$ level. No significant interaction was found between method and session.

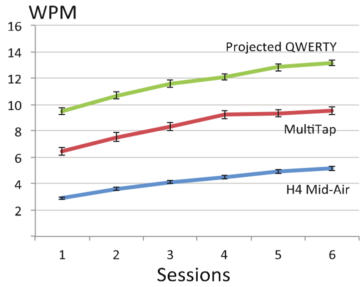

Fig. 6. Mean words-per-minute (WPM) for methods over sessions. Error bars show standard error of the mean

5.2 Error Rate

Fig. 7 shows the error rate measured as Minimum String Distance (MSD), Uncorrected Errors (ErrUC), and Corrected Errors (ErrC). Text entry method was found to have a significant effect on ErrC, $F(2, 10) = 9.14$, $p < .01$, but not on MSD, $F(2, 10) = 1.52$, $p = .27$, or ErrUC, $F(2, 10) = 1.46$, $p = .28$. Pairwise comparisons showed more corrected errors with MultiTap ($M = 5.7\%$, $SD = 6.6\%$) than with Projected QWERTY ($M = 2.8\%$, $SD = 4.6\%$), $p < .05$. No significant different were found between MultiTap and H4 Mid-Air ($M = 4.0\%$, $SD = 4.9\%$), $p = .35$, or between H4 Mid-Air and Projected QWERTY, $p = .20$. Session was also found to have a significant effect on ErrC, $F(2.175, 10.874) = 4.19$, $p < .05$, but not on MSD, $F(1.853, 9.264) = 2.205$, $p = .17$, or ErrUC, $F(1.922, 9.609) = 1.962$, $p > .19$. Overall, ErrC declines over sessions, which was expected because participants make fewer errors as they become increasingly familiar with a text entry method.

5.3 Subjective Measures

Fig. 8 shows the SMEQ scores across input methods and sessions; recall that lower SMEQ scores represent lower mental effort. Mental effort differs significantly across input methods, $F(2,10) = 6.25$, $p < .05$. Pairwise comparisons showed that mental

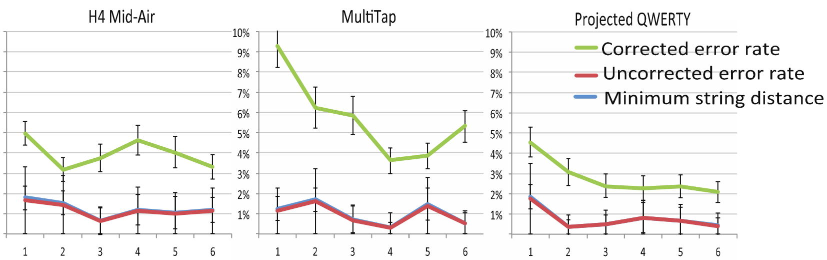

Fig. 7. Error rates (mean) over sessions and text entry method. Error bars show standard error of the mean.

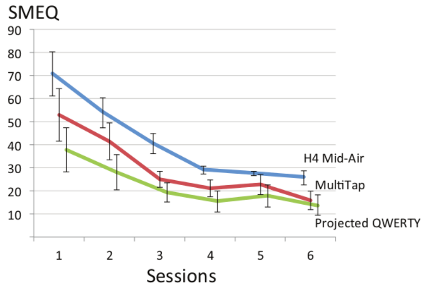

Fig. 8. Mean SMEQ scores over sessions for the three text entry methods. Error bars show standard error of the mean.

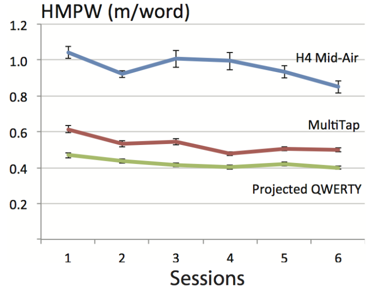

Fig. 9. Mean HMPW for the three text entry methods over sessions. Error bars show standard error of the mean.

effort with Projected QWERTY is lower than with H4 Mid-Air. In the final session, MultiTap and Projected QWERTY lead to comparable mental effort. As expected, we found a significant effect of session, $F(5, 25) = 17.19$, $p < .01$. In particular H4 Mid-Air improved; from session 1 to 6, its SMEQ ratings decreased with 63%.

