

QB-Gest: Qwerty Bimanual Gestural Input for Eyes-free Smartphone Text Input

Linghui Ye¹, Frode Eika Sandnes^{1,2[0000-0001-7781-748X]} and I. Scott MacKenzie³

¹ Department of Computer Science, Oslo Metropolitan University, N-0130 Oslo, Norway

² Institute of Technology, Kristiania University College, N-0153 Oslo, Norway

³ Department of Computer Science, York University, Toronto, Ontario, Canada, M3J 1P3
yelinghui1987@gmail.com, frodes@oslomet.no, mack@cse.yorku.ca

Abstract. We developed QB-Gest, a bimanual text entry method based on simple gestures where users drag their thumbs in the direction of the desired letter while visualizing the Qwerty-layout. In an experiment with four sessions of testing, 20 users achieved text entry rates of 11.1 wpm eyes-free and 14.1 wpm eyes-on. An expert user achieved an eyes-free rate of 24.9 wpm after 10 rounds of entering the-quick-brown-fox phrase. The method holds potential for users with low vision and certain types of reduced motor function.

Keywords: mobile text entry, eyes free text, gestures, Qwerty.

1 Introduction

Smartphones have become an important tool in modern society by facilitating communication independent of time and place. Many smartphone tasks require text input, such as searching the web, sending emails, or messaging. Text input typically uses a default virtual keyboard. Yet, many users find it hard to use virtual smartphone keyboards compared to physical desktop keyboards [1], because input requires accurately hitting keys without tactile feedback. Virtual smartphone keys are smaller than physical desktop keys; so, input is both visually intensive and requires careful eye-motor coordination [2]. Techniques that require limited visual demand, such as touch typing, are desirable in many situations. Eyes-free operation is an absolute necessity for blind users [3, 4]. In addition to situational impairments, visual acuity and pointing accuracy reduce with age. Another goal is to free up the valuable display real estate occupied by soft keyboards for displaying other information.

Gestures have been proposed as a means of reducing visual demand in text entry tasks [5, 6, 7]. Instead of hitting keys at absolute positions, the user inputs simple gestures as relative finger movements. Such gestures, often resembling the graphical shape of letters, can be input eyes-free. Usually, gesture alphabets must be learned, and this requires the user to invest effort. In the proposed approach we instead rely on users' familiarity with the Qwerty layout, thus reducing the need to learn new gestures. Users visualize the Qwerty layout and move a thumb in the direction of the required character within either the left or right half of the layout (see Fig. 1).

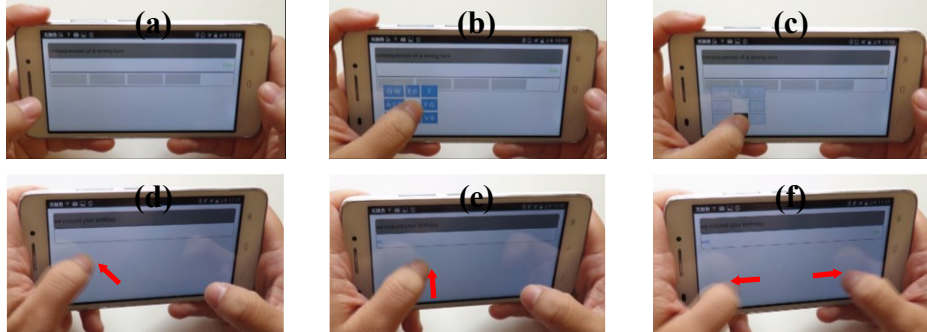


Fig. 1. Top: Input of *c* in eyes-on mode by (a) pressing left thumb, (b) displaying the left side of the Qwerty keyboard centered around the thumb, (c) dragging the thumb down towards *c* and releasing the thumb. Bottom: Inputting the word *we* in eyes-free mode by (d) the left thumb northwest for *w*, (e) left thumb north for *e*, (f) both thumbs outwards for *space*.

High text entry rates can be achieved when both hands are used collaboratively to input text. Expert smartphone users often input text with two thumbs [8]. However, it is difficult to input complex gesture shapes bimanually as users typically need their dominant hand for fine motor tasks. Since the proposed method only requires simple directional gestures it is envisaged that bimanual input is possible. An experiment was designed to determine if users who are familiar with Qwerty can visualize the layout and produce the corresponding directional gestures. These observations were contrasted against a group of users who used the technique with visual feedback.

In the next section we review work related to smartphone text entry. This is followed by a description of the proposed QB-Gest prototype. Then, we describe a user study to test the prototype followed by an analysis of the results.

2 Related work

Qwerty is the most common layout on smartphones, although alternatives have been proposed [9]. Since users resist change [10, p. 187], research focuses on compromises that leverage users' familiarity and the reduced visual search afforded by both alphabetical [11] and Qwerty [12] layouts. Other work experimented with adding keys dynamically according to current input using a language model [13].

Virtual or soft keyboards on smartphones are small with limited space for each key. Key width is especially small when the device is in portrait mode. Clearly, to successfully hit small keys, visual cues are required for key locations. Ordinary smartphone virtual keyboards are a challenge for individuals with low vision and or reduced motor function [14]. As vision and motor control reduce with age, smartphone text entry is particularly challenging for older individuals [1].

Input on a virtual keyboard is a pointing task where users move a finger or stylus to hit a key. With physical keyboards, the fingers feel the keys before acting; with practice, ten-finger touch-typing is possible. However, virtual keyboards have no such

tactile feedback: The first touch is recorded as a keypress. Input using virtual keyboards is therefore a visually intensive target acquisition task. Fitts' law predicts the relationship between the distance to the target, the width of the target, and the speed [15]; hence, the faster the pointing task is performed, the less likely is the user to hit the target. With smaller targets, accurate target acquisition requires slowing down. Hence, Fitts' law explains why smartphone text entry is slower than using ordinary physical keyboards.

Experiments on ten finger touch-typing on touch surfaces have used Markov-Bayesian models [16]. Although entry rates were high, at 45 wpm, achieving adequate accuracy was a challenge. Attempts have also been made to input text through eyes-free thumb input relying on motor memory [17]. Again, accuracy is a challenge.

The small size of smartphones means that ten finger touch typing is not possible. Instead, input uses a single index finger, two thumbs, or a single thumb. Azenkot and Zhai report that text entry rates reach approximately 50 wpm with two thumbs, 36 wpm with a single index finger, and 33 wpm with one thumb [8]. Two thumbs are faster as the keyboard area is divided in two and each thumb traverses shorter distances. Two thumb input is also associated a higher error rate (11%) compared to using a single index finger (8%) and or a single thumb (7%). The authors also found that the participants hit consistently below the targets, with larger horizontal offsets at the edges. This implies that participants tend to minimize the finger distance travelled. Approaches for supporting input on tiny virtual keyboards include zooming in using a callout where an enlarged version of the acquired key is displayed above the finger. Callouts may also include neighboring keys [2].

