Gaze and head pointing for hands-free text entry: Applicability to ultra-small virtual keyboards

Yulia Gizatdinova
University of Tampere, Finland
julia.kuosmanen@uta.fi

Oleg Špakov
University of Tampere, Finland
oleg.spakov@uta.fi

Outi Tuisku
Lappeenranta University of Technology, Finland
outi.tuisku@lut.fi

Matthew Turk
University of California, Santa Barbara, USA
mturk@cs.ucsb.edu

Veikko Surakka
University of Tampere, Finland
veikko.surakka@uta.fi

ABSTRACT

With the proliferation of small-screen computing devices, there has been a continuous trend in reducing the size of interface elements. In virtual keyboards, this allows for more characters in a layout and additional function widgets. However, vision-based interfaces (VBIs) have only been investigated with large (e.g., full-screen) keyboards. To understand how key size reduction affects the accuracy and speed performance of text entry VBIs, we evaluated gaze-controlled VBI (g-VBI) and head-controlled VBI (h-VBI) with unconventionally small (0.4°, 0.6°, 0.8° and 1°) keys. Novices (N = 26) yielded significantly more accurate and fast text production with h-VBI than with g-VBI, while the performance of experts (N = 12) for both VBIs was nearly equal when a 0.8-1° key size was used. We discuss advantages and limitations of the VBIs for typing with ultra-small keyboards and emphasize relevant factors for designing such systems.

CCS CONCEPTS

Human-centered computing → Human computer interaction (HCI); *Interaction techniques*; Text input;

KEYWORDS

Gestural input, vision-based interfaces, eye typing, head tracking, hands-free text entry, assistive technologies

ACM Reference format:

Yulia Gizatdinova, Oleg Špakov, Outi Tuisku, Matthew Turk and Veikko Surakka. 2018. Gaze and head pointing for hands-free text entry: Applicability to ultra-small virtual keyboards. In *Proceedings of ACM Symposium on Eye Tracking Research & Applications (ETRA'18)*, ACM, Warsaw, Poland, June 2018, 9 pages. DOI: 10.1145/123 4

1 INTRODUCTION

Vision-based interfaces (VBIs) utilize visual cues of a user's body for direct computer control. VBIs support hands-free interaction and, therefore, have been widely leveraged in developing assistive technologies for disabled and/or elderly users who have difficulties in handling traditional keyboards and computer mice due to impairments in motor coordination of hands and weak muscle power [Porta 2002; Sears and Young 2003].

Hands-free text entry by means of voluntarily produced and controlled gestures and movements of eyes, face and head has received special attention in the research community. Thus, writing electronic texts with gaze-controlled VBIs (g-VBIs), so-called eye typing, has been thoroughly studied [Majaranta 2012; Majaranta and Räihä 2007; Räihä 2015]. Eye typing often involves gazing at a key of a virtual keyboard and dwelling the gaze on that key for about 1 s in order to activate it. For novices, a typical speed of eye typing with a static unambiguous virtual keyboard is 5-8 wpm [Majaranta 2012; Majaranta and Räihä 2007]. With a key press [Gizatdinova et al. 2012b] or other fast dwell-free key activation methods [Tuisku et al. 2016; Urbina and Huckauf 2007], the speed of eye typing for novices can increase up to 11 wpm. Experienced eye typists can set shorter dwells of 0.3-0.4 s and achieve the speed of 20 wpm, while individual typists may reach the speed of up to 24 wpm [Majaranta et al. 2009; Majaranta and Räihä 2007; Räihä and Ovaska 2012; Tuisku et al. 2016].

Head-controlled VBIs (h-VBIs) for text entry employ head motion (i.e., tracked position of the head, face or a facial feature from a video stream) as a pointing cue (a.k.a. camera mouse) and, frequently, a dwell for key selection. Text entry with h-VBIs is referred to as face typing in the following text. While technological solutions for h-VBIs have existed for a while, relatively recently researchers have started looking deeper into usability and user experience properties of h-VBIs. The speed of h-VBIs with static unambiguous virtual keyboards is in the range of 4-7 wpm when 0.5 s dwell or a key press is used for key activation [Betke et al. 2002; Gizatdinova et al. 2012b].

The adoption of a particular VBI for text entry is primarily defined by the physical abilities of an individual. Thus, g-VBI may be a preferable or even the only available means of writing electronic texts for users with extreme disabilities such as cerebral palsy or amyotrophic lateral sclerosis, which only allow these individuals to move their eyes [Sears and Young 2003; Su et al. 2005]. At the same time, certain eye conditions (e.g., nystagmus) may seriously compromise the voluntary control over eye movements, making interaction by gaze difficult or impossible [McDonald et al. 2013]. Imperfections of gaze-point estimation (e.g., calibration drifts) can make gaze-based interaction cumbersome. Head-controlled interfaces, on the other hand, are dependent on a user's capacity to perform fine head motions and require that at least some control over the neck (and sometimes even the torso) is preserved. Head-based interaction may be considered as an alternative for those users who are used to typing text, for instance, with a mouth stick [Koester and Levine 1996; Simpson 2013].

Earlier studies on hands-free text entry with g-VBI and h-VBI have been primarily performed with able-body participants due to the unavailability of large samples of impaired users. It was shown that able-bodied participants prefer eye typing over face typing due to its superior speed of text production, less physical fatigue, ease- and pleasantness-of-use [Gizatdinova et al. 2012b]. Head-controlled interfaces, on the other hand, appeared to be less prone to errors of text entry regardless of the used key layout or key activation mechanism, and availability of error correction function [Gizatdinova et al. 2012b; Hansen et al. 2004].

With the advent of portable computing devices with small screens, there has been a continuing trend in reducing the size of interface elements. In virtual keyboards, this allows for more keys in the layout and additional widgets such as menus and function keys. Small keyboards are desirable if a user wishes to see a large portion of the text (a paragraph or the whole page instead of few lines of text) without the need to scroll the text up or down, or to simultaneously interact with other on-screen applications (e.g., having two windows visible on a screen). Therefore, it is sometimes preferable that a keyboard occupies a rather small part of the screen, and a VBI supports comfortable text entry in these conditions. Gaze and head position tracking are already available in portable devices [Cuaresma and MacKenzie 2017; Kar and Corcoran 2017; Neel et al. 2012; Roig-Maimo et al. 2018; Zhang et al. 2017], opening up numerous possibilities for impaired users in accessing communication and information in hands-free mobile contexts.

