Effect of Target Size on Non-Visual Text-Entry

André Rodrigues[1], Hugo Nicolau[2], Kyle Montague[3], Luís Carriço[1], Tiago Guerreiro[1]
[1]LaSIGE, Faculdade de Ciências, Universidade de Lisboa
[2]INESC-ID; Instituto Superior Técnico, Universidade de Lisboa
[3] Open Lab, Newcastle University
afrodrigues@fc.ul.pt, hman@inesc-id.pt, kyle.montague@ncl.ac.uk, lmc@di.fc.ul.pt, tjvg@di.fc.ul.pt

ABSTRACT
Touch-enabled devices have a growing variety of screen sizes; however, there is little knowledge on the effect of key size on non-visual text-entry performance. We conducted a user study with 12 blind participants to investigate how non-visual input performance varies with four QWERTY keyboard sizes (ranging from 15mm to 2.5mm). This paper presents an analysis of typing performance and touch behaviors discussing its implications for future research. Our findings show that there is an upper limit to the benefits of larger target sizes between 10mm and 15mm. Input speed decreases from 4.5 to 2.4 words per minute (WPM) for targets sizes below 10mm. The smallest size was deemed unusable by participants even though performance was in par with previous work.

Author Keywords
Blind; Text-entry; Performance; Touchscreen; Key Size;

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces - *Input devices and strategies*. K.4.2 Computers and Society: Social Issues – *Assistive technologies for persons with disabilities.*

INTRODUCTION
Touchscreens are the primary interaction technology on current smartphones. In recent years, touchscreen devices have endured a diversification of screen size from 10-inch tablets to 1.5-inch smart watches. Previous work assessed how target size affects input performance on smartphones [7, 10, 15] with the purpose of inferring an optimal key size. More recently, authors have investigated *tiny* keyboards for wearable devices proposing new visual techniques and language models [11, 17]. Although this topic has been studied for years, it has been limited to sighted users and techniques that rely on visual feedback. Visually impaired people resort to a different touch interaction paradigm (*Explore by Touch*) using accessibility services, such as *VoiceOver* [1] and *TalkBack* [5].

MobileHCI '16, September 06-09, 2016, Florence, Italy
© 2016 ACM. ISBN 978-1-4503-4408-1/16/09□$15.00
DOI: http://dx.doi.org/10.1145/2935334.2935376

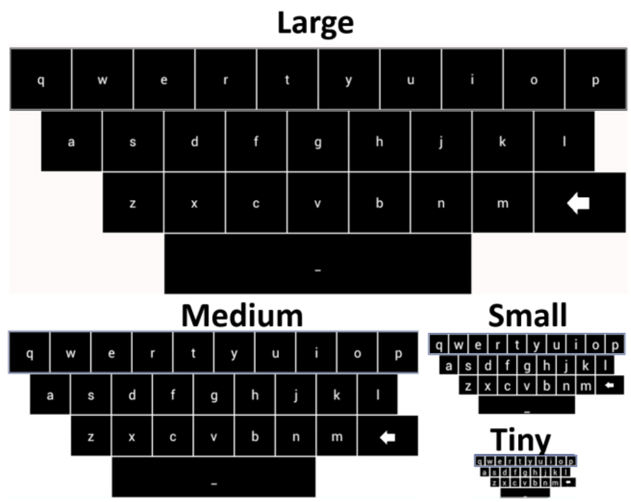

Figure 1- The keyboard conditions and their relative sizes.

In *Explore By Touch* users explore the interface elements by dragging their fingers while receiving audio feedback of the options underneath; lifting the finger or double tapping (depending on the device's settings) makes a selection. Given the variety and fast evolution of devices' dimensions, it is of the utmost importance to understand the effect of target size in current non-visual input methods. Yet, there is no knowledge on how performance changes with target size, as research in non-visual input methods has been limited to a single size [2, 4, 6, 14, 16]. Targeting smaller keys is a visually demanding task, but screen reader users rely on a drag to key interaction. Thus, smaller targets reduce the space users have to explore and hear auditory feedback. It is not clear at what target size we start to see a negative effect on performance. On the other hand, larger sizes increase the distance between targets and can provoke an adverse effect in performance.