Overall, we found a significant effect of input method on the 12 satisfaction questions, $F(2, 10) = 5.63$, $p < .05$; session and interaction between input method and session were not significant. Key factors in the main effect of input method were perceived difficulty of use (H4 Mid-Air is perceived as more difficult to use and more uncomfortable to use) and mental effort (H4 Mid-Air is perceived as requiring more effort).

Participants' ranking of input methods showed a clear pattern. In the first session, five participants ranked Projected QWERTY first, four ranked MultiTap second, and four ranked H4 Mid-Air last. This is a significant difference, $\chi 2$ (4, N = 6) = 12.5, $p < .05$. This pattern changed only minimally in the last session, where just one participant changed ranking, giving Projected QWERTY first for four participants (H4 Mid-Air first for two). The ranking was still significant, incidentally with the same $\chi 2$ score.

Participants' comments after each session and in a comparison of methods after session 6 support the above data. Generally, users commented that H4 Mid-Air was hard to use and tired them. Two participants, however, noted that H4 Mid-Air was fun and made them think of typing as a game.

5.4 Hand Movement

Fig. 9 shows average hand movement (HMPW) for the text entry methods across sessions. We found a main effect for text entry method, $F(1.018, 5.092) = 31.95$, $p < .01$, but not for session, $F(2.339, 11.696) = 3.020$, $p = .08$. Pairwise comparisons showed more hand movement for H4 Mid-Air ($M = 0.96$ m, $SD = 0.28$ m) than for MultiTap ($M = 0.53$ m, $SD = 0.12$ m), which in turn was more than for Projected QWERTY ($M = 0.42$ m, $SD = 0.09$ m), all with $p < .05$.

The three methods differ in their use of motor space. Projected QWERTY uses absolute mapping of all buttons into rectangular areas in motor space; H4 Mid-Air

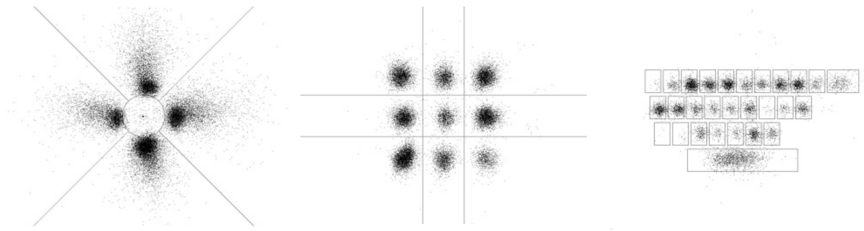

Fig. 10. Distribution of taps in motor space for each text entry method: H4 Mid-Air (left), MultiTap (middle), Projected QWERTY (right). Motor space proportions are identical for the interfaces. Activation zones are indicated by gray lines.

divides the motor space into four slices surrounding a 4cm dead area; MultiTap separates the motor space into a central 4cm x 4cm square activation zone, surrounded by eight infinite activation zones.

Fig. 10 shows a plot of all taps for the three text entry methods. Even though H4 Mid-Air and MultiTap both have open-ended activation spaces, they used motor space differently. For H4 Mid-Air, the shape of the tap-cloud indicate that users are aiming for the activation space close to the center, but often overshooting takes place, resulting in taps in the parts of the activation area further away from the center. Perhaps users may be relying less on the visible feedback of the radar and more on proprioception. However, tap-clouds for MultiTap indicate that participants targeted the center of the keys in the on-screen keyboard. We hypothesize that this difference is primarily related to the visual design: using buttons may discourage the use of open-ended activation areas. This suggests that the radar feedback of H4 Mid-Air facilitates fast, but inaccurate movements that could potentially be based on muscle memory rather than visual feedback. In contrast, the button-based design of the MultiTap keyboard may motivate users to perform accurate pointing and tapping rather than quick movements based on muscle memory.