The effects of key size on the accuracy of gaze pointing and head pointing have been studied in simple target acquisition tasks (see Table 1). In general, a deterioration of gaze pointing performance is expected for objects smaller than 1-2° [Miniotas et al. 2004]. MacKenzie [2012] studied the effects of object size reduction on error rate of target acquisition by gaze. The reported error rate of gaze was more than 35% for 1.2° targets and more than 75% for 0.6° targets. As he emphasized, the test condition with 0.6° targets approached the limits of human eye physiology and led to a dramatically increased error rate as well as a slower speed of gaze pointing as compared to the condition with the larger targets.

On the other hand, head pointing was shown to perform well with objects as small as 0.9° [De Silva et al. 2003] (0.4° with head-mounted enhancements for head tracking [Hansen et al. 2004, Radwin et al. 1990]). Darrell et al. [2002] investigated positional accuracy of pointing techniques and reported the results in terms of standard deviation from a predefined trajectory path as follows: mouse (0.06°), trackball (0.12°), 3D head tracker (0.18°), 2D head tracker (0.59-1.01°) and eye/gaze input (0.61-1.21°). The accuracy of modern 2D head trackers was reported as 0.2-0.3° in favorable conditions [Gizatdinova et al. 2012a; 2012b]. As Table 1 shows, gaze pointing with large targets [MacKenzie 2012] is faster than head pointing [De Silva et al. 2003]. (A direct comparison of the results is possible for studies with similar results for the baseline mouse condition, which indicates a similar methodology that the studies applied.) Importantly, the speed of head pointing seems to become comparable to or even higher than that of gaze pointing if smaller targets (0.6-1.2°) or activation times longer than 0.5 s are used.

Writing electronic texts is arguably a more control-demanding task than a simple target acquisition task. It is unclear how the findings from Table 1 generalize to the task of hands-free text entry with ultra-small keyboards, which implies pointing at small and densely located objects. To date, research on VBIs for hands-free text entry, especially eye typing studies, have been primarily performed with large (e.g., full-screen) keyboards and a key size is frequently not even reported. VBI developers merely make a key size large enough for eye trackers in use, which are usually labelled by manufacturers as providing the accuracy of 0.5-1° (6-10 mm when viewing from the distance of 65 cm). Note that the 0.5° accuracy requires nearly ideal tracking conditions, which are difficult to achieve in real-word interaction scenarios.

Despite attempts to study this issue, a dependency of error rate on the key size for text entry VBIs has not been clearly defined until now. Gizatdinova et al. [2012b] investigated face typing and eye typing with key sizes of 1.9°, 2.4° and 2.9° but found no statistically significant effects of key size reduction on the error rate of text entry for either g-VBIs or h-VBIs. Nevertheless, they hypothesized that the performance of g-VBI will likely deteriorate with a further decrease of key size, while such effect will not take place for h-VBI. The aim of this work was to continue this research to 1) systematically evaluate the effects of a key size reduction (0.4°, 0.6°, 0.8° and 1°) on the accuracy and speed performance of text entry with g-VBI and h-VBI and 2) discuss possibilities and directions of future research regarding the development of text entry VBIs for ultra-small keyboards.

Table 1: Impact of Gaze and Head Pointing on Throughput* for Small Sparsely Located Targets

Multidirectional tapping task	Partici-pants	Pointing	Selection**	Pointing distance***, °	Target size***, °	Throughput, bits/s
8 radial directions [Jagacinski and Monk 1985]	8 able-bodied	head tracker	0.34 s dwell	2.5, 4.3, 7.5	0.4, 0.7, 1.2	5
		joystick			0.3, 0.5, 0.9	7-8
8 radial directions [Radwin et al. 1990]	10 able-bodied	head tracker	0.63 s dwell	1.9, 10.5	0.3, 0.8, 2.3	4.7
		mouse				7.7
16 radial directions [De Silva et al. 2003]	8 able-bodied	head tracker	key press	10****	0.9****	2.0
		mouse				4.7
16 radial directions [MacKenzie 2012]	16 able-bodied	gaze	0.75 s dwell	7.7, 9.9	2.1, 2.8	2.3
		gaze	0.50 s dwell			3.1
		gaze	key press			3.8
		mouse				4.7
	12 able-bodied	gaze	0.5 s dwell	9.2, 18.3	0.6, 1.2	1.8
		gaze	0.5 s eye blink			1.2
		mouse	key press			4.8
20 radial directions [Roig-Maimo et al. 2018]	12 able-bodied	mobile head tracker	touch	4, 8, 16*****	1, 2*****	1.42
		device movement				1.2
11 radial directions [Cuaresma and MacKenzie 2017]	12 able-bodied	mobile head tracker	smile, blink, 2.5 s dwell	1.9, 3.7	0.4, 0.7	0.6, 0.59, 0.53

[*] Note that in different studies, a throughput might have been calculated using different Fitts' law variants.

[**] Clicking a regular mouse button or pressing a key of the physical keyboard takes 0.05 s on average [Betke et al. 2002].

[***] If not specified in the publication, a viewing distance of 60 cm was assumed between a user and a monitor.

[****] 15″ screen was assumed.

[*****] Distance of 40 cm was assumed

2 METHODS

2.1 Participants

Thirty-eight students and staff members from the University of Tampere volunteered to take part in the experiment. We divided participants into two groups based on their previous experience of eye tracking technology. The first group (N = 26) was composed of those participants who had no prior first-hand experience of eye tracking and were considered as novices regarding text entry tasks of the current study. This group included 12 males and 14 females of average age M = 26.4, SD = 6.6 (range 19-38 years). All had normal or corrected to normal vision (6 wore eye-glasses).

The second group (N = 12) was composed of mainly researchers of our gaze interaction group (7 males and 5 females of average age of M = 45.7, SD = 13.9; range 31-65 years; 6 participants wore eye-glasses), who were highly experienced in developing and evaluating various eye/gaze-based systems. Some of them had extensive (several hours) experience of eye typing. Some of them had about 30 min prior experience of typing with a large virtual keyboard using the VBIs under investigation. Experts were familiar with an eye tracker used in this study and knew the "tricks" of handling imperfect calibration as, for instance, gazing at a different location on the screen in order to point at a desired key.

All participants were highly experienced computer users and regularly used physical QWERTY keyboard as well as virtual keyboards of tablets and mobile phones. The participants were able-bodied, except one expert who had a severe motor disability (32 years old, female). This participant preserved a control over neck, face and, partly, torso, arms and hands. She used to type text using middle fingers of her both hands and, sometimes, voice recognition software.

