Hindawi
Advances in Human-Computer Interaction
Volume 2023, Article ID 8855764, 23 pages
https://doi.org/10.1155/2023/8855764

Research Article

Vision-Based Interfaces for Character-Based Text Entry: Comparison of Errors and Error Correction Properties of Eye Typing and Head Typing

Yulia Gizatdinova,[1] **Oleg Špakov**,[1] **Outi Tuisku**,[1,2] **Matthew Turk**,[3] **and Veikko Surakka**[1]

[1]*Research Group for Emotions, Sociality and Computing, TAUCHI Research Center, Faculty of Information Technology and Communication Sciences, Tampere University, 33014 Tampere, Finland*
[2]*School of Industrial Engineering, Tampere University of Applied Sciences, 33520 Tampere, Finland*
[3]*Four Eyes Lab, Department of Computer Science, University of California, Santa Barbara, CA 93106-5110, USA*

Correspondence should be addressed to Yulia Gizatdinova; julia.kuosmanen@tuni.fi

Received 8 May 2023; Revised 5 September 2023; Accepted 10 October 2023; Published 22 November 2023

Academic Editor: Marco Porta

We examined two vision-based interfaces (VBIs) for performance and user experience during character-based text entry using an on-screen virtual keyboard. Head-based VBI uses head motion to steer the computer pointer and mouth-opening gestures to select the keyboard keys. Gaze-based VBI utilizes gaze for pointing at the keys and an adjustable dwell for key selection. The results showed that after three sessions (45 min of typing in total), able-bodied novice participants ($N = 34$) typed significantly slower yet yielded significantly more accurate text with head-based VBI with gaze-based VBIs. The analysis of errors and corrective actions relative to the spatial layout of the keyboard revealed a difference in the error correction behavior of the participants when typing using both interfaces. We estimated the error correction cost for both interfaces and suggested implications for the future use and improvement of VBIs for hands-free text entry.

1. Introduction

Vision-based interfaces (VBIs) actively and unobtrusively perceive visual cues about users, user actions, and the surrounding environment from real-time video frames captured by a camera [1, 2]. This study concentrates specifically on VBIs that perform video-based processing of the human face and/or head areas, which are referred to as vision-based *face interfaces*. Face interfaces recognize facial identities, estimate emotions from facial behaviors, and analyze gaze and head movements [3]. Technical progress in recent decades has significantly improved the accuracy and robustness of camera-based processing. Current vision-based face interfaces show potential for implicitly providing feedback or adjusting computer systems to the user's needs, for instance, as soon as a spontaneous user behavior has been recognized. Explicit or direct control over software applications through dedicated and voluntarily controlled gestures–for instance, head nodding or eye blinks as confirmation commands–has been integrated into several consumer products.

The hands-free interaction properties of VBIs are especially attractive when designing assistive and rehabilitation systems targeted at persons with motor disabilities and elderly users with functional impairments in mobility control, muscle power, and motor coordination of the hands [4, 5]. With VBIs, such individuals can use their preserved voluntary eye, face, and head motions to access technology-mediated information, communication, entertainment, and environmental control. Face interfaces can successfully emulate the pointing and clicking functions of conventional input devices. This enables their utilization in

writing electronic texts, which is considered one of the most desirable technology-enhanced activities [6–8]. *Hands-free text entry* with VBIs involves one or both of the following operations: (1) *camera-based pointing* at the elements of an on-screen spelling application (e.g., a virtual keyboard) and (2) *activation* of commands such as key selection, menu navigation, and changing character sets.

Text entry with *gaze-based* VBIs, also known as *eye typing*, typically requires looking at a key and then dwelling the gaze on that key for about one second or less to activate it [9]. Other hands-free alternatives for key activation exist, such as eye blinks [10, 11], facial movements detected by electrode measurement technology [3, 12–14], and switches and foot pedals [7]. Such methods have shown good potential in eye typing but gained less popularity than well-established dwell-time protocols. This may be because these methods are not yet fully noninvasive, robust, or comfortable enough to achieve wide adoption among researchers and end users of eye typing. In general, eye typing has evolved rapidly in the past and accumulated a substantial body of knowledge, methodologies, and empirical results [15–17].

Since 2000, evidence has accumulated that *head-based* VBIs can enhance or fully substitute for the functionality of gaze-based VBIs. Writing electronic text using head-based VBIs, referred to as *head typing* in the following text, usually employs head movements to steer a computer pointer. For this purpose, the position of the face or a facial feature (such as the nose tip) is usually tracked in the video. Key selection and other activation commands are executed either by dwelling or, alternatively, by head and/or face gestures detected from the video stream. The latter is naturally a special area of research in head typing. Owing to the variability of facial expressions (as well as head poses) that can be controlled voluntarily, a rich set of commands can potentially be designed [3, 12, 13], whereas dwell time (when used without additional aids, such as graphical toolbars) can only substitute for a single command, typically a mouse button click. Noteworthy, previous research has shown that text entry commands based on facial behavior can be executed independently and simultaneously (such as scrolling keyboard rows by lowering and raising eyebrows and making key presses through a mouth-opening gesture [18]).

In this study, we examine whether and how error correction with gaze- and head-based VBIs compromises text entry performance and user satisfaction, with a special focus on head-controlled VBI. This is because user evaluations of text entry with head-based VBIs are still rare, and errors and error correction properties of head typing have not been thoroughly investigated, as it is discussed in Section 2. At the same time, errors of eye typing have been investigated relatively well in the past [9, 15, 16, 19] and can serve as a reference point for head typing. Such a comparison is useful for obtaining insights into the limitations and advantages of both types of VBIs in different settings and for different typists. Thus, we specifically aimed to (i) verify the ability of head- and gaze-based VBIs to support error-free performance of text entry, (ii) study strategies for identifying and correcting errors and estimate the relative cost of error correction for both VBIs, and (iii) analyze errors relative to the spatial keyboard layout to reveal whether certain layout characteristics are more likely to lead to errors than others. Because the field lacks a comprehensive analysis of state-of-the-art in camera-based head typing, we review relevant research from the past two decades and present the results in Section 2. In addition, we discuss the applicability of the examined VBIs for text entry and identify factors for optimizing such systems.

2. Head Typing: Text Entry with Head-Based VBI Technology

Tables 1 and 2 show state-of-the-art in head typing, namely, methodological details and key findings of user studies from the last two decades. We concentrated solely on head-based VBI technology that applies camera-based processing to calculate a head pointer and/or facial activations (e.g., head motion calculated via inertial sensors of AR/VR helmets is, therefore, outside the scope of the current review). We analyze the literature based on the keyboard layouts used and adopt keyboard classification by Poláček et al. [45]. Table 1 presents the results for *static* (unambiguous) keyboards that support direct selection techniques (a single character is assigned to each key, and a single keystroke is sufficient to enter any character from a given layout). For example, a well-known QWERTY keyboard with a standard static layout was used in studies #2–5, 7–10, and 12 (Table 1).

Table 2 overviews head typing with dynamic keyboards (ambiguous, encoding, and scanning layouts) [45]. Dynamic keyboards are usually implemented with (i) the reduced number of keys (so that more than 1 "click" is needed to enter a single character of text (studies #1, 4, 5, 7, and 12, Table 2), (ii) dynamic change of key position (or size) in the layout (study #2, Table 2), (iii) a scanning interface in which a desired character (or word) appears or is highlighted automatically while the user selects it through head/face gestures (studies #3, 7, 8, 10, and 14, Table 2), (iv) gesture-based interfaces that adopt head gestures to draw letters directly or predict a word based on the head-pointer trajectory while it scans over the keyboard (study #13, Table 2), or (v) a binary spelling interface, such as Morse code, in which sequences of head gestures denote dots and dashes of the encoded communication (studies #9 and 11, Table 2).