Zhu et al. [18] explored if users were able to enter text on a smartphone eyes-free using their memory of the Qwerty layout. Experiments with their Invisible Keyboard achieved 31.3 wpm during the first session and 37.9 wpm during the third session without visual feedback. An invisible smartphone keyboard relies on skill transfer from smartphone text entry. It is not likely that ordinary Qwerty typing skills will transfer to the smartphone form factor. Text entry methods such as Apple's VoiceOver or Android's TalkBack support blind or low-vision individuals. Users move their finger over the keyboard and receive speech feedback to explore the layout until the desired letter is spoken. One of the first accounts of this technique is found in the Slide Rule project [19]. Such exploratory strategies are slow with a study on blind users revealing a text entry rate of 2.1 wpm with VoiceOver using a Qwerty layout [4]. Proposed improvements include two-handed input with separate male and female voices for each hand [20] and using pseudo-force to separate exploration and taps, where the touch area of the fingertip is used as a measure of finger pressure [6]. An alternative to keyboard input is speech, with rates about 5× faster than VoiceOver [3]. Moreover, speech input was preferred by blind users [3].

There are comparatively fewer studies addressing smartphone text entry for individuals with reduced motor function. Graphical approaches such as Dasher [21] are visually intensive. The literature on text entry with reduced motor function is dominated by keyboard scanning where virtual keyboards are automatically traversed in regular steps and once the desired key or key-group is highlighted a selection is made with a single key or switch [22]. Hence, the user only makes a time-dependent selec-

tion and there is no need to hit an absolute target. Clearly, scanning is slow compared to other techniques. Chording [23, 24] offers potentially high text entry rates and is suitable for certain types of reduced motor function, such as loss of a hand, and can also be performed eyes-free. Chording has been used for eyes-free numeric input on smartphones [5]. Unfortunately, learning the chords is required and chording is therefore not widely adopted despite effective mnemonic aids such as using the graphical shape of letter [24] and Braille codes [25].

Another avenue of text entry research is gestures [26, 27]. Gestures are simple dragging motions performed on the touch display. The motions are relative and do not require hitting specific targets. Thus, gesture input is applicable to users with low or no vision. Also, gestures can be employed by users with reduced motor function who are unable to accurately hit targets. Approaches such as Graffiti [26, 28] and Edge-Write [27, 29] rely on gestures that resemble the shape of letters. It is easier to remember gestures that mimic symbols users already know. On the other hand, these gestures can resemble complex shapes with twists and turns. Simpler gestures such as UniStrokes are faster (15.8 wpm) than the more complex Graffiti gestures (11.4 wpm) as there is less distance for the fingers to travel [28].

Attempts exist to input text using single-stroke gestures, for example navigating menu hierarchies to retrieve a specific letter using multiple simple gestures [30, 31]. With Swipeboard [31] the user selects letters by navigating menus on very small touch displays using multiple swipe gestures. Absolute pointing tasks and gestures have also been combined, such as Swype, where the user drags the finger along the keys of a keyboard using a continuous stroke to produce a word shape. Bimanual mechanisms have been explored but are slower than the methods relying on one hand [32]. Lai et al. [33] investigated if simple gestures with audio feedback could be used for entering text eyes free on a smartphone using one hand. Their system, Thumb-Stroke, presented letters alphabetically in groups around a circle with the thumb dragged in the direction of the desired letter. After 20 sessions users reached 10.85 wpm eyes-free. Banovic et al. [34] combined absolute pointing with relative gestures for eyes-free text entry. With their Escape-Keyboard, the user first points at one of four regions on the display, then performs a simple gesture in one of eight directions. Input used one thumb. Users were given audio feedback and after six training sessions reached 7.5 wpm.

A study where a physical keyboard was compared to a virtual keyboard and gestures [35] found that the physical keyboard (48.6 wpm) was more than 2× faster than the virtual on-screen keyboard (21.0 wpm), which again was faster than using gestures (16.5 wpm). As found in performance measures with smartphones [8], the virtual keyboard and gestures were input with one hand, while the keyboard task was conducted bimanually. However, an 8-inch tablet was used instead of a smartphone.

The study reported herein allows text entry with simple gestures. It was inspired by a text entry method proposed for dual joystick game controllers [36], where the user visualizes the Qwerty keyboard. If the desired key is on the left side of the keyboard, the user uses the left hand and left joystick, and vice versa. In inputting a specific letter the user imagines that the left and the right joysticks are located between D-F and J-K, respectively. To retrieve a specific letter, the respective joystick is moved in

the direction of the letter. Dictionaries resolve ambiguities as several of the directions are assigned multiple characters. Experiments with joysticks showed that with little practice participants reached 6.8 wpm [36]. The present study employed the same principle except using finger touch gestures. One benefit is that the gestures require less finger movement than traditional UniStroke-type gestures. Moreover, they are bimanual in nature. There is thus potential for high text-entry speeds. However, touch gestures are different from manipulating dual joysticks as used in the earlier study [36]. Joysticks provide tactile feedback since they move in the direction pushed, and stop once the maximum displacement is achieved, with the stick returning to the home position when released. Moreover, each joystick has a square guide allowing users to feel if they have found the diagonal corners. Thus, joysticks offer tactile feedback. Also, with joysticks the response appears on the display with the user mapping joystick motion with display feedback. With touch input, visual feedback can be provided near the finger, thus allowing for a close mapping between the physical movement and visual feedback. The benefit of building eyes-free text entry on the smartphone is the wide applicability and availability of this technology, where special-purpose input devices such as joysticks are impractical.

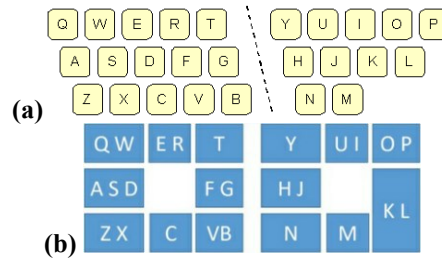


Fig. 2. Qwerty layouts showing (a) customary left-right division for two-handed touch typing and (b) *QB-Gest* letter positions for bimanual gestural input.



Fig. 3. QB-Gest with visual hints in the eyes-free mode.