2.2 Apparatus

The following hardware was used: desktop computer (Intel Core 2 quad, 2.66 GHz, 3 GB RAM), Tobii T60 eye tracker (60 Hz sampling rate) with its integrated 17″ monitor (1280 × 1024 pixel resolution) and Logitech Webcam Pro 9000 camera (320 × 240 pixel resolution, 25 frames per second (fps) capture rate).

2.2.1 Pointing and selection techniques of g-VBI and h-VBI. Gaze data from only left eye was used for pointing. This decision was based on the fact that gaze pointer computed as an average of gaze data from both eyes may have unwanted jumps in case the tracker loses/restores one of the eyes (there is always some difference between the estimated points for each eye separately). Although not very common, we supposed that this effect could have a negative impact when pointing at small keys. The distance from a participant's face to the monitor was about 65 cm during eye typing, therefore one degree of a visual angle corresponded to 1 cm on the screen surface (40 pixels). Akkil et al. [2014], using the same eye-tracker, same table-chair setup, same lightning, etc., reported that the median gaze point offset in the given conditions was 0.7 degrees (ranging from 0.4° to 1.2°).

Head-controlled interface used in this study was evaluated previously in real-time interaction scenarios [Gizatdinova et al. 2012b; Ilves et al. 2014]. The head pointer control was implemented based on a continuous face tracking from a video stream, using two tracking methods [Gizatdinova et al. 2012a]. Based on the results of pilot tests, h-VBI supported 25 fps and allowed to select targets as small as 0.2-0.3° (5-10 pixels), assuming favorable illumination conditions. The camera was fixed at the top border of the monitor, which was lowered down so that participants' eyes were located approximately at the camera level. Such camera setup helped to capture nearly frontal-view facial images and, therefore, supported good performance of computer vision methods for face processing in h-VBI. In addition, a non-invasive light source was located in front of a participant's face to further improve the performance of h-VBI.

As we were interested to isolate the performance of pointing from that of key activation and to remove possible errors caused by dwell-based activation mechanisms (e.g., a falsely activated key while in pursuit of the target key), key selection was executed by pressing the SPACE key of a physical keyboard (0.05 s duration [Betke et al. 2002]).

2.2.2 Virtual keyboard. A virtual keyboard[1] [Räihä 2015] layout included 29 letters of Finnish language, punctuation marks and additional controls: SPACE, BACKSPACE and READY (as illustrated in Fig. 1). The keys were visually represented as circles separated by a spatial gap of 0.5° (20 pixels). The pointing-sensitive areas of the keys were squares without any gaps in between (e.g., refer to a black bounding box around the key "f" in Fig. 1). On the periphery, the key pointing-sensitive areas were prolonged by approximately a visual size of the key in all possible directions (as it is shown for the key "a" in Fig. 1). The borders of the pointing-sensitive areas were not visible during the experiment.

The key sizes of 0.4°, 0.6°, 0.8° and 1° (15, 25, 32, and 40 pixels, correspondently) were empirically selected for the experiment. In face typing, the pointer was displayed as a dark red square of 0.25° (10 pixels) size. In eye typing, there was no visible pointer. This decision was based on the earlier findings showing that a visible pointer may distract users in gaze-based interaction, causing prolonged reaction times, false alarms and misses during visual letter search [Majaranta et al. 2006].

Visual feedback shown on the keys included three states: "neutral key," "focused key" (gaze hover the key) and "pressed key" (shown for 150 ms after the key selection), which all differed by the shades of a blue colour. The key selection was accompanied by a short "click" sound. If gaze or head pointer estimation failed, the keyboard appeared inactive until the pointer control was restored.

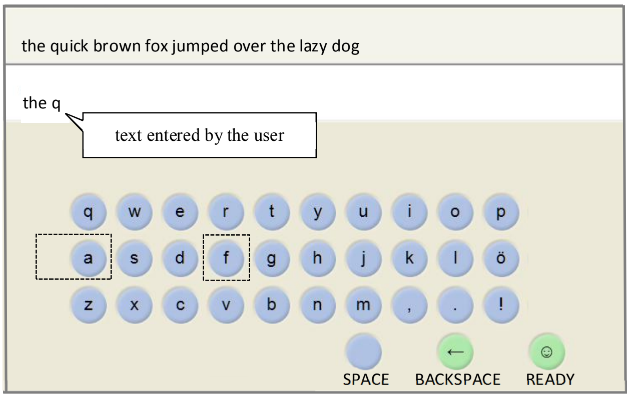
2.3 Experiment design, procedure and task

The experiment was a 2×4 within-subject repeated measures factorial design with the following independent variables and levels: interface (g-VBI and h-VBI) and key size or keyboard (0.4°, 0.6°, 0.8° and 1°).

The experiment consisted of two typing blocks (i.e., eye typing and face typing), each included transcribing a single phrase once with all four keyboards. The target phrase was "the quick brown fox jumped over the lazy dog," which involves pointing at and selecting the majority of the keys of the keyboard. The order of VBI presentation was counterbalanced. The keyboards were presented starting with the largest size, gradually moving towards the smallest size. The total number of phrases was 38 participants × 2 interfaces × 4 key sizes = 304.

Taking into account that a minimum rate of 3 wpm was defined for text entry interfaces to support a tolerable interactive conversation [Darragh and Witten 1992], the typing time for the phrase was limited to 4 minutes (i.e., the expected minimum speed of text entry was a bit less than 3 wpm).

After a participant arrived to the lab, s/he filled in consent and questionnaire forms and was explained the aims of the study, VBIs and virtual keyboard. In each typing block, a VBI was calibrated for each participant and a participant had a short practice trial of writing his/her own name. In eye typing blocks, the eye tracker was calibrated with 5 points. The average gaze point offsets from each calibration point were visually inspected by an experimenter right after the calibration was finished. Re-calibration was performed if multiple gaze points were further than the supposed maximum distance of 1 degree (i.e., not within a circle drawn around the calibration point). High accuracy was achieved by many participants, and there were no participants with >1 degree offsets dominating. For the details of h-VBI calibration procedure, refer to [Gizatdinova et al. 2012a; 2012b].



the quick brown fox jumped over the lazy dog

the q

Figure 1: A virtual keyboard with static QWERTY layout used in the study.

---

[1] AltTyping is available for download at www.sis.uta.fi/~csolsp/downloads.php.