It is interesting that while hands-free text entry systems primarily target users with disabilities, 21 out of 26 interfaces reviewed in Tables 1 and 2 were only tested with able-bodied participants. This comes from an assumption that individuals who can potentially use head pointing as an input method typically preserve a relatively large neck range of motion and do not have strong head tremors. These users can use alternative assistive technologies that imply movements of the head, and partly the torso, such as mouth/head sticks [4, 5, 7]. As a result, able-bodied participants are considered a good representative of this target group of people with disabilities in terms of head pointer control. Still, those studies that directly compared the performance of able-bodied typists and those with motion disabilities reported lower text entry rates for participants with disabilities as in studies #6a, b and 7 in Table 1 [5, 45].

Table 1: Speed and accuracy of head typing with static keyboards.

Spelling application	Authors	Participants[i]	Pointing method	Activation method	Words[ii] of transcribed text per participant	Speed[iii]	Accuracy	
							Error rate[iv]	KSPC[v]
#1 Camera mouse: alphabetical layout; no letter/word prediction; error correction allowed	[20]	20	Head (eyes, nose, or lip)	0.5 s dwell	8	6.3	—	—
#2 QWERTY layout; no letter/word prediction; error correction forced	[21]	24	Head (face)	Pneumatic switch	56	4.5 (for trials with no mistakes left)	34 corr	—
#3 ScreenDoors 2000: QWERTY layout; no letter/word prediction; error correction allowed	[22]	8	Head	Unknown (alternatives are puff on the tube, switch, or dwell)	Unknown	4.6	5.4	—
#4 Windows XP keyboard: QWERTY layout; no letter/word prediction; error correction allowed	[23]	10	Head (eyes)	Adjustable dwell (1 s default)	1	2.2	—	—
#5 QWERTY layout; no letter/word prediction; error correction allowed	[24]	5	Head (face, nose)	Adjustable dwell (1 s default)	2-3	2.6–3.2	—	—
#6a ScreenEdge: One-row alphabetical layout; no letter/word prediction; error correction allowed	[25]	6	Head (face)	Head gestures	~16	3.5		
		11 disabled				2	—	—
#6b Gnome keyboard: alphabetical layout; no letter/word prediction; error correction allowed	[25]	6	Head (face)	0.05 s key press[vi]	~16	2.9		
		11 disabled				1.6	—	—
#7 QWERTY layout; no letter/word prediction; error correction allowed	[26]	17	Head (eyes) Method 1	Mouth open	2	7.8		
			Method 2	0.5 s dwell		4.3	—	—
		17 disabled	Method 1	Mouth open		3.6	—	—
			Method 2	0.5 s dwell		2.1	—	—
#8 QWERTY layout; no letter/word prediction; no error correction	[27]	15	Head (face)	0.05 s key press[vi]	90	4.4	3.8	—
			Gaze			11	8	—
		13	Head (face)	Brows up	30	2.9	21	—
				Mouth open		3.1	6	—
#9 QWERTY layout with an extra small key size of 0.4°; no letter/word prediction; error correction allowed	[28]	26	Head (face)	0.05 s key press[vi]	9	4.6	0.4	1.1
			Gaze			1.6	91	4.2
		12 eye tracking experts	Head (face)			4.8	1.7	1
			Gaze			2.5	55	3.6
#10 QWERTY layout; no letter/word prediction; error correction allowed	[29]	5 experienced	Head	Eye blink	50	20–30 (55 for experts)	1.35	1.24
#11 8pen-based circular layout: no letter/word prediction; error correction forced	[30]	8	Head (face)	Head gestures	1.6	4.5	0.2 corr	
#12 HGaze: QWERTY layout, word prediction, error correction allowed	[31]	11	Head (eyes)	Head gestures	114	11.5	0.37	—
				0.6 s dwell		9	0.21	

[i]Participants are able-bodied novices if not specified otherwise. [ii]One "word" equals five characters, including spaces and punctuation marks. [iii]Speed measure (WPM) is obtained by multiplying characters per second by 60 (seconds per minute) and dividing by 5 (characters per word). [iv]Mean uncorrected error rate (%) is usually calculated as a total number of errors left uncorrected in the text divided by the total number of typed characters. Note that for techniques with forced error correction, the error rate is computed using all errors made during text entry (marked as "corr" in the table). [v]Keystrokes per character (KSPC) is usually calculated as a ratio of keystrokes used to correct errors to a total number of keystrokes used to type a sentence. [vi]Clicking a regular mouse button or pressing a key of the physical keyboard takes 0.05 s on average [20]. [vii]The entire word is entered by a single head gesture: two selections are used per word, one to identify the beginning of the word and another to identify the end of the word.

Only user studies with experimental results are included in Tables 1 and 2. Whenever possible, we compared empirical results of head typing and eye typing in terms of common standards of text entry evaluation, focusing on the overall error rate (as a percentage of misspelled characters left uncorrected in the text) and speed of text entry (traditionally measured in words per minute (WPM) where one "word" equals five symbols including spaces and punctuation marks) [45]. Some results provided in the tables were available in the original papers as numerical data, and some were inferred from the figures and/or computed by us. It should be noted that the results of text entry evaluation depend on multiple factors, such as video processing methods, experimental setups, key layout designs, phrase corpora, and population samples. Owing to the variability of these factors across different publications, the results of the individual studies presented in the tables should be interpreted and compared cautiously.

2.1. Errors of Head Typing. Error correction in VBIs is executed without the use of hands and may require significant effort from users, including cognitive (planning) and motor operations [8, 46]. In text entry studies, participants usually transcribe a set of short model phrases and correct errors immediately after they appear in the text with a backspace key. In real-world text entries, however, errors may be distributed throughout the text. Correcting errors in somewhat longer text segments is virtually indistinguishable from general text editing [46], which often requires, for example, relocating a pointer to an arbitrary place in the text, selecting parts of the text, cut/copy/paste operations, or performing undo/repeat functions. To address this challenge, some authors proposed additional keys and gestures to cover editing functionalities or even introduced text-editing tools with extended graphical interfaces (e.g., [47]). Research is still needed in this area, as additional graphical aids and increased gesture vocabularies may

Table 2: Speed and accuracy of head typing with dynamic keyboards.

Spelling application	Authors	Participants[i]	Pointing method	Activation method	Words[ii] of transcribed text per participant	Speed[iii]	Accuracy	
							Error rate[iv]	KSPC[v]
#1: SpeechStaggered: dynamic spelling layout; two-level selection; no letter/word prediction; error correction allowed	[32]	10	Head (eye, nose, or lip)	0.5-1.5 s dwell	3.6	4.9	—	—
#2: Dasher layout: letter prediction model	[33]	2	Head (nose)	—	160	7.3	—	—
		1 expert				12	—	—
#3: BlinkLink: On-screen keyboard with a two-level scanning; letters in alphabetical order; no error correction	[10]	15 disabled	—	2 eye-blinks	2	1.3	—	—
#4: MouthType: telephone keypad (two-level selection); error correction forced	[34]	2 experts	—	Mouth shape + 0.05 s key press[vi]	~500	12	3.1 corr	—
				Hand			1.9 corr	—
#5: GazeTalk: dynamic layout; letter/word prediction; two-level selection; error correction allowed	[35]	12	Head	0.5 s dwell	7.1	6.1	0.5	3.5
			Gaze			6.3	1.1	4.3
#6: FaceMouse: dynamic layout with char prediction; derivative pointing; error correction allowed	[36]	10 disabled	Head (nose)	Adjustable dwell	5	2.7	—	—
#7: Spelling board with three-level scanning; no error correction	[37]	4	—	3 eye-blinks (>~0.4 s)	1	0.5	—	—
#8: BlinkWrite2: Spelling board with a two-level scanning; letters arranged according to occurrence frequency; word prediction; error correction allowed	[38]	12	—	2 eye-blinks (>0.85 s)	~900	5.3	—	—
#9: Morse code: encoding application with a double-switch code; no error correction	[39]	2	—	Tongue protrusions to left/right	140	2	5–28	—
#10: b-Link: On-screen keyboard with two-level scanning; letters arranged according to their frequency of occurrence; no error correction	[40]	49 (12 disabled)	—	2 eye blinks (>0.20 s; <0.25 s)	1.6	1	0.4	—
#11: Morse code: error correction allowed	[41]	20	—	Head gestures + smile, mouth opening for deletions	90	4.9	—	—
#12: Three-level dynamic layout with visible or thermal imaging and hierarchical letter selection; no letter/word prediction; error correction forced	[42]	14	Head (face)	Head gestures	5.6	2	0.5 corr	—
#13: Nosype: smartphone dynamic layout, word prediction; error correction allowed	[43]	10	Head (nose)	Nose gesture [vii]	~15	6.5	—	—
#14 Scanning interface with QWERTY layout; word prediction; error correction allowed	[44]	10		Head gestures	6	2.9	—	—