3 QB-Gest System Description

QB-Gest was implemented in Java and tested on a Huawei C8817E smartphone with a 5-inch display running Android 4.4 KitKat. The UI included a text region and a gesture region, along with two entry modes: eyes-free and eyes-on. The text region

contained fields for the presented and transcribed text, and a word region showing four suggested words. The word region was only implemented for the eyes-on mode.

The gesture region was located below the text region and was divided into a left and right side. The left side was for letters on the left side of the Qwerty layout and the right side was for letters on the right side of the Qwerty layout. Fig. 2a shows the usual Qwerty left-right division of input for two-handed touch typing. Fig. 2b shows the QB-Gest left-right positions for gestural input. Although practical, the assignment of letters to direction is not optimal from an information theoretic perspective. In particular, left-west is assigned three high-frequency letters, ASD, while left-south is assigned just the letter C. The rationale was to exploit users' familiarity with Qwerty. Consequently, the letter-to-direction assignments are limited. An alternative is to change the groups ASD:FG to AS:DFG, but then one would also need to change the right-side assignments from HJ:KL to HJK:L to maintain symmetry across the two hands. There is an obvious trade-off between information theoretic optimization and assignments that leverage users' mental model of the Qwerty keyboard [36].

For eyes-free input, the letter positions in Fig. 2b are hidden and do not appear during input. However, the user can request visual hints using a two-thumb swipe-up gesture. In this case, the letter positions appear for 1.5 seconds. Fig. 3 shows the QB-Gest application in eyes-free mode with visual hints shown.

For eyes-on input, the letter positions are also hidden, but appear immediately when the user's thumb touches the display to make a gesture. On touch, the display shows the letters for the left or right half of the Qwerty layout, depending on the touch location. Each half shows eight boxes around the touch point. By dragging the thumb in one of the eight directions, the respective box of letters is selected once the finger is released. In the eyes-free mode, a short beep is heard each time a letter is entered.

The angle of the gesture was converted to one of eight directions by dividing the space into eight equal sectors centered on compass directions. A trie data structure [37] mapped the sequence of directions to words in a dictionary. The trie is a special tree data structure that stores words associated with each word prefix. In the eyes-on mode, words were suggested based on these word prefixes. Although there are ambiguities caused by words sharing the same sequence of directions, most sequences are unique. Fig. 4 gives an example of text input in the eyes-on mode. In eyes-free mode, entry proceeds similarly except the letter positions do not appear.

3.1 KSPC Analysis

Since QB-Gest positions 26 letters on $2 \times 8 = 16$ keys, the entry of some words is ambiguous. However, the ambiguity is considerably less than that of a phone keypad where 26 letters are positioned on 8 keys.

Keystrokes per character (*KSPC*) is an established metric that captures the key-stroke efficiency of a text entry method [38]. Of course, "keystrokes" is "gesture strokes" in the present context. For ambiguous keyboards, *KSPC* reflects the overhead in resolving the ambiguity when a key sequence corresponds to more than one word.

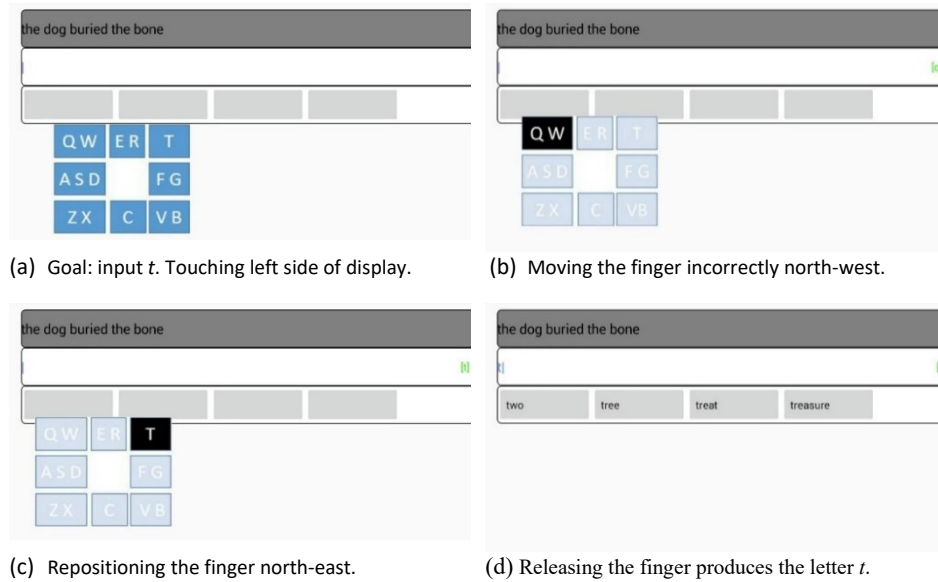


Fig. 4. QB-Gest example input in eyes-on mode.

As an example, for English text entry using a 10,000 word dictionary, a phone keypad has $KSPC = 1.0072$ [38]. In other words, the overhead, on average, is just 7.2 keystrokes per 1000 keystrokes. This is for T9-style input, where the user navigates an ordered list of words when ambiguities occur.

A similar calculation for QB-Gest yields $KSPC = 1.0037$, for an overhead of 3.7 keystrokes per 1000 keystrokes. So, ambiguous words, or "collisions", are rare. Examples for QB-Gest include {king, lung}, {edge, rage} and {rise, ride, rude}. See Fig. 2. Since collisions are rare, we have not implemented a disambiguation method for eyes-free input in our prototype. We expect a full implementation could use linguistic context to resolve the occasional collision that does occur.

3.2 Special Inputs

As well as the two-thumb gesture to request visual hints, we used special inputs for SPACE, BACKSPACE, and ENTER. Two-thumb gestures were chosen to reduce conflicts with letter input. Through some experimentation it was found that symmetric gestures worked better than non-symmetric gestures [39]. The two-thumb gestures are

SPACE	both thumbs moving outward to the sides
BACKSPACE	both thumbs moving inward
ENTER	both thumbs moving down
HINTS	both thumbs moving up

Our research questions are as follows: Can users leverage their knowledge of the Qwerty layout to perform bimanual gestural input on a smartphone? To what degree can users perform eyes-free text entry? Is the error rate affected by visual feedback? Are the gesture dynamics affected by visual feedback? To explore these, we conducted a user study, as now described.

4 Method

We evaluated QB-Gest in both the eyes-free and eyes-on entry modes over four sessions of text entry.