During the experiment, several participants commented that the movement required could be reduced for H4 Mid-Air and MultiTap without loss of text entry performance. In Fig. 10, we see that the tap-clouds for both H4 Mid-Air and MultiTap are clearly separated by areas with few taps. This is less pronounced for Projected QWERTY. It seems that taps on the keys of Projected QWERTY occurs everywhere on the keys. Taking the low error rate and relatively good performance of Projected QWERTY into account, we see no reason to suspect that the motor space of Projected QWERTY is too small for participants to tap. Rather, this pattern could indicate that the button size in Projected QWERTY is approaching a lower limit for the current tracking precision; using this technique with lower-precision equipment (e.g., Kinect) could result in poor performance. Instead, we hypothesize that a reduction of the motor space movement required for H4 Mid-Air and MultiTap could reduce HMPW and potentially improve text entry performance. We do however note that a reduction of motor space would potentially impact the level of visibility-independence of the text entry methods.

6 Discussions

We first discuss our results in relation to design options and criteria in mid-air text entry. Then we discuss potential improvements to each of the methods.

6.1 Results and Design Space

The criteria that the methods were designed for impacts their performance. First, MultiTap and H4 Mid-Air, both designed to work without visual feedback, perform significantly worse than Projected QWERTY in terms of speed; MultiTap also has a higher error rate. This suggests a trade-off between visibility-dependence and performance. However, further empirical studies are needed to actually show whether skilled users can use MultiTap and H4 Mid-Air without feedback. Surprisingly, we did not see any benefit from the open-ended activation areas of MultiTap. On the contrary, users seemed to perform accurate pointing and tapping within the motor space of the keys.

Second, the number of buttons is likely to affect the results. MultiTap had fewer but larger buttons (in motor space) than Projected QWERTY, which might explain why more hand movements were found for MultiTap. We would expect an increase in performance with MultiTap if the activation areas were smaller. However, this could have a detrimental effect on the method's visibility-dependence. Given the wide distribution of taps with H4 Mid-Air (see Fig. 10), we hesitate to make similar speculations about reducing the activation areas for H4 Mid-Air.

6.2 Improving the Methods

The H4 Mid-Air technique did not work well in the present study. It achieved only an average of 5.2 WPM in the last session, compared to 20.4 WPM in the paper describing H4 Writer [21], and 14.0 WPM in a glove based study [2]. It is worth noting, however, that the number of sessions and transcribed phrases per participant in our study were significantly lower than in these two studies. Users' satisfaction was the lowest for H4 Mid-Air among the methods we explored, though SMEQ scores dropped by 63%.

In our view, H4 Mid-Air performance may be improved in several ways. First, the distribution of tap points is elliptical (rather than circular as for the other methods). Thus, users moved their hands much more than they had to, as also shown by the HMPW measure. Second, the original H4 Writer used Huffman coding on four possible choices because a four-button device was used. We can do a HX Mid-Air, where X is the number of discrete zones that the user can actuate in mid-air; it is not clear that four is the right number. For instance, Fig. 10 suggests plenty of space to do H8, resulting in reduced input zones in motor space and shorter Huffman codes for each character. Third, tapping was used to write a character, but as mentioned in the section on design space, many other options exist (e.g., pinching, using depth). Fourth, we reiterate that the feedback method for H4 Mid-Air was designed to facilitate walk-up-and-use. We have not compared it to the feedback in the H4 Writer system [21].

The MultiTap method performed quite good, achieving an average of 9.5 WPM on the last session; almost identical to the 10 WPM performance of typical MultiTap implementations for mobile phones (without text prediction) [24]. One way to improve MultiTap is to minimize the time spent on timeouts between taps; earlier studies have attempted to do this by using an extra button to skip the timeout. However, adding buttons to the MultiTap interface would reduce some of the potential benefits with regards to visibility-independence. Other improvements that do not impact visibility-independence could be the use of bimanual interaction or depth information, even though that was found too complicated with an accelerometer in GesText [18]. Interestingly, MultiTap seems to be successful in generating a feeling of buttons-in-the-air, which means that users make less movement to hit an area (in contrast, tap distributions in H4 Mid-Air were elliptical). As previously mentioned, a reduction in used motor space could also result in performance improvements for MultiTap.

The Projected QWERTY method achieved high text entry rates (13.2 WPM in the final session) and was the preferred system among users. We have a few ideas for further improving this technique. First, the size of the motor space for controlling Projected QWERTY was determined through pilot studies, but it might be further optimized. Second, Projected QWERTY could easily be extended to support input for two hands, which could dramatically increase performance.