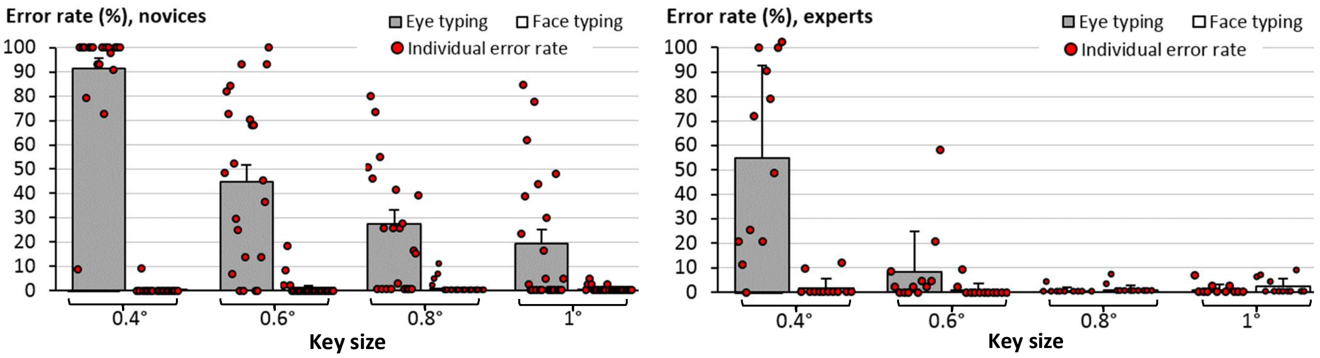


Figure 2: Columns show error rates (bars define 1 SEM from the means) of eye typing and face typing for novices (left) and experts (right). Scatter plots (red circles) account for error rates of individual typists.

Following this, a participant was instructed to read the target phrase, try to memorize it and type the phrase as fast as possible. A participant was told that there is the BACKSPACE key in the layout, but error correction was not explicitly required. A participant then typed the target phrase once with each keyboard and the next typing block proceeded in a similar fashion. A face processing window of h-VBI was visible to all participants during face typing blocks, so that a participant could verify that the software tracked the face. A participant was instructed to stay in the camera view in facial up-right frontal position, and move torso rather than rotate head while pointing at the keyboard keys.

No formal subjective assessments were collected, but many participants were willing to share their thoughts about the interfaces and their experiences of text entry in informal talks. We will refer to some of these notes in the discussion section.

2.4 Text entry metrics

Error rate was calculated as a ratio between a Levenshtein string distance and a total character (char) count in the target phrase. The Levenshtein string distance [Soukoreff and MacKenzie 2001] was defined as the minimum number of single-char edits required to change the transcribed phrase into the target phrase. The distance value was composed of the three types of errors: *deletions* (missed chars like 'e' in 'dsktop'), *insertions* (extra chars like 'floewer') and *substitutions* (erroneous chars like 'constraction') computed based on the detailed inspection of the Levenshtein matrixes. In counting single-char edits, a prioritization was given to insertions when error type was ambiguous, which did not change the resulting number of single-char edits. *Error-free performance* was defined strictly as the ability of a typist to output correct text after transcribing the phrase during 4 minutes. *Keystrokes per char* (KSPC) was measured as a total count of key presses (excluding the READY keystrokes) divided by the length of the transcribed text. *Text entry speed* in words per minute (wpm) was computed over a time interval between the first and the last entry of a char in the target phrase. Note that one word equals five chars, including SPACE and punctuation marks.

3 RESULTS

Four novices did not enter text with 0.4° key size due to technical problems and their results were omitted from the analysis of text entry metrics in this condition. Otherwise, no outlier removal procedure was applied to the data collected.

3.1 Error analysis

In this section, we present the results on errors and error correction first for novices, followed by the results for experts.

3.1.1 Novices. Fig. 2 (left) shows that novices wrote the target phrase rather correctly during face typing as compared to eye typing. Thus, error rate of novice face typists was 0.4-1.2% throughout all tested conditions. Noteworthy, 20 face typists (76%) were able to type the phrase without a single mistake using the smallest keyboard at the end of the experiment.

In contrast, error rate of novice eye typists was high, starting at 20% with the key size of 1° and reaching 91% with the key size of 0.4°. A half of novices rejected to continue eye typing as impossible to perform with the key size of 0.4°. None of novices was able to type the phrase without mistakes during the permitted 4 minutes of eye typing in this condition. Among novices (35%) who spent 4 minutes for entering the phrase with the smallest keyboard (i.e., did not reject the task), some wrote only few chars and did not proceed much further due to entering wrong chars and deleting those repeatedly. Some novices did not attempt to correct mistakes at all during eye typing because pointing at the BACKSPACE key by gaze was difficult. In such cases, the written phrase was 3-4 times longer than the target phrase. This was reflected in rather low KSPC values of novices during eye typing, as Fig. 3 (left) illustrates. During eye typing, novices also sometimes accidently pressed the READY key right in the beginning of writing the phrase.

For error rate of novices, a two-way 2 (interface: g-VBI and h-VBI) × 4 (key size: 0.4°, 0.6°, 0.8° and 1°) ANOVA showed significant main effects of interface ($F (1, 18) = 104.9$, $p < 0.001$, $\eta2 = 0.8$), key size ($F (3, 54) = 47.7$, $p < 0.001$, $\eta2 = 0.7$) and interface × key size interaction ($F (3, 54) = 48.38$, $p < 0.001$, $\eta2 = 0.7$). To find out where the interaction effect results from, separate one-way ANOVAs were performed for both interfaces within the key size factor. For gaze typing, a one-way ANOVA revealed a significant main effect of key size: $F (3, 57) = 44.01$, $p < 0.001$,

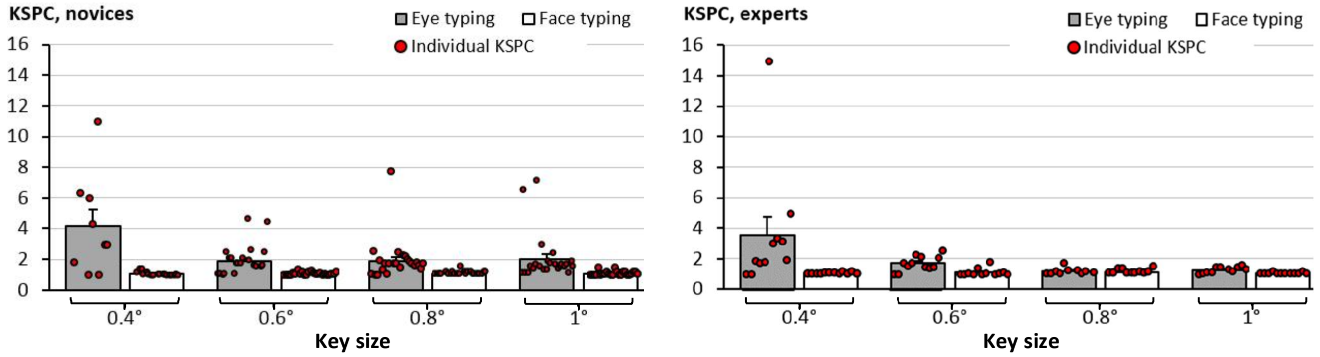


Figure 3: Columns show KSPC (bars define 1 SEM from the means) of eye typing and face typing for novices (left) and experts (right). Scatter plots (red circles) account for KSPC values of individual typists.