[i]Participants are able-bodied novices if not specified otherwise. [ii]One "word" equals five characters, including spaces and punctuation marks. [iii]Speed measure (WPM) is obtained by multiplying characters per second by 60 (seconds per minute) and dividing by 5 (characters per word). [iv]Mean uncorrected error rate (%) is usually calculated as a total number of errors left uncorrected in the text divided by the total number of typed characters. Note that for techniques with forced error correction, the error rate is computed using all errors made during text entry (marked as "corr" in the table). [v]Keystrokes per character (KSPC) is usually calculated as a ratio of keystrokes used to correct errors to a total number of keystrokes used to type a sentence. [vi]Clicking a regular mouse button or pressing a key of the physical keyboard takes 0.05 s on average [20]. [viii]The entire word is entered by a single head gesture: two selections are used per word, one to identify the beginning of the word and another to identify the end of the word.

deteriorate text entry productivity and users' experiences (e.g., deleting text by mistake) [8]. So far, as long as easy and comfortable error correction remains an unresolved issue in text entry VBIs, error-free properties of such interfaces appear to be highly desirable, at least among certain user populations.

Little attention has been paid to systematically analyze errors of text entry and error correction strategies of head typists. As Tables 1 and 2 show, earlier studies mainly concentrated on the speed of head typing and overlooked its accuracy. Less than a half out of 26 studies reviewed in Tables 1 and 2 analyzed errors of head typing and reported simple quantitative characteristics such as error rates. In studies #8 (Table 1) and #5 (Table 2), which directly compared eye and head typing, eye typing resulted in approximately twice as many errors as head typing, regardless of the activation method, layout used, and availability of error correction function. Furthermore, in study #9 (Table 1), the keyboard size clearly affected the error performance of text entry VBIs. Their participants typed rather correct text during head typing, even with very small keys, while eye typing became nearly impossible in this condition owing to a high error rate. Similar findings were reported in study #5 (Table 2) (as well as by Jagacinski and Monk [48] and Radwin et al. [49] for directional tapping tasks by the gaze and head). Only few studies measured the effort required to generate text using head-based entry methods. Studies #9-10 (Table 1) and #5 (Table 2) reported the keystrokes per character (KSPC) measure, which was higher for eye typing than for head typing. This may indicate that, while eye typists

tend to write electronic texts faster than head typists, they may need to make more corrections to the typed text at the end. Interestingly, study #2 (Table 1) reported a significant improvement in the speed of text entry, while the number of errors did not change significantly with practice. A possible explanation here is a mental trade-off between speed and errors (when users sacrifice accuracy for speed [50]). However, it would be interesting to identify specific sources of errors, study the error correction process and its effects on text entry productivity, and obtain insights for optimizing the interaction methods and layouts used.

To conclude, the evidence from Tables 1 and 2 suggests that head-based VBIs may be less error-prone than gaze-based VBIs, despite their slower text production speed. This observation could open prospects for head-based VBI utilization in scenarios where error-free text entry performance is critical. In this study, we further extend the investigation of errors of head typing initiated in studies #8 and 9 (Table 1) with in-depth evaluation of spatial locations of errors relative to keyboard layout, computation of numerous metrics such as KSPC, error-free performance, and backspace corrections, and estimation of the relative cost of error correction for both gaze- and head-based VBIs. We present extensive results regarding error correction behavior of head typists, including a person with a disability.

2.2. Speed of Head Typing. Speed of head typing was researched well in the past. We summarize the findings in this section and compare those to the speed of eye typing. As Table 1 shows, head typists enter electronic text with a speed of 2–8 WPM without character/word prediction and 11.5 WPM if prediction models are used (study #12, Table 1). Research has recently been conducted to eliminate the need for camera-based face detection per se while preserving the use of video-based techniques for pointer control. In study #10 (Table 1), a camera was placed on the user's head to capture an image of the surrounding environment (i.e., computer screen). Head movements resulted in changes in the camera view, which was analyzed to compute the position of the head relative to the screen. The speed of head typing was reported as 20–30 WPM for five experienced users (reached 55 WPM with practice). This interface allows for fast text entry but may not function if there are moving objects in the background view of the camera.

For comparison, a speed of 22 WPM was theorized for eye typing with static unambiguous layouts without the use of prediction models, assuming 0.5 s dwell and 0.04 s average saccade duration [15]. Dwell-free eye typing for such keyboards was theorized to reach 46 WPM [51]. These simulations imply monotonous text entry, entering text one character after another without an active visual search of the keyboard layout, inspection of the written text, or error correction. In practice, however, the typical speed of dwell-based eye typing with static keyboards for novice users is 5–10 WPM [52], which can increase with adjustable (cascading) dwells or other fast dwell-free activation methods (such as pressing a physical key) up to 11–20 WPM [11, 27].

Table 2 shows speed parameters of head typing with dynamic keyboards as varying in a range of 1–12 WPM. Thus, head-controlled Dasher supported 7 WPM (up to 12 WPM for experienced head typists) in study #2 (Table 2). The authors theorized a typing speed of 24 WPM for their system with experienced head typists and a well-optimized letter/word prediction model. For comparison, gaze-controlled Dasher (without prediction) allows novice users, after some practice, to write text with an average speed of 17 WPM (23 WPM for experienced users) [53], with further increase possible when letter/word predictions are used.

These numbers suggest that eye typing tends to outperform head typing in terms of text production efficiency (if not considering error rates). However, only rarely a direct comparison between eye typing and head typing was performed in the past. The results differ among different studies. Study #8 (Table 1) reported a significant speed superiority of eye typing compared to head typing on a static keyboard with large keys, without an error correction option, and a key press as an activation command. Opposite results were achieved on a keyboard with small keys in study #9 (Table 1). In study #5 (Table 2), nearly equal speeds for eye and head typing were reported using a dynamic layout with word prediction and an error correction function.

3. Methods

3.1. Participants. Thirty-three unpaid university students without motor disabilities (24 males and 9 females) aged between 18 and 47 years ($M = 26.7$, SD = 7.5) volunteered to participate in the experiment. Thirty participants were native Finnish language speakers, and three participants were non-Finnish speakers who had previously taken basic courses in Finnish. All participants had normal or corrected-to-normal vision (seven participants wore eyeglasses). The participants had no prior experience with the VBIs under investigation and were considered novices regarding text entry tasks in the current study. All participants were highly experienced computer users and regularly used physical QWERTY as well as virtual keyboards on tablets and mobile phones.