4.1 Participants

Twenty participants were recruited, 10 each assigned to the eyes-free and eyes-on groups. Participants were recruited among students from the university campus of the first author. There were 11 males and 9 females split equally (approximately) between the eyes-free and eyes-on groups. Ages ranged from 10 to 54 years with most participants between 25 to 34 years. The mean ages for the two groups was approximately equal. All the participants had normal or correct-to-normal vision.

The participants were screened for text input skill and Qwerty familiarity as this was a prerequisite for participating in the experiment. Responses were self-assessed on a six-point Likert scale, with higher scores for greater skill/familiarity. The responses are summarized thus:

Text input skill

Eyes-on group ($M = 4.3$, $SD = 0.68$)

Eyes-free group ($M = 4.6$, $SD = 0.52$)

Qwerty familiarity

Eyes-on group ($M = 4.2$, $SD = 1.14$)

Eyes-free group ($M = 5.2$, $SD = 0.79$)

Using a Mann Whitney U test, there was no significant difference between the responses of the two groups in terms of self-assessed text entry skill ($z' = -1.023$, $p = .306$). However, the slightly higher mean response for the eyes-free group on Qwerty familiarity was statistically significant ($z' = -2.053$, $p = .0401$). Although the two groups exhibit different Qwerty skills which may confound results, our main focus was to study eyes-free text entry with the eyes-on entry as a reference.

4.2 Task

The participants performed a text copy task using a standard 500-phrase set [40]. Phrases were selected at random and appeared in the presented text field. The user entered the phrase using QB-Gest in the assigned mode with the result appearing in

the transcribed text field (see Figs. 3 and 4). At the end of a phrase, participants employed the ENTER gesture to move to the next phrase. The participants in the eyes-free group could use the HINTS gesture to receive a visual hint wherein the full Qwerty letter pattern was shown for 1.5 seconds.

4.3 Procedure

Testing was done in a quiet room. The first session included a briefing where participants signed a consent form and completed a questionnaire asking for demographic information and a text entry skill self-assessment. Next, participants practiced QB-Gest for 10 minutes to enter text using either the eyes-free or eyes-on entry mode, depending on the participant's group. After practicing, the measured text entry sessions began. Sessions were time-limited to 20 minutes with the measured text entry part of the session taking about 5 minutes. The number of phrases entered in a session varied from 6 in session 1 to 11 in session 4.

The four sessions for each participant were a few days apart in a concentrated time-period to avoid confounding effects. The entire experiment involved 80 sessions and ran over three months. The participants were asked about their opinions on QB-Gest using a 6-point Likert scale after the fourth session. No monetary reward was given.

4.4 Apparatus

The hardware and QB-Gest user interface described earlier were used in the experiment. The software logged all the interactions performed on the smartphone during the experiment, including the spatial and temporal details of the individual gestures as well as high-level statistics such as text entry speed (in words per minute), error rate (percent of incorrect letters) [41], and requests for hints.

To minimize word collisions (particularly for the eyes-free mode), QB-Gest was configured with a small dictionary that contained only the 1168 unique words in the phrase set. In this configuration, the T9-style $KSPC = 1.0023$.

4.5 Design

The experiment was a 2×4 mixed design with the following independent variables and levels:

Entry mode	Eyes-free, eyes-on
Session	1, 2, 3, 4

The assignments were between-subjects (entry mode) and within-subjects (session). Entry mode was assigned between-subjects to avoid interference between the two conditions.

The dependent variables were text entry speed in words per minute (wpm), error rate (%), hint requests (count per character), output/input gain resulting from using word suggestions (difference in number of output and input symbols over number of

output symbols) and gesture length (cm). Error rate was measured using the minimum string distance metric, comparing the presented and transcribed text phrases.

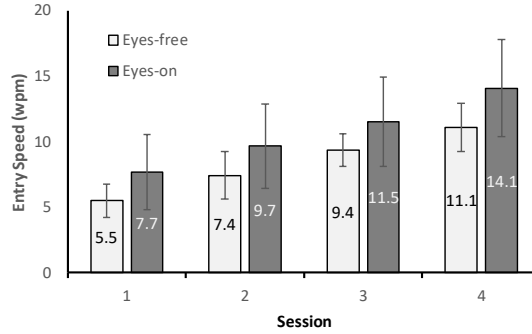


Fig. 5. Text entry speed (words per minute) by session and entry mode. Error bars show ± 1 SD.

5 Result and discussion

5.1 Performance

The grand mean for text entry speed was 9.55 wpm. By session, the means increased from 6.60 wpm (session 1) to 12.60 wpm (session 4). By entry mode, the means were 8.35 wpm (eyes-free) and 10.75 wpm (eyes-on). The results show a significant improvement in text entry performance with practice ($F(3, 54) = 106.5, p < .001, \eta^2 = .851$) and also a significant difference between entry modes ($F(1, 18) = 4.917, p = .040, \eta^2 = .215$). Text entry speed was about 29% higher for the eyes-on entry mode (see Fig. 5). Bonferroni post-hoc tests showed that all the sessions are significantly different from each other ($p < .001$).

Table 1. Eyes-free text mobile text entry results (*Results with a MacBook).

Mode	Hands	Entry Speed (wpm) by Session				
		1	2	3	4	> 5
Invisible Keyboard [18]	2	31.3	35.7	37.9		
Blindtype (PR) [17]	1	22.8				
Graffiti (delayed) [7]*	1	11.1				
Escape [34]	1	6.9	9.3	10.4	11.1	
QB-Gest	2	5.5	7.4	9.4	11.1	24.9
Thumb Stroke [33]	1	7.2				10.8
EdgeWrite [33]	1					7.8
Qwerty [4]	1	2.1				
NoLookNotes [46]	1	1.32				
VoiceOver [46]	1	0.66				

The improvement with practice appears close to linear. It is thus likely that further practice would yield further improvements. However, with prolonged training one would expect the improvement to be logarithmic, that is, with increased practice the

improvement becomes smaller [10, p. 274]. A logarithmic regression model suggests that the mean entry speed may exceed 20 wpm after 18 sessions.

The performance in the eyes-free entry mode seems to match that of the preceding session for the eyes-on entry mode. Another interesting observation is that the spread for eyes-on is larger than the spread for eyes-free. One would expect the spread to be somewhat higher with higher mean values. Also, one could expect a higher spread for the eyes-free mode since the task is more difficult. We suspect that the difference in spread is a result of between-group sampling bias and not necessarily an effect. There was no Session \times Entry Mode interaction effect on entry speed ($F(3, 54) = 0.63$, ns).