η2 = 0.7. Bonferroni corrected post-hoc pairwise comparisons showed that novices made significantly more errors with the key size of 0.4° than with the key size of 0.6° (MD = 51.71, p < 0.001), 0.8° (MD = 69.84, p < 0.001) and 1° (MD = 76.61, p < 0.001). In addition, they made more errors with the key size of 0.6° than with the key size of 0.8° (MD = 18.14, p < 0.05). For face typing, a one-way ANOVA for key size was not statistically significant, which indicates that key size did not impacted error rate of novices significantly.

The KSPC values for novices remained about 1 during face typing in all tested conditions (see Fig. 3, left), which corresponds to the fact that novices did not make mistakes during face typing (and the target phrase did not include capital letters). For KSPC value of novices, a two-way 2 × 4 ANOVA showed significant main effects of interface (F (1, 7) = 31.36, p < 0.01, η2 = 0.8), key size (F (3, 21) = 3.38, p < 0.05, η2 = 0.3) and interface × key size interaction (F (3, 21) = 3.56, p < 0.05, η2 = 0.3). For eye typing, a one-way ANOVA showed a significant main effect of key size: F (3, 24) = 3.29, p < 0.05, η2 = 0.3. Bonferroni corrected post-hoc pairwise comparisons were not statistically significant. For face typing, a one-way ANOVA was not statistically significant.

Fig. 4 (left) breaks down errors that novices left uncorrected in the final text output into deletions, insertions and substitutions for both interfaces as normalized relatively to a total char count in the target phrase. It is visible that novices frequently missed chars during eye typing (i.e., were not able to complete entering the phrase).

3.1.2 Experts. As Fig. 2 (right) illustrates, error rates of exerts for the key size of 0.8-1° were equally low during eye typing and face typing. Interestingly, the disabled participant (expert) produced a completely error-free text in the 0.6-1° conditions. The difference in the performance started to show with the key size of 0.6° and, especially, 0.4°. In the 0.4° condition, five experts (43%) ended up writing the phrase by gaze with error rate of less than 25%, while only one of them (the author of the current study) entered the phrase by gaze without a single mistake. Differently from novices, whose resulting text was hardly readable in case of ultra-small keyboards, some experts were able to start typing quite accurate text by gaze with the key size of 0.4°, although not always they had time to complete typing the phrase during the permitted 4 minutes. At the same time, ten experts (83%) did not leave any mistakes in the transcribed text using the smallest virtual keyboard during face typing (including the disabled participant).

For error rate of experts, a two-way 2 × 4 ANOVA showed significant main effects of interface (F (1, 8) = 10.59, p < 0.001, η2 = 0.7), key size (F (3, 24) = 15.06, p < 0.001, η2 = 0.7) and interface × key size interaction (F (3, 24) = 14.79, p < 0.001, η2 = 0.6). For eye typing, a one-way ANOVA showed a significant main effect of key size: F (3, 27) = 16.69, p < 0.001, η2 = 0.7.

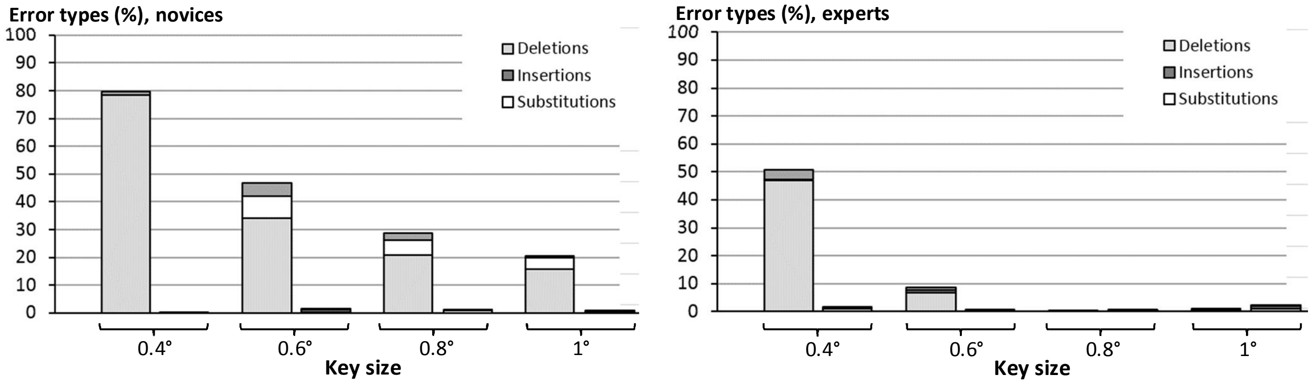


Figure 4: Columns show deletions, insertions and substitutions of eye typing (left bar for each key size) and face typing (right bar for each key size) for novices (left graph) and experts (right graph).

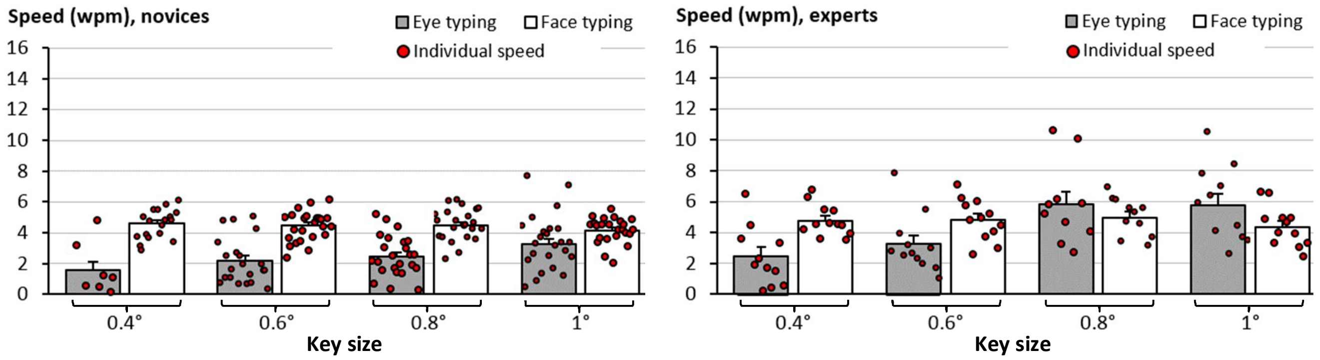


Figure 5: Columns show speed (bars define 1 SEM from the means) of eye typing and face typing for novices (left) and experts (right). Scatter plots (red circles) account for average speed of individual typists.