In addition, a person with a motor disability (32 years old, female, native Finnish) participated in this experiment. This participant maintained good control over her neck, face, and, partly, arms and hands. She was an expert in eye tracking and had approximately 30 min of prior experience typing with both VBIs under investigation (eye-tracking experts are users who have previous experience with gaze-based VBIs and know, for example, how to handle imperfect calibration of an eye tracker by gazing at a slightly different location on the screen to point at the desired target).

3.2. Apparatus. The following hardware was used: a desktop computer (Intel Core 2 quad, 2.66 GHz, 3 GB RAM), a Tobii T60 eye tracker (60 Hz sampling rate) with a 17″ monitor (1280 × 1024 pixels), and a Logitech Webcam Pro 9000 camera (320 × 240 pixels, 25 fps).

3.2.1. Gaze-Based VBI. In gaze-based VBI, the gaze point was calculated by using $\{x, y\}$ position of the pupil in the left eye. As reported by the manufacturer, the accuracy of the eye tracker is 0.5–1° (1° corresponds to approximately 1 cm on a computer monitor viewed at 65 cm), assuming a nearly perfect calibration.

Differently from other works which typically use averaging and smoothing filters to compute the gaze pointer, we utilized a dwell accumulation algorithm to define which key was currently "in focus" and to further execute key activation as described in [16, 19]. Simply put, a voting mechanism was applied, in which the keys competitively collect a predefined number of "votes"[36]. Each keyboard key has its own dwell accumulation counter, which is set to zero $S(t)_i = 0$, $i = 0 \div 39$ (keys in the layout) at t_0. Each time a gaze sample arrives from the eye tracker, it is mapped naively on the keyboard layout and the dwell accumulation counter of the key, which area was hit by the gaze, increases by the amount of time Δt passed from the previous gaze sample n to the current one $(n + 1)$:

$$S(t)_{\text{current}} = \sum_{n=0}^{N} \Delta t, \tag{1}$$

where $\Delta t = t_{n+1} - t_n$ and N is a number of raw gaze samples needed to activate a key. All other dwell accumulation counters simultaneously decrease by the same amount of time or remain at zero:

$$S(t)_i = -\sum_{n=0}^{N} \Delta t, S(t)_i \geq 0, \quad i \neq \text{current}. \tag{2}$$

A key with the biggest dwell accumulation counter at that moment is visualized as "focused" (as described in Section 3.2.3). Once a counter exceeds a predefined dwell time D, $S(t)_i \geq D$, the corresponding key becomes activated, and all counters are reset to zero $S(t)_i = 0$. Note that because the keys collect gaze points competitively, the time T needed for key activation may exceed a fixed dwell time D. Thus, the total number of gaze samples needed for key activation is $N = T/\Delta t$.

3.2.2. Head-Based VBI. A head-controlled interface was previously described and evaluated in real-time interaction scenarios [18, 27, 54]. Head pointer control was based on continuous face tracking from a video stream (25 Hz sampling rate) using two tracking methods [18, 55], as shown in Figure 1. Based on pilot tests, the head pointer allowed the selection of targets as small as 5–10 pixels (0.1–0.3°), assuming favorable illumination conditions.

The mouth-opening gesture served as key activation. The selection of the gesture was based on the consideration that the face, in addition to the hands, is well represented in the cortical sensorimotor strip of the human brain. The lower face (the lips, jaw, and tongue) is richly represented in the brain's sensorimotor cortex, is better innervated, and has more complex sensory and motor connections than the upper face (the forehead, eyes, and brows) [56]. This allows for a more voluntary and learned control of the lower face, which is required, for instance, for mastication, speech production, and articulation. This suggests that lower face gestures may serve well as activation commands in text entry applications. In the past, mouth and tongue gestures were used for hands-free text entry (Tables 1 and 2). Gizatdinova et al. [27] studied two facial expressions as activation mechanisms in the context of text entry and reported that mouth opening was significantly more accurate than brows-up activation (although the speeds of both methods were similar). Mouth opening was rated highly by the participants as being used for frequent key selections. In addition, from a technical perspective, mouth opening produces a visual pattern that is relatively easy to detect by computer vision methods compared to, for instance, brow-up gestures that are barely distinguishable from the neutral state for some individuals [27].

Mouth-opening gesture detection was implemented using a segmented region of the lower face, as shown in Figure 1. The false-positive and false-negative misdetection rates were below 10% [18]. Considering that the average duration of a voluntary gesture, such as mouth opening, is 500 ± 200 ms in the key activation context [18], it was assumed that the gesture detector will not cause noticeable latencies in text entry. Also, study #7 (Table 1) showed that mouth-opening gesture resulted into a faster text entry as compared to 0.5 s dwell for both able-bodied users and users with disabilities.

3.2.3. Target Phrases and Virtual Keyboard. Following a standard methodology of text entry evaluation, target phrases were taken from a large representative corpus of approximately 500 by MacKenzie and Soukoreff [59]. Examples are "what a monkey sees a monkey will do" or "I can see the rings on Saturn." The first sentence consists of eight words of the English language and, at the same time, 35 characters, which make exactly seven "words" in total. Note that the definition of a "word" here is a segment of text five characters long, including spaces and punctuation marks. The length of the second sentence is 30 characters or six "words" (seven words of the English language).

The phrase corpus was first published by MacKenzie and Soukoreff [58], and then translated to Finnish by Isokoski and Linden [59], resulting into 14 565 characters, of which 425 are capital letters. As it was shown that writing in native language is preferred for optimal text entry performance and fewer errors [59], the Finnish corpus was used in the experiment with our Finnish-speaking participants. The phrases of the Finnish corpus consist of, on average, 28 characters, which amount to approximately six "words" per phrase. The frequency of characters of the Finnish phrase corpus corresponds to the character frequency of common Finnish texts with the most frequent characters "*a*" (10.2%), "*i*" (9.4%), SPACE (9.3%), "*t*" (8.1%), "*n*" (7.0%), "*e*" (6.4%), and "*s*" (6.1%).

As stated by Poláček et al. [45], static keyboards require less cognitive effort than dynamic keyboards because static layouts do not change with the context and users may memorize key distributions of static keyboards relatively easily. Because our participants were assumed to be experienced users of a conventional QWERTY layout, and the use of unfamiliar layouts was previously shown as provoking errors of text entry [60], a virtual keyboard with a QWERTY layout [16] was used. The layout included letters of the Finnish language, punctuation marks, and additional controls: SHIFT key, a dwell-time display and a control widget, SPACE key, BACKSPACE key, and READY "☺" key (refer to Figure 2(a)), altogether 39 keys. An adjustable dwell time was used for key selection during eye typing [52]. A dedicated control widget of the virtual keyboard allowed users to change the default dwell value $D = 1$ second [15].

Language models for word prediction are useful for improving text entry speed. However, none were used in this study because we aimed to compare the efficacy of character-level text entry of gaze- and head-based VBIs, which implies extensive use of keyboard layouts. Moreover, in transcription typing with word prediction, the use of backspace deletes the last input, which results in the deletion of either a single character or an entire word, thereby compromising the error correction analysis of the two VBIs under investigation.

To perform a fair comparison between the VBIs, a key size that was large enough to compensate for possible inaccuracies in gaze pointing was selected. Informed by previous studies [27, 28, 61], the key size was set to 55 pixels (1.38°), resulting in a total keyboard size of 605×220 pixels ($15.1° \times 5.5°$). The keys were visually represented as circles separated by a spatial gap of 20 pixels (0.5°). Pointing-sensitive areas of the keys were squares without any gaps in between (e.g., refer to a black-bounding box around the key "ö" in Figure 2(a)). On the periphery, the key pointing-sensitive areas were prolonged by approximately the visual size of the key in all possible directions, as shown for the keys on the right side of the keyboard in Figure 2(a). The borders of the pointing-sensitive areas were not visible during the experiment.