We compared our results to other studies of eyes-free mobile text entry. See Table 1. Clearly, the Invisible Keyboard yields text entry speeds much above the other methods as it directly leverages users' Qwerty skills. However, such high mobile text entry speeds are mostly found among expert users, typically young individuals. Older smartphone users typically yield lower performance [1]. The Invisible Keyboard is also the only other reported bimanual input method. Very fast rates were also obtained with an enhanced version of Graffiti. QB-Gest has a similar performance to the Escape-keyboard at the fourth session, while QB-Gest has more rapid improvement. QB-Gest performs marginally better than Thumb Stroke. All these recent methods yield higher performance than older methods such as EdgeWrite and VoiceOver. Comparison of error was not possible as the studies report different error metrics.

5.2 Errors

Error rate was calculated from the minimum-string distance between the presented and transcribed text phrases [41]. The grand mean for error rate was 10.8%. By session, the means decreased from 15.9% (session 1) to 6.90% (session 5). By entry mode, the means were 15.2% (eyes-free) and 6.3% (eyes-on). The error rates by session and entry mode are shown in Fig. 6.

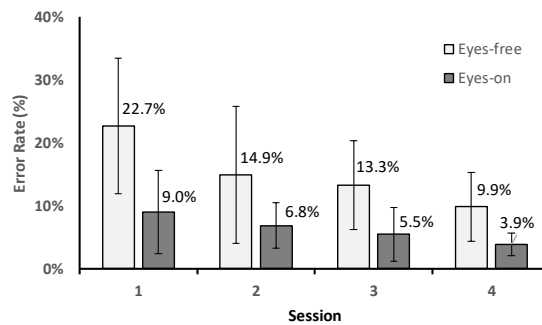


Fig. 6. Error rates (%) by session and entry mode. Error bars show ± 1 SD.

As seen, the error rates with eyes-free entry are nearly twice as high as with eyes-on entry. Higher error rates for eyes-free entry agree with other studies [19, 4].

Training also affects the error rate as there is a reduction from the first to last session for both eyes-free and eyes-on entry. A Levene's test showed that the data lacked

equality of variances. The measurements were therefore transformed using the aligned rank transform (ART) [45]. Mauchly tests showed that the data aligned according to session did not satisfy the assumption of sphericity and a Greenhouse-Geisler correction was therefore applied. There was a significant effect on practice (session) for eyes-on entry ($F(2.156, 38.813) = 13.941, p < .001, \eta^2 = .426$). Bonferroni post-hoc tests show that the error reduction from the first to the second session is significant ($p = .013$) while not from session 2 to 3 and from session 3 to 4. The reduction in error is also significant from session 2 to 4. There is also a significant effect of mode ($F(1, 18) = 13.34, p = .002, \eta^2 = .426$) as well as a significant interaction between session and mode ($F(3, 54) = 3.874, p = .014, \eta^2 = .167$).

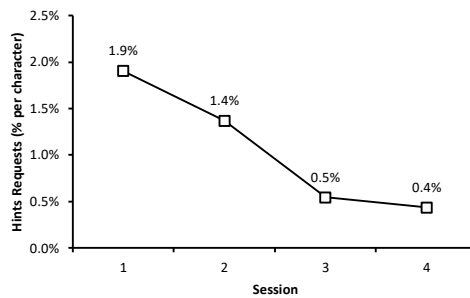


Fig. 7. Percentage visual hints requests by session (% per character).

5.3 Visual Hints

In the eyes-free mode, requesting visual hints is an important feature in learning QB-Gest (see Fig. 3). The need for hints reduces as users get accustomed to which thumb to use and in which direction to swipe. We logged the occurrence of hint requests by session and by character. See Fig. 7. The requests dropped from 1.9% per character in the 1st session to 0.4% in the 4th session, clearly showing an improvement. A Friedman test confirmed that the effect of practice was statistically significant ($\chi^2(3) = 11.61, p = .009$). Conover's post hoc tests revealed that there was no significant difference from one session to the next, but the difference was significant from session 1 to 3 ($p = .007$) and from session 2 to 4 ($p = .013$).

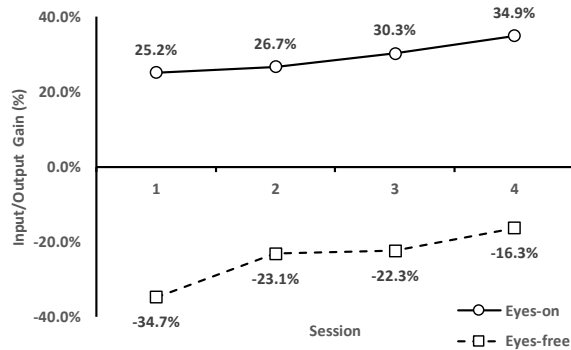


Fig. 8. Output/input gain resulting from word suggestions and BACKSPACE use.

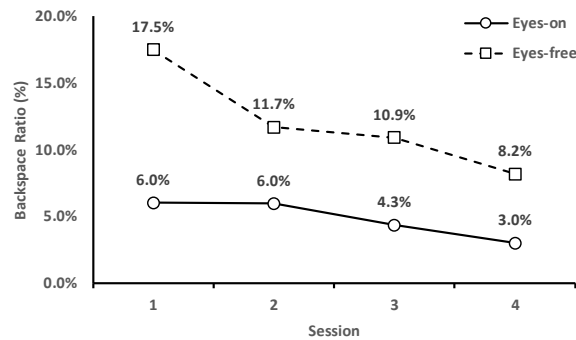


Fig. 9. Backspace ratio.

5.4 Word Suggestions

In the eyes-on mode participants could select words based on prefixes by directly pressing the displayed suggestion. To assess the effect of the suggestions, the input gain for both the eyes-on and eyes-free modes were calculated. We define input gain as the difference between the number of outputs (characters) and the number of inputs (gestures) over the total number of inputs (gestures). All the participants in the eyes-on experiment utilized word suggestions. As illustrated by Fig. 8, the suggestions resulted in a gain of 25.2% in the first session, increasing to a gain of 34.9% in the fourth session. The eyes-free mode yielded a negative gain of -34.7% during the first session that decreased to -16.3% during the fourth session. Clearly, practice had a significant effect on output/input gain ($F(3, 54) = 8.626, p < .001, \eta^2 = .309$). However, Bonferroni post-hoc testing revealed that only the first and last sessions were significantly different ($p < .001$). There was a significant difference between the output/input gains for the eyes-on mode vs. the eyes-free mode ($F(1, 18) = 151.1, p < .001, \eta^2 = .894$).