Bonferroni corrected post-hoc pairwise comparisons showed that experts made significantly more errors with the key size of 0.4° than with the key sizes of 0.6° (MD = 45.76, $p < 0.05$), 0.8° (MD = 53.43, $p < 0.05$) and 1° (MD = 52.77, $p < 0.05$). For face typing, a one-way ANOVA for key size was not statistically significant.

The KSPC values for experts remained about 1 during both eye typing and face typing in the 0.8-1° conditions, see Fig. 3 (right). Similar to novices, experts made on average 4 KSPC during eye typing in the 0.4° condition. For KSPC value of experts, a two-way 2 × 4 ANOVA was not statistically significant.

Error patterns in Fig. 4 (right) show that a high error rate of experts during eye typing in the 0.4° condition primarily originated from a high number of deletions (i.e., experts were not able to complete writing the phrase).

3.2 Speed analysis

3.2.1 Novices. The speed of face typing for novices remained on average about 4-5 wpm in all tested conditions (4.2 wpm for the 1° condition and 4.6 wpm for the 0.4° condition), refer to Fig. 5 (left). The best individual typing speed of face typing reached 6 wpm. It is worth noting that novices wrote the phrase virtually without mistakes with this speed. The speed of eye typing was 3.3 wpm for the key size of 1°, which decreased to 1.6 wpm for the key size of 0.4°. It should be noted that text written by novices during eye typing contained many mistakes and was either too short or excessively long. Even with the largest key size of 1°, the error rate of novices was still quite high on average (20%, as Fig. 2 (left) shows). For this reason, we consider that the results of novice eye typists may not describe reliably the actual typing speed.

A two-way 2 × 4 ANOVA showed significant main effects of interface ($F(1, 7) = 38.08$, $p < 0.001$, $\eta 2 = 0.8$) and interface × key size interaction ($F(3, 21) = 5.12$, $p < 0.01$, $\eta 2 = 0.4$). For gaze typing, a one-way ANOVA showed a main effect of key size: $F(3, 24) = 3.42$, $p < 0.05$, $\eta 2 = 0.3$. Bonferroni corrected post-hoc pairwise comparisons showed that novices typed text significantly faster with the key size of 0.8° than with the key size of 0.4°: MD = 1.35, $p < 0.05$. For face typing, a one-way ANOVA showed a main effect of key size: $F(3, 57) = 3.85$, $p < 0.05$, $\eta 2 = 0.2$. Bonferroni corrected post-hoc pairwise comparisons showed that novices typed text faster with the key size of 1° than with the key size of 0.6°: MD = 0.48, $p < 0.01$.

3.1.2 Experts. In face typing blocks, experts typed text with the speed of 4-5 wpm on average throughout the test. The best typing speed of experts during face typing reached 7 wpm. The speed of experts during eye typing was notably slower than during face typing in the 0.4-0.6° condition, see Fig. 5 (right). In the 0.8-1° condition, however, the speed of text production of experts during eye typing was about 6 wpm, which is a bit higher than that of face typing. The best individual speed of expert eye typists was 11 wpm.

A two-way 2 × 4 ANOVA showed significant main effects of key size ($F(3, 24) = 8.67$, $p < 0.001$, $\eta 2 = 0.5$) and interface × key size interaction ($F(3, 24) = 8.82$, $p < 0.001$, $\eta 2 = 0.5$). For eye typing, a one-way ANOVA showed a main effect of key size: $F(3, 24) = 10.97$, $p < 0.001$, $\eta 2 = 0.58$. Bonferroni corrected post-hoc pairwise comparisons showed that experts typed text faster with the key size of 1° than with the key size of 0.6° (MD = 2.49, $p < 0.05$) and 0.4° (MD = 3.52, $p < 0.05$). For face typing, a one-way ANOVA was not statistically significant.

4 DISCUSSION AND CONCLUSIONS

Today, communication between people moves more and more into social media applications on mobile platforms. In terms of text entry, this means that people increasingly use mobile instant messaging applications such as WhatsApp (www.whatsapp.com) or Messenger (www.messenger.com) with devices having a small screen size, for example, 4″-6″. In order to address this shift in communication methods and support accessibility of portable devices for all user groups, it is important to explore hands-free text entry in mobile environments. This study was one of the first attempts to investigate the effect of ultra-small keyboards on the performance of text-entry VBIs.

The results on error analysis revealed that the performance of h-VBI was virtually not affected by a key size reduction. The error rate of face typing remained low in all tested conditions, being 0.4-1.2% for novices and 0.8-2.2% - for experts. The average age of experts was higher than that of novices and this might have affected the results. Notably, error-free performance of face typing was achieved by 76% of novices and 83% of experts at the end of the test. In general, the obtained low error rate of face typing with ultra-small keyboards is comparable to the earlier results reported for similar but larger static keyboards [Gizatdinova et al. 2012b; Hansen et al. 2004].

On the contrary, a deteriorating effect of ultra-small keys on error-free performance of g-VBI was observed. A highly accurate commercial eye tracker used in this study did not support eye typing with ultra-small keys by novices. Noteworthy, many experts were able to write correct text by gaze even with the key size as small as 0.6°, which suggests that the limitations of human eye physiology reported in [MacKenzie 2012] could be, at least partly, compensated by sufficient practice of eye typing. However, for a majority of experts writing the phrase became much more challenging with the key size of 0.4°.

In eye typing, we expected that the participants would experience problems when selecting even the largest keys of 1° (R=0.5°), given the median offset was 0.7°. It was also expected that experts will apply some strategy to overcome calibration inaccuracies. Indeed, experts successfully applied it when targeting keys as small as 0.6° (R=0.3°). The common strategy was a "trial-and-error" strategy, meaning that after selecting a wrong char experts estimated the offset direction and length, and tried to compensate the inaccuracy by looking aside from the required char according to these estimations.