Figure 2(b) illustrates visual feedback shown on the keys. A key "in focus," that is, a key with the largest dwell time accumulated (in eye typing) or simply hit by the head pointer (in head typing), was visualized as a bulged key. The pressed key displayed a slightly darker blue shade for 150 ms after activation. Key selection was accompanied by a short "click" sound. In head typing, the pointer was displayed as a dark red square with a size of 10 pixels (0.25°), whereas in eye typing, there was no visible pointer because earlier findings showed that a visible pointer distracts users in gaze-based interactions, causing prolonged reaction times, false alarms, and character misses during visual letter searches [62]. Instead, a visualization of the elapsing dwell time was used: the key "in focus" displayed a growing red arc that helped in estimating the time the user needs to gaze at the key to activate it (refer to Figure 2(b)). If gaze or head pointer estimation failed, the keyboard appeared inactive until pointer control was restored.

3.3. Procedure. The experiments were conducted under controlled laboratory conditions; the participants typed text in a test room, while the experimenter was in an adjacent room with a one-way observation mirror. The participants' progress in text-typing tasks was monitored using a duplicate monitor. The study consisted of altogether three typing sessions, which were separated by an interval of not less than one and not more than two weeks. Each typing session lasted one hour.

The Ethics Committee of the Tampere Region gave a positive statement to this research (statement 36/2018). In the first session, the participants were informed about the study and completed a consent form and background questionnaire. The session continued with the first typing block, in which the keyboard layout was explained and one of the interfaces was calibrated (for details on the calibration procedure, refer to Gizatdinova et al. [27]). The participants received a demonstration of the typing technique and practiced briefly by typing their own names.

During head typing, the participants were instructed to avoid strong head rotations and tilts and to move the torso to ease head pointer control. A video stream captured by the camera with an overlaid face-processing output (Figure 1) was visible below the on-screen keyboard. The camera was fixed at the top border of the monitor, and the participants were seated in a way that makes his/her eyes level with the camera. This helped capture nearly frontal-view facial images and, therefore, supported the performance of computer vision methods used for face processing. In addition, a noninvasive light source was placed in front of the participant's face to further improve the performance of head-based VBI.

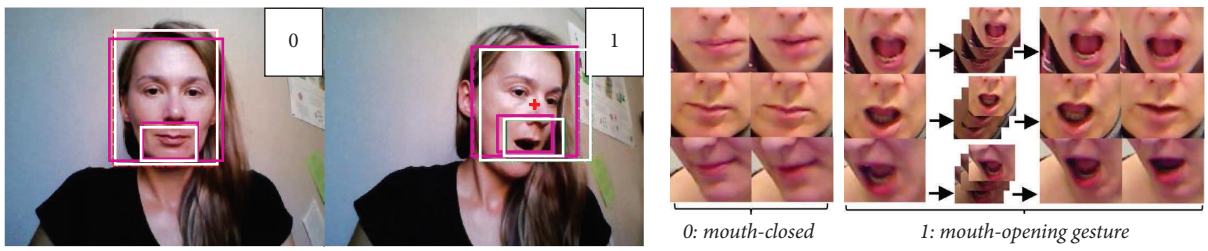
For eye typing, the participants were seated approximately 65 cm from the monitor, and their eyes approximately level with its center. The participants were instructed not to move their heads significantly because large head movements are known to worsen the calibration of the eye tracker. No special equipment (e.g., headrest) was used.

After calibration, the participants received instructions regarding the actual typing task, which emphasized that *the correctness of text is more important than the speed of typing*. Nevertheless, the participants were allowed to make errors and decide whether to correct errors that occurred, for example, at the beginning of a sentence. This encouraged typical typing behavior and enabled the analysis of the entire input stream, including errors and error corrections.

Thus, the participants were asked to correct their mistakes using the BACKSPACE key whenever they noticed an error; they were also instructed to memorize the target phrase at the beginning of each typing task so that they would not spend time repeatedly looking at the phrase. This was done to improve typing speed and decrease the unintentional selection of keys during eye typing, which may occur if eye typists make frequent glances across the keyboard. Next, the participants typed phrases randomly selected from the phrase corpus for 15 minutes.

After completing the first typing block, the participants rated their subjective experiences using bipolar rating scales (see Section 3.4). After this, the participants proceeded with the calibration, practice, text entry tasks, and ratings of the second typing block using another VBI. The order of VBIs was counterbalanced between the participants and their lab visits. At the end of the session, the participants compared their overall typing experiences with both VBIs using the pairwise preference form (see Section 3.4) and underwent a free-form interview (see Section 3.4).

3.4. Design. The experiment had a 2×3 within-subjects design. The independent variables and their levels were as follows:



(a) (b)

Figure 1: (a) Facial regions detected by two face trackers (white and violet-bounding boxes) and a combined pointer (red cross). (b) Sequences of no gestures (labelled as 0) and mouth-opening gestures (labelled as 1).

(i) Interface: head-based (head typing with a mouth gesture) and gaze-based (eye typing with adjustable dwell)

(ii) Session: 1st, 2nd, and 3rd sessions

It would have been interesting to test two additional conditions, namely, head typing with adjustable dwell and eye typing with a mouth gesture. However, we predicted that adding two more conditions would have extended the length of the experiment beyond tolerance of the participants. For this reason, the experiment design was limited to two conditions only.

The average length of the phrases transcribed in this study was in the range of 22–41 characters ($M = 28$, SD = 3.5), which is equivalent to 4–8 words per phrase. With 34 participants, the total number of characters typed was $34 \times 2 \times 3 \times 28 = 5\ 712$.

The following dependent variables were examined. We analyzed the effectiveness of text entry based on the number of words (i.e., chunks of five characters long) transcribed by the participants. Accuracy measures were defined as follows. *Error-free performance* was defined strictly as the ability of a typist to output the correct text after transcribing phrases for 15 minutes. The *error rate* accounted for the uncorrected errors in the transcribed text. It was calculated as the ratio between the Levenshtein string distance and the total character count in the target phrase. The Levenshtein string distance [63] was defined as a minimum number of single-character edits required to transform a transcribed phrase into a target phrase. The distance value was the sum of the three types of errors: *deletions* (e.g., missed characters such as "*e*" in a word "dsktop"), *insertions* (e.g., extra characters such as "floewer"), and *substitutions* (e.g., erroneous characters such as "constraction") [64]. In addition to the standard accuracy measures, we made in-depth analysis of spatial distribution of different errors relative to the keyboard layout as inspired by Räihä and Ovaska [19].

We defined the measures of text production efficiency as follows: *Text entry speed* in words per minute (WPM) was computed over the time interval between the first and last entries of a character in a given phrase. The *keystrokes per character* (KSPC) metric [65] was measured as the total count of key presses (excluding the READY and dwell-time control keys) divided by the number of entered characters (including the SHIFT and BACKSPACE keys).

In text-transcribing studies, a predominant number of errors (e.g., about 99% [46]) are corrected with a backspace key, even if other methods are available such as keyboard shortcuts, navigation and deletion keys, or mice. Based on this consideration, a new metric called *corrective action* was introduced and analyzed relative to each entered character (corrective actions per character (CAPC)). Corrective action occurs when a typist notices an error (not necessarily the last typed character) and attempts to correct it using a backspace key. Hence, the length of the corrective action is equal to the number of consecutive (uninterrupted) backspace key presses required to remove the erroneous character(s) (multiple erroneous characters can be removed within a single corrective action). The higher the CAPC values, the more frequently error correction distracted the typist from typing. Short corrective actions indicate that the typist made frequent error checks during typing. Long lengths would likely indicate that the typist checked errors, for example, only after entering the entire phrase.