6.3 Next Steps

Our results indicate that the adaptions of successful text entry methods from other domains are indeed possible, and that some of these adapted techniques can provide acceptable text entry speeds. This paper provides a performance baseline of how a set of adapted text entry methods from other domains may perform in mid-air. Based on the experiences from our experiments, we agree with Shoemaker et al. [33] that the terms distance- and visibility-dependence describe important properties of mid-air text entry well, and we see a need to further study how aspects of the mid-air design space affect these properties.

We are aware that the availability of 6-DoF and precise tracking could potentially open up for new and innovative text entry methods. The guidelines for designing character-entry systems in 3D user interfaces presented by Bowman et al. [7] combined with detailed empirically based analyses of the mid-air design space, such as that attempted in the present paper, provides starting points for future research.

Acknowledgements. This work has been supported in part by the Danish Council for Strategic Research, grant 10-092316.

References

1. Bailly, G., Vo, D.-B., Lecolinet, E., Guiard, Y.: Gesture-aware remote controls. In: Proc. ICMI 2011, pp. 263–270. ACM, New York (2010)
2. Bajer, B., MacKenzie, I.S., Baljko, M.: Huffman Base-4 Text Entry Glove (H4 TEG). In: Proc. ISWC 2012, pp. 41–47. IEEE (2012)

3. Ball, R., North, C., Bowman, D.A.: Move to Improve: Promoting Physical Navigation to Increase User Performance with Large Displays. In: Proc. CHI 2007, pp. 191–200. ACM, New York (2007)
4. Baudel, T., Beaudouin-Lafon, M.: Charade: remote control of objects using free-hand gestures. Communications of the ACM 36(7), 28–35 (1993)
5. Baudisch, P., Cutrell, E., Robbins, D., et al.: Drag-and-Pop and Drag-and-Pick: Techniques for Accessing Remote Screen Content on Touch- and Pen-operated Systems. In: Proc. Interact 2003, pp. 57–64 (2003)
6. Benko, H.: Beyond flat surface computing: challenges of depth-aware and curved interfaces. In: Proc. MM 2009, pp. 935–944. ACM, New York (2009)
7. Bowman, D.A.: 3D user interfaces: theory and practice. Addison-Wesley (2005)
8. Butts, L., Cockburn, A.: An evaluation of mobile phone text input methods. In: Proc. AUIC 2002, vol. 7, pp. 55–59. IEEE, Los Alamitos (2002)
9. Castellucci, S.J., MacKenzie, I.S.: Unigest: text entry using three degrees of motion. In: Proc. CHI 2008, pp. 3549–3554. ACM, New York (2008)
10. Castellucci, S.J., MacKenzie, I.S.: Graffiti vs. unistrokes: an empirical comparison. In: Proc. CHI 2008, pp. 305–308. ACM, New York (2008)
11. Czerwinski, M., Smith, G., Regan, T., et al.: Toward characterizing the productivity benefits of very large displays. In: Proc. Interact 2003, pp. 17–24 (2003)
12. Douglas, D.H., Peucker, T.K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica 10(2), 112–122 (1973)
13. Douglas, S.A., Kirkpatrick, A.E., MacKenzie, I.S.: Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In: Proc. CHI 1999, pp. 215–222. ACM, New York (1999)
14. Forlines, C., Vogel, D., Balakrishnan, R.: HybridPointing: fluid switching between absolute and relative pointing with a direct input device. In: Proc. UIST 2006, pp. 211–220. ACM, New York (2006)
15. Gustafson, S., Bierwirth, D., Baudisch, P.: Imaginary Interfaces: Spatial Interaction with Empty Hands and Without Visual Feedback. In: Proc. UIST 2010, pp. 3–12. ACM, New York (2010)
16. Hilliges, O., Izadi, S., Wilson, A.D., Hodges, S., Garcia-Mendoza, A., Butz, A.: Interactions in the air: adding further depth to interactive tabletops. In: Proc. UIST 2009, pp. 139–148. ACM, New York (2009)
17. Hinckley, K., Wigdor, D.: The Human-Computer Interaction Handbook, pp. 95–132. CRC Press (2012)
18. Jones, E., Alexander, J., Andreou, A., Irani, P., Subramanian, S.: GesText: accelerometer-based gestural text-entry systems. In: Proc. CHI 2010, pp. 2173–2182. ACM, New York (2010)
19. Khan, A., Fitzmaurice, G., Almeida, D., Burtnyk, N., Kurtenbach, G.: A remote control interface for large displays. In: Proc. UIST 2004, pp. 127–136. ACM, New York (2004)
20. Kristensson, P.O., Nicholson, T., Quigley, A.: Continuous recognition of one-handed and two-handed gestures using 3D full-body motion tracking sensors. In: IUI 2012, pp. 89–92. ACM, New York (2012)
21. MacKenzie, I.S., Soukoreff, R.W., Helga, J.: 1 thumb, 4 buttons, 20 words per minute: design and evaluation of H4-writer. In: Proc. UIST 2011, pp. 471–480. ACM, New York (2011)
22. MacKenzie, I.S., Soukoreff, R.W.: Text Entry for Mobile Computing: Models and Methods, Theory and Practice. Human–Computer Interaction 17(2), 147–198 (2002)