We note that experts emphasized eye dryness and the need for a high (many defined it as a painful) visual concentration during eye typing with the smallest key size. Some experts doubted the feasibility of prolonged eye typing with ultra-small keyboards as such due to strong eye fatigue.

We note also that apart from calibration drifts, another reason for difficulties of eye typing in this study could have been insufficient visual feedback. Thus, a difference between different shading of the blue colors of "neutral," "focused" and "pressed" key states during eye typing were difficult to visually distinguish especially in the 0.4° condition. We anticipate that the performance of g-VBI could be improved in case of a more pronounced and optimized visual feedback. However, it is difficult to predict whether an improved visual feedback will help eye typists achieve error-free performance in the 0.4° condition.

Despite a good error-free performance with h-VBI, many experts emphasized that face typing felt inconvenient, weird and fatiguing. The fact that experts were used to working with gaze-based interaction might have biased them to consider h-VBI as an alien technique. Novices did not express any particular preferences for either of the interfaces during this short typing practice but were impressed by the fact that h-VBI supported text production with ultra-small keyboards. On the other hand, the earlier studies with larger static keyboards have shown that novices prefer g-VBI over h-VBI as more efficient, less physically tiring, and easier- and more pleasant-to-use [Gizatdinova et al. 2012b]. A high speed of text production of g-VBI in those studies likely influenced the user ratings. It may be also that eye typing as such has many attracting qualities as compared to face typing, one of which, namely no need for physical motion, can be critical for disabled users. It would be interesting to see whether and how ultra-small keyboards would change user ratings of eye and face typing.

In eye typing, text entry speed has been often studied with longitudinal experiments to find out the maximum speed that participants are able to reach [Majaranta et al. 2009; Räihä and Ovaska 2012; Tuisku et al. 2016]. These earlier works with larger keyboards reported that novices reach the text entry speed of 4-10 wpm in the first session, and improve own performances up to 20+ wpm after intense practice. In the current study, the speed of text production of g-VBI was slower. Novices reached 2.5-3.3 wpm, while experts notably outperformed novices and achieved about twice faster speed of 6 wpm (with the best individual speed of 11 wpm). In the case of face typing, both novices and experts of eye typing (who are both novices of face typing) typed the phrase with the same speed of 4-5 wpm throughout the experiment.

It should be noted that a challenge of selecting a desired key is smaller with hand input as compared to dwell or some other activation mechanisms like visual gestures (e.g., eye blink). As a few experts mentioned, manual key activation was especially helpful during eye typing, because it was possible to "catch" the exact moment when the right key was in the activated state and make a fast key press by hand. This implies that g-VBI with a dwell- or gesture-based key activation may show even slower speed of text entry than in the current study. At the same time, the results of the face typing speed (varying from 4 to 5 wpm) appear to be a bit higher than those reported earlier for a similar but larger keyboard and a manual key selection, for instance, 3-4 wpm in Gizatdinova et al. [2012b]. This may indicate that pointing at smaller keys by head is easier and faster (e.g., less physical motion is required to move a pointer) than pointing at larger (1° and larger) keys.

The current study has a number of limitations. Firstly, our participants entered only one phrase with each keyboard, whereas in other studies larger phrase corpuses are usually used. This makes a direct comparison to the earlier studies difficult. Thus, to overcome this obvious limitation, we plan to run a formal longitudinal study in the future to study learning effects and systematically collect user experiences along with performance metrics of hands-free text-entry VBIs with ultra-small keyboards. In these future studies, it would be important to have a formal calibration verification procedure to assess the dependency of novices' ability to type text by gaze on the calibration quality of the eye tracker.

Secondly, our study was made in desktop settings. Nevertheless, we argue that the results are likely to be transferrable to mobile environments, although a loss of performance in terms of speed and, perhaps, accuracy may be expected. In the future, we plan to utilize existing head tracking APIs to implement and test text entry VBIs for portable devices. It would be beneficial, in our view, to consider combining gaze pointing and face pointing a.k.a. HMAGIC pointer [Kurauchi et al. 2015] for typing with ultra-small keyboards. In particular, fast but less accurate eye movements could be used to roughly position a pointer on a target, while slow but small in amplitude and accurate head motion could be used to accurately re-adjust the pointer and keep it on the small target until key activation is executed [Špakov et al. 2014].

To conclude, we investigated the effects of a key size reduction on the accuracy and speed performance of text entry with g-VBI and h-VBI. The results showed that h-VBI supported rather accurate and relatively fast text entry for both novices and experts even with the smallest key size of 0.4°. Our results suggest that computer users could potentially learn quickly to type correct text with ultra-small keyboards using h-VBIs. Moreover, there is a chance that computer users may be able to type quality text using h-VBIs even with keys that are smaller than 0.4°. Gaze-controlled interface supported a good performance of text entry only for experts of eye typing in case of the key size of 0.8-1°. Novice users would need to get sufficient practice in order to achieve similar performance as experts. We note, however, that keyboards with the key size of 0.4° are likely to be manageable only by h-VBIs. We see a promising performance potential in combining head and gaze input for text entry with small-screen portable devices, and we plan to perform a larger study in the future to derive specific guidelines for designing VBIs for ultra-small virtual keyboards.

ACKNOWLEDGMENTS

The authors thank Jenny and Antti Wihuri Foundation, Academy of Finland (decision #308929), University of Tampere, Lappeenranta University of Technology's Research Platform on Smart Services for Digitalization and University of California, Santa Barbara for support. The authors greatly acknowledge the effort of anonymous reviewers in improving this manuscript.

REFERENCES

Akkil D., Isokoski P., Kangas J., Rantala J., Raisamo R. 2014. TraQuMe: a tool for measuring the gaze tracking quality, Proceeding Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '14). ACM, 327-330.

Betke M., Gips J. and Fleming P. 2002. The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access for People with Severe Disabilities. IEEE Trans. Neural Systems and Rehabilitation Engineering, 10, 1, 1-10.

Cuaresma J. and MacKenzie I.S. 2017. FittsFace: Exploring navigation and selection methods for facial tracking. In Proceedings of the 19th International Conf. on Human-Computer Interaction, 403-416.

Darragh J.J. and Witten I.H. 1992. The reactive keyboard. Cambridge: Cambridge University Press.

Darrell T., Checka N., Oh A. and Morency L.-P. 2002. Exploring Vision-Based Interfaces: How to Use Your Head in Dual Pointing Tasks, AI Memo 2001-2002 (MIT Technical Report).