The negative output/input gains observed with the eyes-free mode can be partly explained by BACKSPACE use. If a user inputs a , followed by BACKSPACE and b , that's 3 gestures producing just 1 character. To explore BACKSPACE further, the ratio of BACKSPACE inputs to all inputs is plotted in Fig. 9. The observations did not show equivalence of variances and the observations were therefore transformed using ART [45]. The eyes-free mode is associated with significantly more use of BACKSPACE than the eyes-on mode ($F(1, 18) = 15.72, p < .001, \eta^2 = .466$). During the first session the eyes-free mode exhibited a mean rate of 17.5% BACKSPACE inputs, while the eyes-on mode only exhibited 6.0% BACKSPACE inputs. Practice had a significant effect on the use of BACKSPACE ($F(3, 54) = 13.196, p < .001, \eta^2 = .415$). Post-hoc tests show that the two first sessions were significantly different ($p = .041$) but not the other two consecutive sessions. Again, session two and four were significantly different ($p = .003$). During the fourth session the BACKSPACE ratio dropped to 8.2% in the eyes-free mode and 3.0% in the eyes-on mode. A significant interaction was also observed between practice and mode ($F(3, 54) = 4.718, p = .005, \eta^2 = .195$).

Having word suggestions partly explains the higher text entry rates in the eyes-on mode compared to the eyes-free mode. In hindsight, the participants should not have been given suggestions in the eyes-on mode to keep this experimental condition constant for both groups. However, as a practical consideration, using word suggestions is clearly beneficial, even expected, for eyes-on text entry.

The eyes-on mode utilized a mixture of relative gestures for character input and absolute targets for selecting words. Marking menus [42] may be one way of providing users with word suggestions without having to rely on direct pointing. The Marking menus approach is based on presenting menu items radially when the user touches the display. Users select an item by making a gesture in the direction of the menu item. Hierarchical selections are also possible. This approach relies on relative motions instead of absolute pointing. With practice, users select items without looking. To avoid confusing word selections with character input, a different region of the display could be allocated for these selections, for example the top middle of the display.

The incorporation of effective word suggestions in eyes-free mode is still an open problem. The user needs feedback on the word suggestions while entering text. Without visual feedback this information must be conveyed using other modalities. The most obvious modality is audio. Further research is needed to uncover how such audio feedback might interfere with the text entry task.

5.5 Gesture Dynamics

To compare the gesture dynamics of the eyes-on and eyes-free modes, detailed gesture information were extracted from the logs and aggregated into mean gesture lengths (cm) and mean gesture angles. See Figs. 10 and 11. There are several noticeable differences between the two modes. First, the gesture lengths in the eyes-on mode are shorter ($M = 1.0, SD = 0.2$) than the eyes-free gestures ($M = 1.2, SD = 0.2$), and the difference is significant ($F(1, 10) = 5.474, p = .041, \eta^2 = .354$). There is no significant difference in gesture length across the two hands for the eyes-on mode ($F(1, 3) = 4.889, p = .114$). A visual inspection of the eyes-on gestures in Fig. 10 shows that

the four diagonal gestures are slightly longer than the four horizontal/vertical gestures. The endpoints of the gestures appear to fall on the boundary of an approximate square for both hands. One explanation is that the visual feedback on the display guides the fingers toward the displayed targets and hence constrains finger movement. Fig. 11 shows that the gesture patterns in eyes-free mode are different from those in eyes-on mode. The two gestures along the adjacent diagonals are longest while the gestures along the perpendicular diagonals are shortest. This pattern mirrors across the hands, explained as follows. There are no visual guides in the eyes-free mode; the user therefore executes the gestures more freely in a comfortable manner. The left thumb has a higher dexterity along the northwest-southeast diagonal as it involves abductions and adductions, while moving the left thumb along the southwest-northeast direction is anatomically more difficult as it involves flexion and extensions. These southwest-northeast motions are thus smaller. The same holds for the right thumb although the patterns are mirrored vertically resulting in longer gestures along the southwest-northeast diagonal.

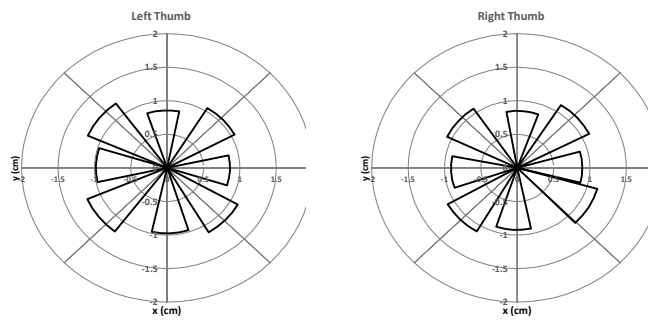


Fig. 10. Mean gesture lengths (cm) and angles in eyes-on mode.

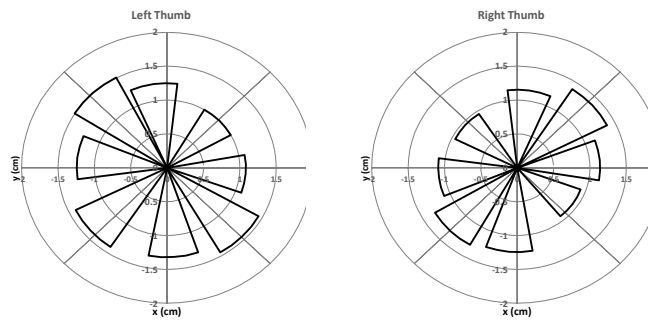


Fig. 11. Mean gesture lengths (cm) and angles in eyes-free mode.

The gesture angles were not noticeably different across the eyes-on and eyes-free modes. However, the plots reveal that the vertical up-down gestures tilted slightly left for the left thumb and slightly right for the right thumb with both modes (an offset of

3-9 degrees). This is probably a result of the thumb origin being in the bottom left corner for the left hand resulting in arc-like shapes and vice versa for the right hand.

These results may help improve the accuracy of gesture recognition. In addition to using just the angle and a minimum distance to detect gestures, the length may help improve the discrimination of neighboring eyes-free gestures. The vertical detection angles should also be adjusted accordingly. Another possibility is to redesign the visual feedback in the eyes-on mode to better align with the hand ergonomics.