In this paper, we contribute with an understanding of the effect of target size on non-visual text input with touchscreens. We conducted a user study involving twelve blind participants with four virtual keyboards with keys sizes of 15mm, 10mm, 5mm and 2.5mm. Results show a significant decrease in input speed from the two larger key sizes to the two smaller ones. Although users are slower, and incurred in more errors during the typing process with the smaller keyboards, the resulting sentences have similar quality across all sizes. Moreover, as key size gets smaller,

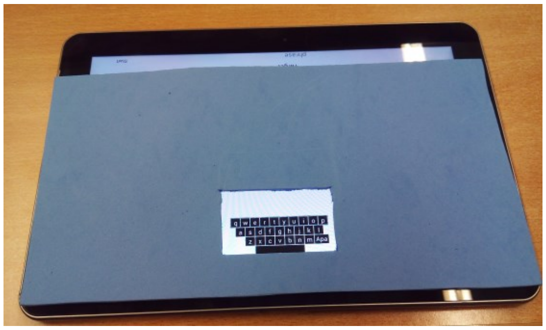

Figure 2 - Experimental setup for the medium size keyboard with the sponge cutout fixed to the tablet.

the relative landing distance from targets increases and users perform less efficient paths on the keyboard. All users considered the smaller size inaccessible.

USER STUDY

We recruited 12 blind users with at least one month of prior smartphone usage to participate in our IRB approved user study. Participants had light perception at most and their ages ranged between 24 and 59 (M=38.5, SD=11.1); three were female; none owned smartwatches; four had a tablet and all were familiar with *Explore by Touch*.

Apparatus

The study was conducted on a Samsung Galaxy Tab 2 with a 10.1-inch (149ppi) screen, running Android 4.1. We developed an app to log the touchscreen interactions and control study tasks. Four keyboard conditions were defined - *Tiny; Small; Medium* and *Large;* with key sizes of 2.5mm; 5mm; 10mm; and 15mm, respectively. These sizes mimic possible keyboards of a smart watch, smartphone in portrait and landscape orientations, and tablet landscape orientation.

To eliminate external device factors and isolate the effect of target size, we used a single tablet device fixed to a table. Each keyboard was rendered in the bottom center, had square keys, with no spacing between them, composed of the letters from A to Z, and included a spacebar (4 keys width) and backspace (1.5 key width right of *the 'm'* key) (Figure 1).

In order to limit the interaction space on the screen, we relied on a physical border created with a sponge cutout overlay (Figure 2). This ensured participants could easily locate the virtual keyboard without the need to explore the whole screen. Moreover, if users dragged their finger outside the keyboard layout, they heard a single beep. To enter text, participants dragged their finger across the keyboard and the letters were read out loud via the device's TTS (Text-to-Speech engine); upon lifting, the focus letter was inserted and users heard a keystroke click sound. Upon entering a space character, the last inserted word was read out loud. Split-Tap [9] was not available due to its limitations on corner targets and smaller screen sizes.

Procedure

The study took approximately 90 minutes per participant. Participants were seated and could take breaks between trials. First, participants were briefed on the procedure of the study, followed by a questionnaire where they were asked about their device usage and proficiency with touchscreens and QWERTY layouts. Participants then interacted with each of the keyboard conditions in a counterbalanced order. For each condition, they had a five-minute warm-up, followed by a 10-minute text-entry session or until five trials were completed (which ever limit was reached first). Each trial consisted of a single sentence, comprised of five words with an average size of five characters. Sentences were randomly selected from a language corpus built following the approach of MacKenzie et al. [12]; each sentence had a minimum correlation of 0.97 with the language. In order to capture their natural typing behavior, users were allowed to delete and correct errors. We instructed participants to write as fast and accurate as possible. The screen was occluded during the study. After each condition, participants were debriefed.

Measures

We calculated input speed in *words per minute (WPM*) as in [19]. Accuracy measures included *uncorrected, corrected,* and *total error rates* [18]. *Uncorrected errors* are those remaining in the transcribed sentence. The *Corrected error Rate* show the percentage of erased characters that were in fact incorrect - a high percentage means most erased characters were errors -, and *total error* rate represents the percentage of erroneous characters that were entered, including those that were corrected. *Substitution Error Rate* related to the number of incorrect characters instead of the intended one; *Omission Error Rate* is the ratio of omission to character presentations; *Insertion Error Rate* are the characters that were mistakenly added. Touch movement measurements included *Path Length* (total movement distance travelled on screen to reach the target) and *Task Axis Length* (straight-line distance between first touch location and the centroid of the target) in pixels. In order for the distances to be comparable across sizes we divided them by the respective key size reaching an adjusted relative distance. To assess the *Path Efficiency* of the movement interactions, we computed the ratio between the *Path Length* (PL) and *Task Axis Length* (TAL) as PL/TAL. We consider *Slips* incorrect entries where the last visited key was the intended target. *Reentries* are the number of times the participant entered the intended key after it was first visited. We also calculated landing accuracy on: intended target, adjacent keys, key row and column.