Subjective ratings were collected using nine *bipolar rating scales*, a *pairwise preference questionnaire*, and a *free-form interview*. The scales were *general evaluation*, *difficulty*, *quickness*, *accuracy*, *pleasantness*, *efficiency*, *distractibility*, *mental effort*, and *physical effort*, varying from −4 (negative evaluation) to 4 (positive evaluation). A pairwise preference questionnaire was used to assess which interface the participants favored as *better in general*, *more difficult*, *quicker*, *more accurate*, *more pleasant*, *more efficient*, *more distracting*, *more mentally difficult*, and *more physically tiring for text entry*. Finally, the pairwise preference questionnaire had three alternative forced choices: (1) *I prefer gaze-based VBI*, (2) *I prefer head-based VBI*, and (3) *I have no preferences about the current interfaces*. Common questions of the free-form interview were, for instance, "*What was easy/difficult about gaze-based VBI and head-based VBI*?," "*Would you type text using the proposed interfaces in public places where other people can see you*?," and "*Would you like to use the interfaces in applications other than text typing*?"

4. Results

Data from one participant were excluded from the analysis in the third session because of technical problems. The collected data were analyzed for outliers using Grubbs' exclusion criterion [66] as follows: if a participant's

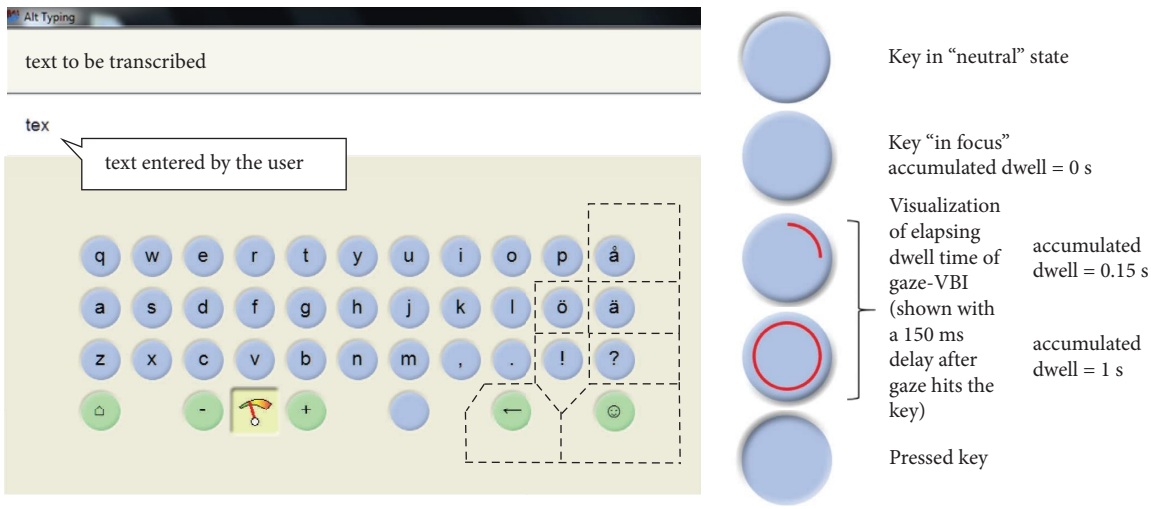


Figure 2: (a) QWERTY keyboard layout used in the study. (b) Visual feedback shown on the keys.

individual error rate averaged over a typing block was larger than three standard deviations from the mean value calculated from the data of all participants in this block, individual data points were considered outliers and excluded from the analysis of text entry metrics and subjective evaluation of this block. Across all sessions, the exclusion analysis revealed six outliers for eye typing and six outliers for head typing. The excluded data were mainly from the participants who accidentally pressed the READY key at the beginning of writing a phrase.

In the following, the results are expressed as mean values ± standard errors of the means (SEMs) and standard deviation (SD). A two (interfaces: head-based VBI and gaze-based VBI) × three (sessions: 1st, 2nd, and 3rd sessions) two-way repeated measures analysis of variance (ANOVA) was used to compare quantitative metrics of text entry. The Bonferroni-corrected t-test was used for post hoc pairwise comparisons. Estimates of effect size r were categorized as follows: 0.2: small effect, 0.5: medium effect, and 0.8: large effect [67], and they are reported together with a mean difference (MD) and a 95% confidence interval (95% CI). For the main (interface × session) interaction effect, one-way within-subjects ANOVAs were run separately on the eye-typing and head-typing data within the session factor.

The Friedman test was used to compare subjective ratings for eye typing and head typing. In case of a statistically significant effect, the Wilcoxon signed-rank test was used for pairwise comparisons. The Bonferroni correction was applied to p values (i.e., for a significance level of $p < 0.05$, the p value needed to be $0.05/15 = 0.003$ or less for the pairwise comparison to be statistically significant). To shorten the text, only significant results are reported numerically. The results for the participant with motor disability are presented separately in Section 4.7.

4.1. Error Analysis

4.1.1. Error Rate. The average error rate over all three sessions was $1.3 \pm 0.2\%$ (SD = 2.0) for eye typing and $0.4 \pm 0.09\%$ (SD = 0.8) for head typing. The error rates for each interface, averaged over the three sessions, are shown in Figure 3. ANOVA showed a statistically significant main effect of interface: $F(1, 20) = 16.4$, $p < 0.01$, $\eta_p^2 = 0.5$. Post hoc pairwise comparisons of the interface showed that the participants left significantly more errors in the transcribed text during eye typing than during head typing (MD = 0.9, 95% CI (0.4, 1.3), $p < 0.01$, $r = 0.6$). Figure 3 also shows the relative proportions of deletions, insertions, and substitutions for both interfaces, computed based on a detailed inspection of the Levenshtein matrixes.

4.1.2. Error-Free Performance. The circles in Figure 3 illustrate the participants' individual error rates. At the end of the last session, 6 participants (18%) using gaze-based VBI and 19 participants (58%) using head-based VBI produced the correct text without a single mistake (error rate = 0.0). The total number of error-free phrases is listed in Table 3.

4.1.3. Error Types. Figure 4 shows the deletions, insertions, and substitutions for both interfaces normalized relative to the total character count in the target phrase. For deletions, ANOVA showed a statistically significant main effect of the interface factor: $F(1, 20) = 9.6$, $p < 0.01$, $\eta_p^2 = 0.3$. Post hoc pairwise comparisons of the interface factor showed that the participants made significantly more deletions during eye typing than during head typing (MD = 0.7, 95% CI (0.2, 1.1), $p < 0.01$, $r = 0.6$). For substitutions, ANOVA showed a statistically significant main effect of the session factor: $F(2, 40) = 5.3$, $p < 0.05$, $\eta_p^2 = 0.2$. Post hoc pairwise comparisons of the session factor were not statistically significant.

4.2. Spatial Distribution of Uncorrected Errors. The spatial distributions of deletions (i.e., character misses) and erroneous selections (i.e., insertions and substitutions combined) relative to the keyboard layout are shown in Figures 5 and 6, respectively. The size of the blobs is proportional to the number of errors and normalized with respect to the total length of the transcribed text for each interface in each session. This accounts for the fact that eye typists wrote twice longer text than head typists and allows for a direct comparison of the results between the figures. The total count of errors for each session is depicted in the figures.