5.6 Subjective Assessment

After completing four sessions of testing, participants were asked several questions on their impressions of QB-Gest. Responses were on a 6-point Likert scale. Participants in each group were asked if they felt the entry method (eyes-free or eyes-on) was "easy to use". A Mann Whitney U test revealed no significant difference between the responses from the two groups ($z' = -1.336, p = .182$). They were also asked to self-assess their typing skill on the entry mode they were using. Again, there was no difference in the self-assessed performance between the two entry modes ($z' = -1.413, p = .158$). All the responses were in the upper part of the Likert scale from 3 to 6, indicating a good overall impression with QB-Gest. A positive but non-significant correlation was found between how easy the participants found the method to use and errors ($r_s(20) = .402, p = .079$).

5.7 Improvement with practice

The longitudinal performance measurements suggest that prolonged training will yield further improvements as there was no sign of the flattening as typical in longitudinal studies. It is thus likely that the observed performance is not representative of what is practically possible. For reference, we also measured the peak performance of an expert QB-Gest user in the eyes-free mode (one of the authors) who managed to input the phrase "the quick brown fox jumps over the lazy dog" 10 times at an average rate of 24.9 wpm. The mean performance observed in the user study, namely 14.1 wpm with visual feedback wpm in session 4, is by no means exceptional compared to the 50 wpm obtained with two-thumb text entry [8] with visual feedback. However, in terms of eyes-free text input, our results are better than 2.1 wpm reported for VoiceOver [4] and similar to the 11.1 wpm obtained with Graffiti with visual feedback [28] and better than the 8.34 wpm obtained with Graffiti in eyes-free mode [7].

The error rate was high in the no-visual feedback mode. This means that a practical system employing this type of text-entry needs robust error correction. Error correction is commonly employed in the text-entry domain, for instance through word-level correction of gesture input [7], full-phrase error correction [43], and word-level chording errors [44].

Users who rely on eyes-free text entry are not likely to continuously monitor and detect mistakes in the inputted text and it is thus appropriate to employ error correction techniques where the entire phrase is checked and corrected instead of individual words [43] as this gives more robustness.

6 Conclusions

We presented QB-Guest – a method for inputting text on small smartphone displays using simple Qwerty bimanual gestures. The user visualizes the Qwerty keyboard and gestures with the left or the right thumb in the direction of the desired character from the gap between D-F and J-K keys as the left and right points of origin.

A user study demonstrated that text could be input both with visual feedback at a rate of 14.1 wpm and eyes-free at a rate of 11.1 wpm during the fourth session. An expert user entered text eyes-free at 24.9 wpm.

The error rate was 9.9% for the eyes-free mode and 3.9% for the eyes-on mode during the fourth session. Participants relied little on visual hints in the eyes-free mode as only 0.4% of the inputted characters required hints during the fourth session. The results are comparable to other results obtained with gestures, but the longitudinal data suggest that higher text entry rates are possible with additional training. All participants in the eyes-on mode relied on word suggestions and during the fourth session this led to an output/input gain of 34.9%. The results also indicate that the visual guides in the eyes-on mode constrained the physical movement of the thumbs.

Text entry on smartphone virtual keyboards is visually demanding as there are small targets and no tactile feedback. QB-Gest holds potential for rapid smartphone text entry with low visual attention, for example, in situations where users are multi-tasking and attention is limited. As QB-Gest can be used eyes-free, it holds potential for blind users. Although the participants herein were not visually impaired, their ability to enter text eyes-free is a convincing indicator; however, further user studies involving blind users are needed.

References

1. Smith, A.L., Chaparro, B.S.: Smartphone text input method performance, usability, and preference with younger and older adults. *Human factors* 57 (6),1015-1028 (2015).
2. Leiva, L.A., Sahami, A., Catala, A., Henze, N., Schmidt, A.: Text Entry on Tiny QWERTY Soft Keyboards. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*, 669-678. ACM, New York, NY, USA (2015).
3. Azenkot, S., Lee, N.B.: Exploring the use of speech input by blind people on mobile devices. In: *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)*, Article 11. ACM, New York, NY, USA (2013).
4. Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J., Gonçalves, D.: Blind people and mobile touch-based text-entry: acknowledging the need for different flavors. In: *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '11)*, 179-186. ACM, New York, NY, USA (2011).
5. Azenkot, S., Bennett, C. L., Ladner, R.E.: DigiTaps: eyes-free number entry on touchscreens with minimal audio feedback. In: *Proceedings of the 26th annual ACM symposium on User interface software and technology (UIST '13)*, 85-90. ACM, New York, NY, USA (2013).
6. Goh, T., Kim, S.W.: Eyes-free text entry interface based on contact area for people with visual impairment. In: *Proceedings of the adjunct publication of the 27th annual ACM*

- symposium on User interface software and technology (UIST'14 Adjunct), 69-70. ACM, New York, NY, USA (2014).
7. Tinwala, H., MacKenzie, I.S.: Eyes-free text entry with error correction on touchscreen mobile devices. In: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordCHI '10), 511-520. ACM, New York, NY, USA (2010).
 8. Azenkot, S., Zhai, S.: Touch behavior with different postures on soft smartphone keyboards. In: Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12), 251-260. ACM, New York, NY, USA (2012).
 9. MacKenzie, I. S., Zhang, S.X.: The design and evaluation of a high-performance soft keyboard. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '99), 25-31. ACM, New York, NY, USA (1999).
 10. MacKenzie, I. S.: Human-computer interaction: An empirical research perspective. Morgan Kaufmann, Waltham, MA (2013).
 11. Zhai, S., Smith, B.A.: Alphabetically biased virtual keyboards are easier to use: layout does matter. In: CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01), 321-322. ACM, New York, NY, USA (2001).
 12. Bi, X., Smith, B.A., Zhai, S.: Quasi-qwerty soft keyboard optimization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10), 283-286. ACM, New York, NY, USA (2010).
 13. Raynal, M.: KeyGlasses: semi-transparent keys on soft keyboard. In: Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility (ASSETS '14), 347-349. ACM, New York, NY, USA (2014).
 14. Kane, S.K., Jayant, C., Wobbrock, J.O., Ladner, R.E.: Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In: Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility (Assets '09), 115-122. ACM, New York, NY, USA (2009).
 15. MacKenzie, I. S., Buxton, W.: Extending Fitts' law to two-dimensional tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '92), 219-226. ACM, New York, NY, USA (1992).
 16. Shi, W., Yu, C., Yi, X., Li, Z., Shi, Y.: TOAST: Ten-Finger Eyes-Free Typing on Touchable Surfaces. In: Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2(1), Article 33. ACM, New York, NY, USA (2018).
 17. Lu, Y., Yu, C., Yi, X., Shi, Y., Zhao, S.: BlindType: Eyes-Free Text Entry on Handheld Touchpad by Leveraging Thumb's Muscle Memory. In: Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1(2), Article 18. ACM, New York, NY, USA (2017).
 18. Zhu, S., Luo, T., Bi, X., Zhai, S.: Typing on an Invisible Keyboard. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Paper 439. ACM, New York, NY, USA (2018).
 19. Kane, S.K., Bigham, J.P., Wobbrock, J.O.: Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In: Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (Assets '08), 73-80. ACM, New York, NY, USA (2008).
 20. Guerreiro, J., Rodrigues, A., Montague, K., Guerreiro, T., Nicolau, H., Gonçalves, D.: TableTS Get Physical: Non-Visual Text Entry on Tablet Devices. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15), 39-42. ACM, New York, NY, USA (2015).