23. MacKenzie, I.S., Soukoreff, R.W.: Phrase sets for evaluating text entry techniques. In: Proc. CHI 2003, pp. 754–755. ACM, New York (2003)
24. MacKenzie, I.S., Tanaka-Ishii, K.: Text Entry Systems: Mobility, Accessibility, Universality. Morgan Kaufman (2007)
25. MacLean, A., Young, R., Bellotti, V., Moran, T.P.: Questions, Options, and Criteria: Elements of Design Space Analysis. Human–Computer Interaction 6(3), 201–250 (1991)
26. Makin, T.R., Holmes, N.P., Ehrsson, H.H.: On the other hand: dummy hands and peripersonal space. Behavioural Brain Research 191(1), 1–10 (2008)
27. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., Mackay, W.: Mid-air pan-and-zoom on wall-sized displays. In: Proc. CHI 2011, pp. 177–186. ACM, New York (2011)
28. Ni, T., Bowman, D., North, C.: AirStroke: bringing unistroke text entry to freehand gesture interfaces. In: Proc. CHI 2011, pp. 2473–2476. ACM, New York (2011)
29. North, C., Dwyer, T., Lee, B., Fisher, D., Isenberg, P., Robertson, G., Inkpen, K.: Understanding Multi-touch Manipulation for Surface Computing. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 236–249. Springer, Heidelberg (2009)
30. Oulasvirta, A., Tamminen, S., Roto, V., Kuorelahti, J.: Interaction in 4-second bursts. In: Proc. CHI 2005, pp. 919–928. ACM, New York (2005)
31. Rico, J., Brewster, S.: Usable gestures for mobile interfaces: evaluating social acceptability. In: Proc. CHI 2010, pp. 887–896. ACM, New York (2010)
32. Sauro, J., Dumas, J.S.: Comparison of three one-question, post-task usability questionnaires. In: Proc. CHI 2009, pp. 887–896. ACM, New York (2009)
33. Shoemaker, G., Findlater, L., Dawson, J.Q., Booth, K.S.: Mid-air text input techniques for very large wall displays. In: Proc. GI 2009, pp. 231–238. Canadian Information Processing Society, Toronto (2009)
34. Soukoreff, R.W., MacKenzie, I.S.: Metrics for text entry research: an evaluation of MSD and KSPC, and a new unified. In: Proc. CHI 2003, pp. 113–120. ACM, New York (2003)
35. Vogel, D., Balakrishnan, R.: Distant freehand pointing and clicking on very large, high resolution displays. In: Proc. UIST 2005, pp. 33–42. ACM, New York (2005)
36. Wilson, A.D.: Robust computer vision-based detection of pinching for one and two-handed gesture input. In: Proc. UIST 2006, pp. 255–258. ACM, New York (2006)
37. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proc. CHI 2009, pp. 1083–1092. ACM, New York (2009)
38. Wobbrock, J.O., Myers, B.A.: Analyzing the input stream for character- level errors in unconstrained text entry evaluations. TOCHI 13(4), 458–489 (2006)
39. Zhai, S., Kristensson, P.-O., Smith, B.A.: In search of effective text input interfaces for off the desktop computing. Interacting with Computers 17(3), 229–250 (2005)