As Figure 5 (top-left) shows, many deletions during the first session of eye typing occurred for the punctuation mark "." and SPACE keys. The participants continued to miss these characters while typing by gaze in session 2 and, to a smaller extent, in session 3. An isolated location of the error cluster ("*a*," "*s*") (see both Figures 5 and 6) suggests that frequently used characters "*a*" and "*s*" were miss-hit interchangeably during eye typing, meaning in some cases the pointer landed on "*a*" instead of "*s*" and vice versa. Similar tendency can be observed during eye typing for other neighboring keys of the layout (e.g., another stable cluster of errors is ("*i*," "*o*," "*l*," ".")). The correlation between spatial locations of uncorrected errors made during the last session of eye typing and the character frequency of the phrase corpus is 0.6 and 0.5 for deletions and erroneous selections, correspondingly.

In head typing, the error cluster ("*a*," "*s*") is also present among the deletion errors, although its scale is smaller than that in eye typing. Misses in the SPACE key and punctuation marks occurred less frequently during head typing than during eye typing. There is a single stable cluster of erroneous selections ("*i*," "*k*," "*l*"). The correlation between uncorrected errors of head typing and the character frequency of the phrase corpus during the last session is 0.8 and 0.4 for deletions and erroneous selections, correspondingly.

4.3. Keystrokes and Corrective Actions. The analysis of committed but corrected (not visible in the output text) errors revealed the grand mean KSPCs averaged over the three sessions were rather similar for eye typing (1.4 ± 0.1 (SD = 0.7)) and head typing (1.3 ± 0.05 (SD = 0.3)). In the last session, the total backspace keystrokes accounted for 6.7% and 6.9% of the total keystrokes (excluding the SHFT, dwell adjustment, and READY function keystrokes) for eye typing and head typing, respectively.

These numbers suggest that the error correction behavior of the participants was similar when typing with both interfaces. However, analysis of the spatial distribution of corrective actions relative to the layout revealed differences between the interfaces. Figures 7 and 8 illustrate the CAPC values and average lengths of corrective actions (backspace counts) for eye typing and head typing. The figures show the characteristics of corrective actions (chains of backspacing) relative to the character that was the target of correction, ignoring all other deleted characters. The blobs in Figure 7 are normalized to the total length of the transcribed text for each interface in each session.

In the first session of eye typing, the participants corrected some characters (e.g., "*g*" and "*h*" in session 1, Figure 7) frequently, but the average length of corrective actions for these characters (refer to Figure 8) was not very long, implying that the participants corrected errors right away after typing these characters erroneously. More than 94% of the time, the average length of corrective actions was two characters or fewer for both interfaces. In contrast, some characters were rarely corrected, but error correction involved the deletion of relatively long portions of the text. The longest corrective action of 34 backspace keystrokes was recorded for "*a*" during the second session of eye typing. The absolute number of corrective actions increased with each session of eye typing; however, as the amount of written text steadily increased, the CAPC values remained nearly the same.

The patterns of the corrective action characteristics appeared stable across all sessions for head typing, as shown in Figures 7 and 8. The longest corrective action of 16 backspace keystrokes was observed for the SPACE key during the second session of head typing. Both the total number of corrective actions and the CAPC values steadily decreased with time for head typing.

4.4. Error Correction Cost. The effort required to write error-free text was approximated based on the prediction model of error correction cost for character-based text entry techniques [46]. The model predicts the extra time (in seconds) required, on average, per character to correct errors, regardless of whether a mistake was made on that character (Figure 9). The following approximations were made: (i) the distribution of the probability to notice and correct errors is exponential, and (ii) WPM accounts for both cognitive (planning and decision-making) and motor timings during text entry:

$$\text{Error–correction cost} = \frac{T_{\text{correct}} * \rho_{\text{error}} * \rho_{\text{char}}}{(1 - \rho_{\text{error}}) * (1 - \rho_{\text{char}})^2}, \quad (3)$$

where T_{correct} predicts the time in seconds necessary to correct an erroneous char in a single attempt:

$$T_{\text{correct}} = \frac{60}{\text{WPM} * 5} * (\text{KSPC} + 1), \quad (4)$$

where ρ_{error} is approximated by the total error rate calculated as the ratio between the total number of incorrect and corrected characters and the total effort required to enter the text. The probability to notice and correct an error ρ_{char} right away (the length of a corrective action equals 1) in our study is 0.8 for eye typing and 0.7 for head typing, which is higher than in the earlier study [46].

4.5. Text Entry Speed. The grand mean of text entry speed averaged over all sessions was 6.8 ± 0.3 WPM (SD = 2.7) for eye typing and 3 ± 0.1 WPM (SD = 0.8) for head typing. At the end, the two eye typists reached a maximum speed of 13 WPM, while the fastest head typist was able to type text at a speed of 5 WPM, as illustrated in Figure 10. The dwell time

for eye typing gradually decreased with increasing typing speed. The average dwell time of eye typing in the last session was 0.7 ± 0.05 s (SD = 0.3). Seven participants typed with dwell less than 0.5 s (two participants set dwell less than 0.3 s), and two participants increased dwell up to 1.2 s.

ANOVA showed a statistically significant main effect of the interface ($F(1, 20) = 133.9$, $p < 0.001$, $\eta_p^2 = 0.9$) and session ($F(2, 40) = 6.7$, $p < 0.01$, $\eta_p^2 = 0.3$). Post hoc pairwise comparisons of the interface factor showed that the participants typed text significantly faster during eye typing than during head typing (MD = 4.4, 95% CI (3.6, 5.2), $p < 0.001$), with a large effect size ($r = 0.9$). Post hoc pairwise comparisons of the session factor showed that the participants typed text significantly faster in session 2 (MD = 1.3, 95% CI (0.5, 2.2), $p < 0.01$, $r = 0.5$) and in session 3 (MD = 1.6, 95% CI (0.2, 2.9), $p < 0.05$, $r = 0.5$) than in session 1.

4.6. Subjective Evaluation

4.6.1. Bipolar Rating Scales. Figure 11 shows the means (as circles) and medians (as dividers within the boxes) of the participants' responses to the bipolar rating scales at the end of session 3. A positive number on the scale defines positive evaluations. The whiskers in the figure indicate the 25% and 75% quartiles, extending to the minimum and maximum scores in each evaluation category (i.e., half of the responses fell within each box). The bold outlines of the boxes indicate responses that fell below the median.

Finally, 82% of the participants reported a higher than neutral general evaluation of eye typing, while 53% rated their general experience with head typing as positive. Figure 11 shows that the mean (and median) scores of quickness and efficiency were both high for eye typing, but ended up on average at the first level of negative evaluation for head typing. The Friedman test showed that there were statistically significant differences between the ratings of general evaluation ($\chi^2(5) = 36.2$, $p < 0.001$), difficulty ($\chi^2(5) = 11.8$, $p < 0.05$), quickness ($\chi^2(5) = 81.3$, $p < 0.001$), pleasantness ($\chi^2(5) = 14.3$, $p < 0.05$), and efficiency ($\chi^2(5) = 56.5$, $p < 0.001$).

For the general ratings, the Wilcoxon signed-rank test showed that the participants rated gaze typing session 1 ($Z = 3.61$, $p < 0.05$, $r = 0.5$), session 2 ($Z = 3.36$, $p < 0.05$, $r = 0.5$), and session 3 ($Z = 3.27$, $p < 0.05$, $r = 0.5$) as significantly better than head typing session 1. They also rated gaze typing session 3 as significantly better than head typing session 2 ($Z = 3.22$, $p < 0.05$, $r = 0.5$).