De Silva G.C., Lyons M.J., Kawato S., and Tetsutani N. 2003. Human factors evaluation of a vision-based facial gesture interface. In Proceedings of IEEE Computer Vision and Pattern Recognition Workshop. IEEE, 52-52.

Gizatdinova Y., Špakov O. and Surakka V. 2012. Face typing: Visual gesture-based perceptual interface for typing with a scrollable virtual keyboard. In Proceedings of the Workshop on the Applications of Computer Vision. IEEE, 81-87.

Gizatdinova Y., Špakov O. and Surakka V. 2012. Comparison of video-based pointing and selection techniques for hands-free text entry. In Proceedings of the International Working Conf. on Advanced Visual Interfaces. ACM, 132-139.

Hansen J.P., Tørning K., Johansen A.S., Itoh K. and Aoki H. 2004. Gaze typing compared with input by head and hand. In Proceedings of the Symposium on Eye tracking Research & Applications. ACM, 131-138.

Ilves M., Gizatdinova Y., Surakka V. and Vankka E. 2014. Head Movement and Facial Expressions as Game Input. Entertainment Computing, 5, 3, 147-156.

Jagacinski R.J. and Monk D.L. 1985. Fitts' Law in Two Dimensions with Hand and Head Movements. Journal of Motor Behavior, 17, 1, 77-95.

Kar A. and Corcoran P. 2017. A Review and Analysis of Eye-Gaze Estimation Systems, Algorithms and Performance Evaluation Methods in Consumer Platforms. IEEE Access, 5, 16495-16519, p14.

Koester H.H. and Levine S.P. 1996. Effect of a word prediction feature on user performance. AAC Augmentative and Alternative Communication, 12, 155-168.

Kurauchi A., Feng W., Morimoto C. and Betke M. 2015. HMAGIC: Head movement and gaze input cascaded pointing. In Proceedings of the International Conf. on Pervasive Technologies Related to Assistive Environments. ACM, 5 pages.

MacKenzie I. S. 2012. Evaluating Eye Tracking Systems for Computer Input. In Majaranta P., Aoki H., Donegan M., Hansen D.W., Hansen J.P., Hyrskykari A. and Räihä K.-J. (Eds.). Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies. Hershey, PA: IGI Global, 205-225.

Majaranta P. 2012. Communication and Text Entry by Gaze. In Majaranta et al. (Eds.), Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies, IGI Global, 63-77.

Majaranta P., Ahola U.-K. and Špakov O. 2009. Fast gaze typing with an adjustable dwell time. In Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems. ACM, New York, NY, USA, 357-360.

Majaranta P., MacKenzie S., Aula A. and Räihä K.J. 2006. Effects of Feedback and Dwell Time on Eye Typing Speed and Accuracy. Universal Access in the Information Society, 5, 2. Springer, 199-208.

Majaranta P. and Räihä K-J. 2007. Text Entry by Gaze: Utilizing Eye-tracking. In MacKenzie I.S. and Tanaka-Ishii K. (Eds.) Text entry systems: Mobility, accessibility, universality. San Francisco: Morgan Kaufmann, 175-187.

McDonald R., Thomacos N. and Inglis K. 2013. Review of Current and Emerging Assistive Technologies for the Reduction of Care Attendant Hours: Cost Effectiveness, Decision Making Tools and Emerging Practices. A Snapshot Review. Melbourne: ISCRR.

Miniotas D., Špakov O. and MacKenzie I.S. 2004. Eye gaze interaction with expanding targets. In Proceedings Extended Abstracts of the SIGCHI Conf. on Human Factors in Computing Systems. ACM, NY, 1255-1258.

Neel J., Kar A. and Cohen M. 2012. Looking at you: fused gyro and face tracking for viewing large imagery on mobile devices. In Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems. ACM, 2211-2220.

Porta M. 2002. Vision-based User Interfaces: Methods and Applications. Int. J. Human–Computer Studies, 57, 1. Elsevier, 27-73.

Radwin R.G., Vanderheiden G.C. and Lin M.L. 1990. A Method for Evaluating Head-Controlled Computer Input Devices Using Fitts' Law. Human Factors. 32, 4, 423-38.

Roig-Maimo M.F., MacKenzie I.S., Manresa-Yee C. and Varona J. 2018. Head-Tracking Interfaces on Mobile Devices: Evaluation Using Fitts' Law and a New Multi-Directional Corner Task for Small Displays. International J. of Human-Computer Studies, 112, 1-15.

Räihä K-J. 2015. Life in the fast lane: effect of language and calibration accuracy on the speed of text entry by gaze. In Proceedings of the Conf. on Human-Computer Interaction. Springer-Verlag, 402-417.

Räihä K.-J. and Ovaska S. 2012. An exploratory study of eye typing fundamentals: dwell time, text entry rate, errors, and workload. In Proceedings of the SIGCHI Conf. on Human Factors in Computing Systems. ACM, New York, 3001–3010.

Sears A. and Young M. 2003. Physical Disabilities and Computing Technologies: An Analysis of Impairments. The Human-Computer Interaction Handbook. L. Erlbaum Associates Inc. Hillsdale, NJ, USA, 482–503.

Simpson, R. 2013. Computer Access for People with Disabilities: A Human Factors Approach. CRC Press, Taylor and Francis Group.

Soukoreff R.W. and MacKenzie I.S. 2001. Measuring errors in text entry tasks: An application of the Levenshtein string distance statistic. Extended Abstracts of the ACM SIGCHI Conf. on Human Factors in Computing System (CHI'01), New York: ACM, 319-320.

Špakov O., Isokoski P. and Majaranta P. 2014. Look and lean: accurate head-assisted eye pointing. In Proceeding of the Symposium on Eye Tracking Research & Applications. ACM, 35-42.

Su M.C., Su S.Y., Chen G.D. 2005. A Low-Cost Vision-Based Human-Computer Interface for People with Severe Disabilities. Biomedical Engineering: Applications, Basis and Communications, 17, 6. World Scientific, 284-292.

Tuisku O., Rantanen V. and Surakka V. 2016. Longitudinal study on text entry by gazing and smiling. In Proceedings of the Eye Tracking Research & Applications ETRA, ACM Press, 253-256.

Urbina M.H. and Huckauf A. 2007. Dwell time free eye typing approaches. In Proceedings of the Conf. on Communication by Gaze Interaction. COGAIN'07, 65-70.

Zhang X., Kulkarni H. and Morris M.R. 2017. Smartphone-based gaze gesture communication for people with motor disabilities. In Proceedings of the 2017 CHI Conf. on Human Factors in Computing Systems. ACM, New York, NY, USA, 2878-2889.