Design and Analysis

We had one independent factor, *Keyboard Size* (*large*, *medium*, *small* and *tiny*). We applied the Shapiro-Wilk test on all dependent measures; for those not normally distributed, we applied $\log_{10}$ or exponential transformations, resulting in normally distributed measures (p >.05). Mixed-effects model analysis of variance was applied to normally distributed measures. Mixed effects models extend repeated measures models, such as ANOVA's, allowing for unbalanced data as ours. We modeled *Keyboard Size* as a

	Large	Medium	Small	Tiny
WPM	4.5 (1.6)	4.1 (1.2)	3.4 (1.2)	2.4 (0.9)
Total Err (%)	7 (4)	9 (6)	10 (4)	15 (8)
Unc Err (%)	3 (3)	5 (5)	5 (4)	7 (5)
Corr Err (%)	54 (20)	55 (29)	61 (16)	68 (30)
Subs (%)	6 (4)	7 (5)	8 (3)	13 (8)
Omi (%)	1 (2)	1 (2)	2 (3)	4 (3)
Ins (%)	1 (1)	2 (2)	2 (3)	2 (2)

Table 1 - Performance results - all values presented in mean (standard deviation) pairs. The measures abbreviated are Total, Uncorrected, Corrected, Substitutions Omissions and Insertion Error Rates.

fixed effect and *trial* as a nested factor within *Keyboard Size*. For the remaining measures not normally distributed, we applied Friedman tests, with the means per participant, and post hoc Dunn-Bonferroni pairwise comparisons [3].

Results

Participants produced 205 sentences distributed per condition as follows: 58 with *Large*, 59 with *Medium*, 52 with *Small,* and 36 with *Tiny*.

Text-Entry Performance

In this section, we report common performance measures, namely regarding input speed and accuracy; detailed results are shown in Table 1.

Large and Medium results were highly participant-dependent. There were no significant differences between *Large* and *Medium* in any of the measures applied. Although, there seems to be an overall decrease in *WPM* as the *Keyboard* size decreases, five participants had higher *WPM* in the *Medium* size than in any other condition.

WPM significantly decrease in Small and Tiny. *Keyboard Size* had a statistically significant effect on *WPM* ($F_{3,185}$ = 10.1, $p < .05$), with *WPM* decreasing with the *Keyboard Size*. Post hoc analysis revealed that *WPM* significantly decreased between *Large* and *Small* ($p < .05$), and *Tiny* ($p < .01$); and between *Medium* and *Tiny* ($p < .01$).

Total errors increases when size decreases. A Friedman test revealed significant differences between the *Keyboard* conditions ($X^2(3) = 10.5$, $p < .05$). The post hoc comparison revealed a significant difference between *Large* and *Tiny* ($p < .05$); *Medium* and *Tiny* ($p < .05$).

Tiny size has more substitution and omission errors. There was a significant effect of *Keyboard Size* on the *substitution* ($X^2(3) = 8.1$, $p < .05$) and *omission error rates* ($X^2(3) = 15.7$, $p < .01$). Post hoc analysis revealed significantly more *substitution*s in *Tiny* than *Large* ($p < .05$), and *Medium* ($p < .05$) sizes. There were also significantly more *omissions* in *Tiny* than *Large* ($p < .05$) and *Medium* ($p < .05$) sizes. However, we found no significant differences regarding insertion error rates.

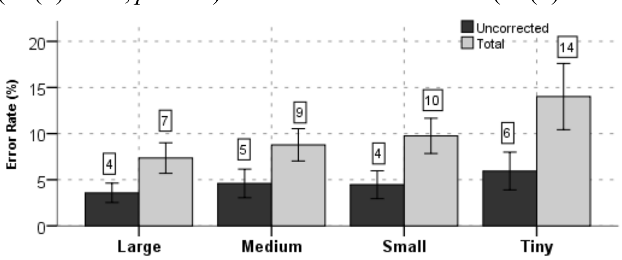

Figure 4 - Total and Uncorrected error rate of the four conditions.