21. Ward, D.J., Blackwell, A.F., MacKay, D.J.C.: Dasher—a data entry interface using continuous gestures and language models. In: Proceedings of the 13th annual ACM symposium on User interface software and technology (UIST '00), 129-137. ACM, New York, NY, USA (2000).
22. Polacek, O., Sporcka, A.J., Slavik, P.: Text input for motor-impaired people. *Universal Access in the Information Society* 16 (1), 51-72 (2017).
23. Lyons, K., Starner, T., Plaisted, D., Fusia, J., Lyons, A., Drew, A., Looney, E. W.: Twiddler typing: one-handed chording text entry for mobile phones. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04), 671-678. ACM, New York, NY, USA (2004).
24. Sandnes, F.E.: Can spatial mnemonics accelerate the learning of text input chords?. In: Proceedings of the working conference on Advanced visual interfaces (AVI '06), 245-249. ACM, New York, NY, USA (2006).
25. Frey, B., Rosier, K., Southern, C., Romero, M.: From texting app to braille literacy. In: CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12), 2495-2500. ACM, New York, NY, USA (2012).
26. Goldberg, D., Richardson, C.: Touch-typing with a stylus. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93), 80-87. ACM, New York, NY, USA (1993).
27. Wobbrock, J.O., Myers, B.A., Aung, H.H., LoPresti, E.F.: Text entry from power wheelchairs: edgewrite for joysticks and touchpads. In: Proceedings of the 6th international ACM SIGACCESS conference on Computers and accessibility (Assets '04), 110-117. ACM, New York, NY, USA (2003).
28. Castellucci, S.J., MacKenzie, I.S.: Graffiti vs. unistrokes: an empirical comparison. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08), 305-308. ACM, New York, NY, USA (2008).
29. Wobbrock, J.O., Myers, B.A., Kembel, J.A.: EdgeWrite: a stylus-based text entry method designed for high accuracy and stability of motion. In: Proceedings of the 16th annual ACM symposium on User interface software and technology (UIST '03), 61-70. ACM, New York, NY, USA (2003).
30. Sandnes, F.E., Tan, T.B., Johansen, A., Sulic, E., Vesterhus, E., Iversen, E.R.: Making touch-based kiosks accessible to blind users through simple gestures. *Universal Access in the Information Society* 11 (4), 421-431 (2012).
31. Chen, X., Grossman, T., Fitzmaurice, G.: Swipeboard: a text entry technique for ultra-small interfaces that supports novice to expert transitions. In: Proceedings of the 27th annual ACM symposium on User interface software and technology, 615-620. ACM, (2014).
32. Bi, X., Chelba, C., Ouyang, T., Partridge, K., Zhai, S.: Bimanual gesture keyboard. In: Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12), 137-146. ACM, New York, NY, USA (2012).
33. Lai, J., Zhang, D., Wang, S., Kilic, I.L., Zhou, L.: A Thumb Stroke-Based Virtual Keyboard for Sight-Free Text Entry on Touch-Screen Mobile Phones. In: Hawaii International Conference on System Sciences. (2018).
34. Banovic, N., Yatani, K., Truong, K. N.: Escape-Keyboard: A Sight-Free One-Handed Text Entry Method for Mobile Touch-screen Devices. *Int. J. Mob. Hum. Comput. Interact.* 5 (3), 42-61 (2013).
35. Armstrong, P., Wilkinson, B.: Text entry of physical and virtual keyboards on tablets and the user perception. In: Proceedings of the 28th Australian Conference on Computer-Human Interaction (OzCHI '16), 401-405. ACM, New York, NY, USA (2016).

36. Sandnes, F.E., Aubert, A.: Bimanual text entry using game controllers: relying on users' spatial familiarity with QWERTY. *Interacting with Computers* 19 (2), 140-150 (2007).
37. Fredkin, E.: 1960. Trie memory. *Commun. ACM* 3(9), 490-499 (1960).
38. MacKenzie, I.S.: KSPC (keystrokes per character) as a characteristic of text entry techniques. In: *Proceedings of the Fourth International Symposium on Human Computer Interaction with Mobile Devices, LNCS*, vol. 2411, 195-210. Springer-Verlag, Heidelberg, Germany (2002).
39. Matias, E., MacKenzie, I. S., Buxton, W.: Half-QWERTY: typing with one hand using your two-handed skills. In: *Conference Companion on Human Factors in Computing Systems (CHI '94)*, Catherine Plaisant (Ed.), 51-52. ACM, New York, NY, USA (1994).
40. MacKenzie, I. S., Soukoreff, R.W.: Phrase sets for evaluating text entry techniques. In: *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 754-755 (2003).
41. MacKenzie, I.S., Soukoreff, R.W.: A character-level error analysis technique for evaluating text entry methods. In: *Proceedings of the second Nordic conference on Human-computer interaction (NordicCHI '02)*, 243-246. ACM, New York, NY, USA (2002).
42. Kurtenbach, G., Buxton, W.: User learning and performance with marking menus. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*, 258-264. ACM, New York, NY, USA (1994).
43. MacKenzie, I. S., Castellucci, S.: Reducing visual demand for gestural text input on touchscreen devices. In: *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*, 2585-2590. ACM, New York, NY, USA (2012).
44. Sandnes, F.E., Huang, Y.-P.: Chording with spatial mnemonics: automatic error correction for eyes-free text entry. *Journal of information science and engineering* 22 (5), 1015-1031 (2006).
45. Wobbrock, J.O., Findlater, L., Gergle, D., Higgins, J.J.: The aligned rank transform for nonparametric factorial analyses using only anova procedures. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 143-146 (2011).
46. Bonner, M.N., Brudvik, J.T., Abowd, G.D., Edwards, W.K.: No-look notes: accessible eyes-free multi-touch text entry. In: *International Conference on Pervasive Computing*, pp. 409-426. Springer, Berlin, Heidelberg (2010).