	Large	Medium	Small	Tiny
On Target (%)	41 (15)	38 (12)	29 (14)	22 (13)
On Adjacent (%)	84 (10)	75 (12)	59 (21)	47 (21)
On Row (%)	73 (18)	68 (18)	59 (21)	48 (20)
On Column (%)	53 (13)	52 (12)	42 (14)	33 (16)
TAL (px)	5 (1)	6 (1)	9 (4)	12 (7)
Path Length (px)	6 (3)	8 (4)	17 (11)	30 (20)
PLTLA	1.1 (0.8)	1.3 (0.5)	1.9 (0.8)	2.7 (1.5)
Visited Keys	2.2 (0.5)	2.2 (0.6)	3.0 (0.8)	4.8 (2.1)
Reentries	0.8 (0.7)	1.0 (0.8)	1.8 (1.2)	2.6 (1.8)
Slip (%)	27 (24)	38 (25)	37 (25)	61 (25)

Table 2 - Touch behavior results - all values reported in mean (standard deviation) pairs.

No significant differences in uncorrected or corrected error rate. Participants tended to correct most of their errors as shown by the low uncorrected errors presented in Figure 4.

Users struggled with Tiny size. In the post questionnaire of the *Tiny* condition, all users remarked that the keyboard was too small. Two of the participants asked explicitly to stop the session during their second trial (about 7 minutes in), due to their fingers continually hitting multiple keys at time.

Touch Movement

We report on hit and movement analysis. Averages and standard deviations for all measures are detailed in Table 2.

Landing accuracy significantly decreases from Large and Medium to Small and Tiny. Participants went from landing on the target *column*, *row* or *adjacent* key on *Large* to landing in unrelated sections of the keyboard in *Tiny*. Landing *on target* accuracy almost halved from *Large* to *Tiny* (Table 2). We found a significant effect of the *Keyboard Size* on all landing metrics with *on adjacent* ($F_{3,185}$= 31.8, $p < .001$), *on row* ($F_{3,185}$= 7.0, $p < .001$), *on column* ($F_{3,185}$= 9.0, $p < .001$) and *on target* ($F_{3,185}$=8.1, $p < .001$). Post hoc analysis revealed *on adjacent* results were statistically different between all conditions ($p < .001$), except from *Small* to *Tiny*. *On row* was significantly different from *Large to Small* and *Tiny* ($p < .01$). *Large* and *Medium* were more accurate *on column* than *Small* and *Tiny* ($p < .05$). Finally, *on*

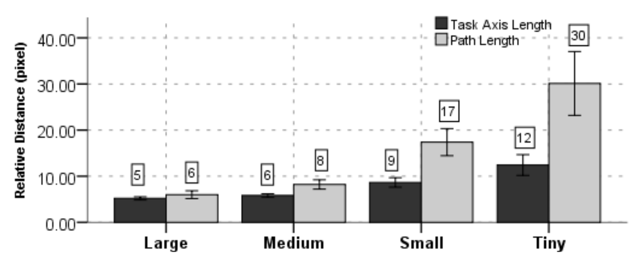

Figure 3- Relative Path Length and Task Axis Length in pixels.

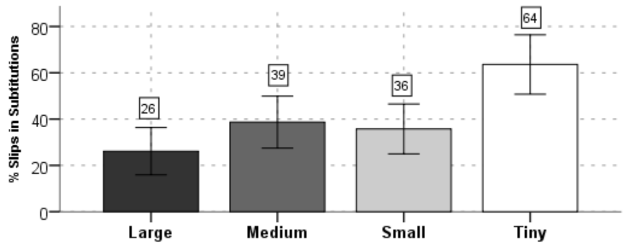

Figure 5 - Percentage of Slip Errors in Substitutions.

target accuracy was statistically lower between *Small* ($p < .05$) and both *Large* and *Medium*; and between *Tiny* and *Large* ($p < .05$).

Exploration efficiency is not affected by larger sizes. The inaccuracy of the touch down translated in landing further from the intended target (Figure 3) with a significant effect of the *TAL* ($F_{3,185}$=21.1 , $p < .001$), a post-hoc pairwise comparisons revealed a significant difference: *Large* and *Medium* with *Small* and *Tiny* ($p < .001$). There was a significant effect of *Keyboard Size* on *PLTLA* ($F_{3,185}$=22.0, $p < .001$), Post hoc tests revealed significant differences between all *Keyboard Sizes* ($p < .001$) except between *Large and Medium; Small and Tiny*. Thus, participants travelled less efficient paths between keys only on the two smaller sizes, illustrated in Figure 3.

Tiny size more than doubles visited keys from Large and Medium. A Friedman test revealed a significant effect of *Keyboard Size* on the number of *visited keys* ($X^2(3) = 27.9$, $p < .01$). *Visited keys* (per character inserted) significantly increased from *Large* and *Medium* to *Tiny* ($p < .001$); and from *Large* to *Small* ($p < .05$).

Smaller keyboards result in higher reentries of intended keys. There was a significant effect of *keyboard size* on *reentries* ($F_{3,185}$=12.5, $p < .001$). A post hoc test revealed a significant difference between *Large* and *Medium* with *Small* ($p < .05$); and *Tiny* ($p < .05$).

Slip errors account for 63% of the substitutions on Tiny. A Friedman test revealed a significant effect of *keyboard size* on *slips* ($X^2(3)$=12.8, p<.005). Post hoc test revealed a significant difference between *Tiny* and both *Large* and *Medium* (p<.05), illustrated in Figure 5.

DISCUSSION

We witnessed a decrease in input speed as *keyboard size* decreased. However, the *uncorrected error* rate was consistent across conditions, even though, in the *Tiny* keyboard condition, participants committed more *substitution* and *omission errors*. The decrease in input speed is the result of not only an increase in typing errors, but poorer landing accuracy and exploration efficiency. Moreover, participants take longer and require multiple key reentries before being able to select the intended target.

The results obtained and behaviors observed enable us to devise implications regarding non-visual interaction and keyboard size. These are of interest to designers, researchers and manufacturers developing devices with new form factors and non-visual text-entry methods:

Bigger is not always better. Input speed on the three larger sizes was highly participant dependent. While most were better as size increase, five users were better on *Medium* than on any other and other five were better on *Small* than in one or more of the sizes above. Several participants were better on several measures (e.g. *WPM*, *on target*, *total error rate*) with *Medium* than with *Large*. Participants felt keys were too far away from each other in the *Large* keyboard, when truly, there were no gaps between keys. Users did not receive feedback while navigating inside a target, only when they crossed its bounds. Future work should explore alternative methods to provide feedback during longer exploration times. Moreover, on larger screen sizes, keyboards should be adaptable to users' expectations and not simply fill the whole screen.

New selection methods for tinier sizes. When the key size decreases it becomes impossible to rest the finger on a single key, making both keyboard exploration and key selection arduous tasks. As previous reported by Holz and Baudisch [8], our participants tended to roll their fingers when lifting them from the screen, often resulting in the selection of adjacent keys. To this end, we observed that most of *substitutions errors* on the *Tiny* condition were slips (M=61%), as opposed to all the other sizes and previous research [13] that reported slips to only account for 38% of the substitutions errors. New selection methods are needed, ones that mitigate the risks of slip errors occurring, such as creating slip guards or touch filtering.

Typing on Tiny QWERTY can be overwhelming. *Tiny* QWERTY keyboard averaged 2.4 WPM, which is faster than when novice blind users input text with QWERTY on touchscreens devices [14]. Yet, users were adamant of their discomfort with the method. Some explicitly asked to stop while others commented, during the final debriefing, that they would never use that keyboard. There was no clear size preference; however, no participant preferred to use the tiny keyboard. Participants landing accuracy was below 50% in all metrics. Relying in just audio feedback brings additional challenges that rendered the tiny size inaccessible.

Spatial independent input techniques. Spatial dependent text-entry methods might not be ideal candidates for writing on the smallest devices. One possible solution could be found in alternative input methods that do not rely on spatial abilities and fine grain movements. Spatial agnostic text-entry methods might prove to be more valuable when applied to smaller devices, such techniques can already be found on prior research with braille input methods [2, 16].

Limitations

We conducted the experiment on a single device to assess the impact of target size, eliminating the effect of device. Using different devices would introduce confounding factors, such as form factor, touchscreen sensor accuracy, "wearabilility" and holding position. However, keyboards size is often a result of the device form factor, as such, future work should explore the effects of the physical form factors of the devices and natural holding positions, on text entry performance.

CONCLUSION

We conducted a study with twelve blind users to assess the effects of target size in non-visual text entry. We compared four *Keyboard* conditions, with key sizes of 15mm, 10mm, 5mm, and 2.5mm. Our results showed an overall decrease in *input speed, landing accuracy* and *movement efficiency* as target sizes decreased. However, this decrease did not significantly affect the uncorrected error rate even though users did commit more errors. Our results suggest target sizes above 10mm do not increase performance. Tiny sizes come with additional challenges; the fat finger problem is exacerbated in these interfaces since one finger touches multiple keys inevitably. Reaching a target is as much of a challenge as it is to select it. Overall, The smallest size was deemed inaccessible by our participants, highlighting the need for further work on non-visual input methods for *tiny* interfaces.

ACKNOWLEDGEMENTS

We thank all the participants and the Fundação Raquel e Martin Sain in Lisbon. This work was partially supported by Fundação para a Ciência e a Tecnologia (FCT) through funding of the scholarship, ref. SFRH/BD/103935/2014, LaSIGE Research Unit, ref. UID/CEC/00408/2013; INESC-ID Research Unit, ref. UID/CEC/50021/2013; and EPSRC award number DERC EP/M023001/1 (Digital Economy Research Centre).

REFERENCES

1. Apple. iOS Accessibility. Retrieved September, 21, 2015 from http://www.apple.com/accessibility/ios/voiceover/
2. Shiri Azenkot, Jacob O. Wobbrock, Sanjana Prasain, and Richard E. Ladner. 2012. Input finger detection for nonvisual touch screen text entry in Perkinput. In Proceedings of Graphics Interface 2012 (GI '12). Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 121-129.
3. Olive J. Dunn. 1964. Multiple Comparisons Using Rank Sums. In *Technometrics*, 6, 241-252.
4. Leah Findlater, Jacob O. Wobbrock, and Daniel Wigdor. 2011. Typing on flat glass: examining ten-finger expert typing patterns on touch surfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). ACM, New York, NY, USA, 2453-2462.
5. Google. Google Android Accessibility Help. Retrieved September 21, 2015, from https://support.google.com/talkback/
6. João Guerreiro, André Rodrigues, Kyle Montague, Tiago Guerreiro, Hugo Nicolau, and Daniel Gonçalves. 2015. 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). ACM, New York, NY, USA
7. Niels Henze, Enrico Rukzio, and Susanne Boll. 2011. 100,000,000 taps: analysis and improvement of touch performance in the large. In Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11). ACM, New York, NY, USA
8. Christian Holz and Patrick Baudisch. 2010. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10). ACM, New York, NY, USA, 581-590.
9. Shaun K. Kane, Jeffrey P. Bigham, and Jacob O. Wobbrock. 2008. 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). ACM, New York, NY, USA, 73-80.
10. Seungyon Lee and Shumin Zhai. 2009. The performance of touch screen soft buttons. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). ACM, New York, NY, USA, 309-318.
11. Luis A. Leiva, Alireza Sahami, Alejandro Catala, Niels Henze, and Albrecht Schmidt. 2015. Text Entry on Tiny QWERTY Soft Keyboards. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA,669-678.
12. I. Scott MacKenzie and R. William Soukoreff. 2003. 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.
13. Hugo Nicolau, Kyle Montague, Tiago Guerreiro, André Rodrigues, Vicki L. Hanson. 2015. Typing Performance of Blind Users: An Analysis of Touch Behaviors, Learning Effect, and In-Situ Usage. In Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '15).
14. João Oliveira, Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves. 2011. 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). ACM, New York, NY, USA, 179-186.

15. Pekka Parhi, Amy K. Karlson, and Benjamin B. Bederson. 2006. Target size study for one-handed thumb use on small touchscreen devices. In Proceedings of the 8th conference on Human-computer interaction with mobile devices and services (MobileHCI '06). ACM, New York, NY, USA
16. Caleb Southern, James Clawson, Brian Frey, Gregory Abowd, and Mario Romero. 2012. An evaluation of BrailleTouch: mobile touchscreen text entry for the visually impaired. In Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12). ACM, New York, NY, USA, 317-326. http://doi.acm.org/10.1145/2371574.2371623
17. Keith Vertanen, Haythem Memmi, Justin Emge, Shyam Reyal, and Per Ola Kristensson. 2015. VelociTap: Investigating Fast Mobile Text Entry using Sentence-Based Decoding of Touchscreen Keyboard Input. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). ACM, New York, NY, USA, 659-668.
18. Jacob O. Wobbrock and Brad A. Myers. 2006. Analyzing the input stream for character- level errors in unconstrained text entry evaluations. ACM Trans. Comput.-Hum. Interact. 13, 4 (December 2006)
19. Jacob O. Wobbrock. 2007. Measures of text entry performance. In Text Entry Systems, MacKenzie and Tanaka-Ishii (eds.). San Francisco: Morgan Kaufmann