For the quickness ratings, the Wilcoxon signed-rank test showed that the participants rated gaze typing session 1 faster than head typing session 1 ($Z = 4.18$, $p < 0.05$, $r = 0.6$), session 2 ($Z = 3.92$, $p < 0.05$, $r = 0.5$), or session 3 ($Z = 3.62$, $p < 0.05$, $r = 0.5$). They also rated gaze typing session 2 faster than head typing session 1 ($Z = 4.33$, $p < 0.05$, $r = 0.6$), session 2 ($Z = 4.05$, $p < 0.05$, $r = 0.5$), or session 3 ($Z = 4.28$, $p < 0.05$, $r = 0.6$). Similarly, gaze typing session 3 was rated faster than head typing session 1 ($Z = 4.39$, $p < 0.05$, $r = 0.6$), session 2 ($Z = 4.37$, $p < 0.05$, $r = 0.6$), or session 3 ($Z = 4.41$, $p < 0.05$, $r = 0.6$). They also rated head typing in session 3 faster than in session 1 ($Z = 3.57$, $p < 0.05$, $r = 0.5$).

Similarly, for the efficiency ratings, the Wilcoxon signed-rank test showed that the participants rated gaze typing session 1 more efficient than head typing session 1 ($Z = 3.24$, $p < 0.05$, $r = 0.4$) and session 2 ($Z = 3.09$, $p < 0.05$, $r = 0.4$). The participants rated gaze typing session 2 more efficient than head typing session 1 ($Z = 4.32$, $p < 0.05$, $r = 0.6$), session 2 ($Z = 4.22$, $p < 0.05$, $r = 0.6$), or session 3 ($Z = 3.95$, $p < 0.05$, $r = 0.5$). Finally, they rated gaze typing session 3 more efficient than head typing session 1 ($Z = 4.05$, $p < 0.05$, $r = 0.6$), session 2 ($Z = 4.11$, $p < 0.05$, $r = 0.6$), or session 3 ($Z = 4.04$, $p < 0.05$, $r = 0.6$).

For the pleasantness ratings, the Wilcoxon signed-rank test showed that the participants rated gaze typing in session 1 ($Z = 3.13$, $p < 0.05$, $r = 0.4$) and session 3 ($Z = 3.21$, $p < 0.05$, $r = 0.5$) as more pleasant than head typing in session 1.

4.6.2. Pairwise Comparison Questionnaire. Figure 12 shows the responses to the pairwise comparison questionnaire that the participants answered at the end of session 3. These responses are generally in line with the bipolar subjective scores shown in Figure 11, favoring eye typing in general, especially in terms of the speed and efficiency of text production. For the final preference judgment about text entry interfaces, most participants (73%) preferred eye typing and 17% preferred head typing.

4.6.3. Final Free-Form Interview. The interviews revealed several issues. First, learning the pointing and selection methods was easy for both VBIs. The participants remembered how to operate the interfaces during sessions 2 and 3. Second, the participants liked that during head typing, they were able to freely inspect the written text. Some participants suggested that this feature of head-based VBI would find better use in applications other than text entry such as web browsing or video gaming. Third, several participants emphasized good learning and speed improvement during eye typing but not head typing. Several participants mentioned that they could type faster with head-based VBI if the mouth-opening gesture had the option of adjusting its speed, similar to how dwell was adjusted during eye typing. In addition, the participants wished to obtain better feedback about the current state of mouth-opening gesture detection (i.e., a clear indication that the mouth was still recognized by the system as open). Fourth, the participants recognized head typing as tiring for the shoulders and neck area, while eye typing as causing eye

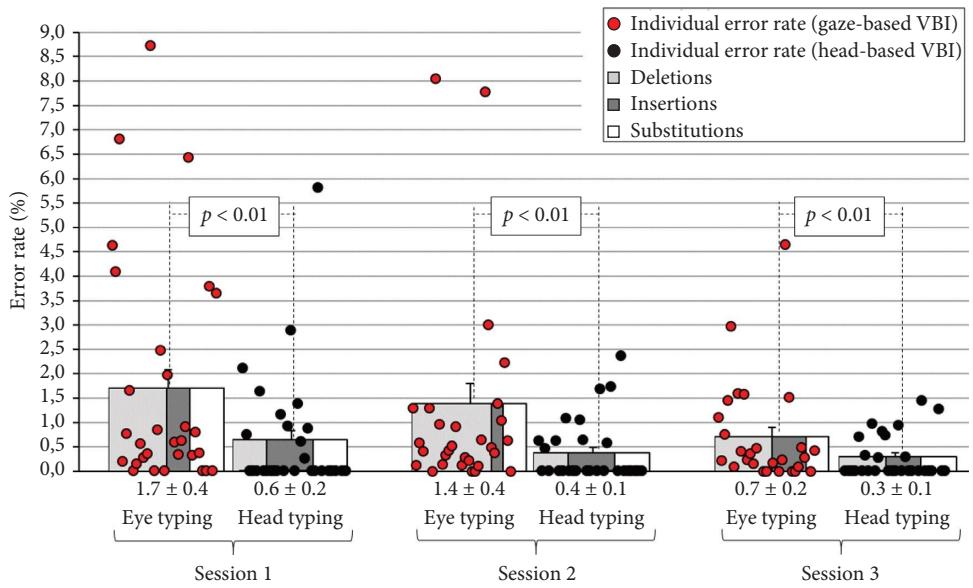


Figure 3: Columns show error rates for the three sessions of text entry (error bars define 1 SEM from the mean values) and relative proportions of deletions, insertions, and substitutions. Scatter plots show individual error rates of participants.

Table 3: Length of the transcribed text and error-free performance.

	Session 1	Session 2	Session 3
Eye typing			
Total phrase count	384	501	513
Total word count	2170	2863	2937
Average words per participant	68 ± 5, SD = 30	93 ± 6, SD = 31.5	105 ± 8, SD = 40
Error-free phrase count	320 (83%)	443 (88%)	465 (90%)
Total error-free word count	1794	2503	2666
Average error-free words per participant	56 ± 6, SD = 33	81 ± 6, SD = 35	95 ± 8, SD = 42
Head typing			
Total phrase count	179	234	244
Total word count	1022	1322	1438
Average words per participant	33 ± 2, SD = 10	44 ± 2, SD = 10	48 ± 2, SD = 12
Error-free phrase count	163 (91%)	219 (94%)	229 (93%)
Total error-free word count	928	1231	1338
Average error-free words per participant	30 ± 2, SD = 12	41 ± 2, SD = 11	45 ± 2.5, SD = 14

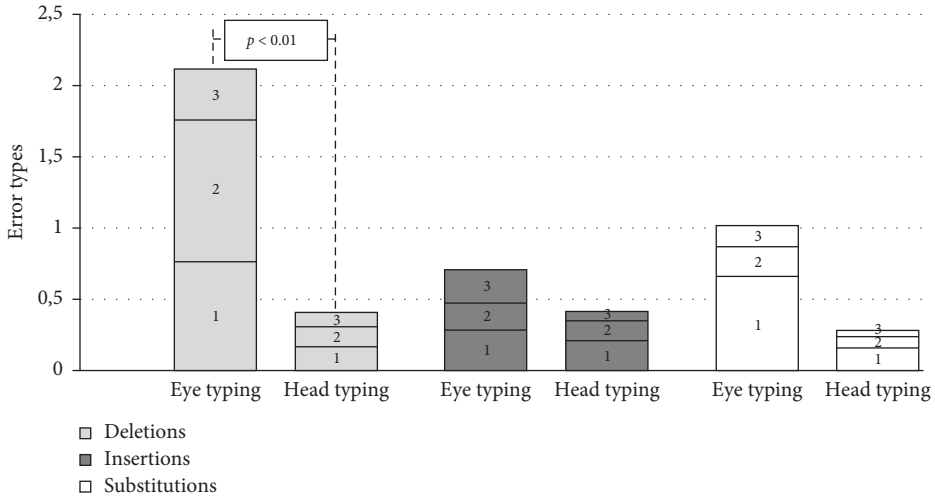


Figure 4: Errors of text entry (deletions, insertions, and substitutions) of eye typing and head typing during sessions 1, 2, and 3.

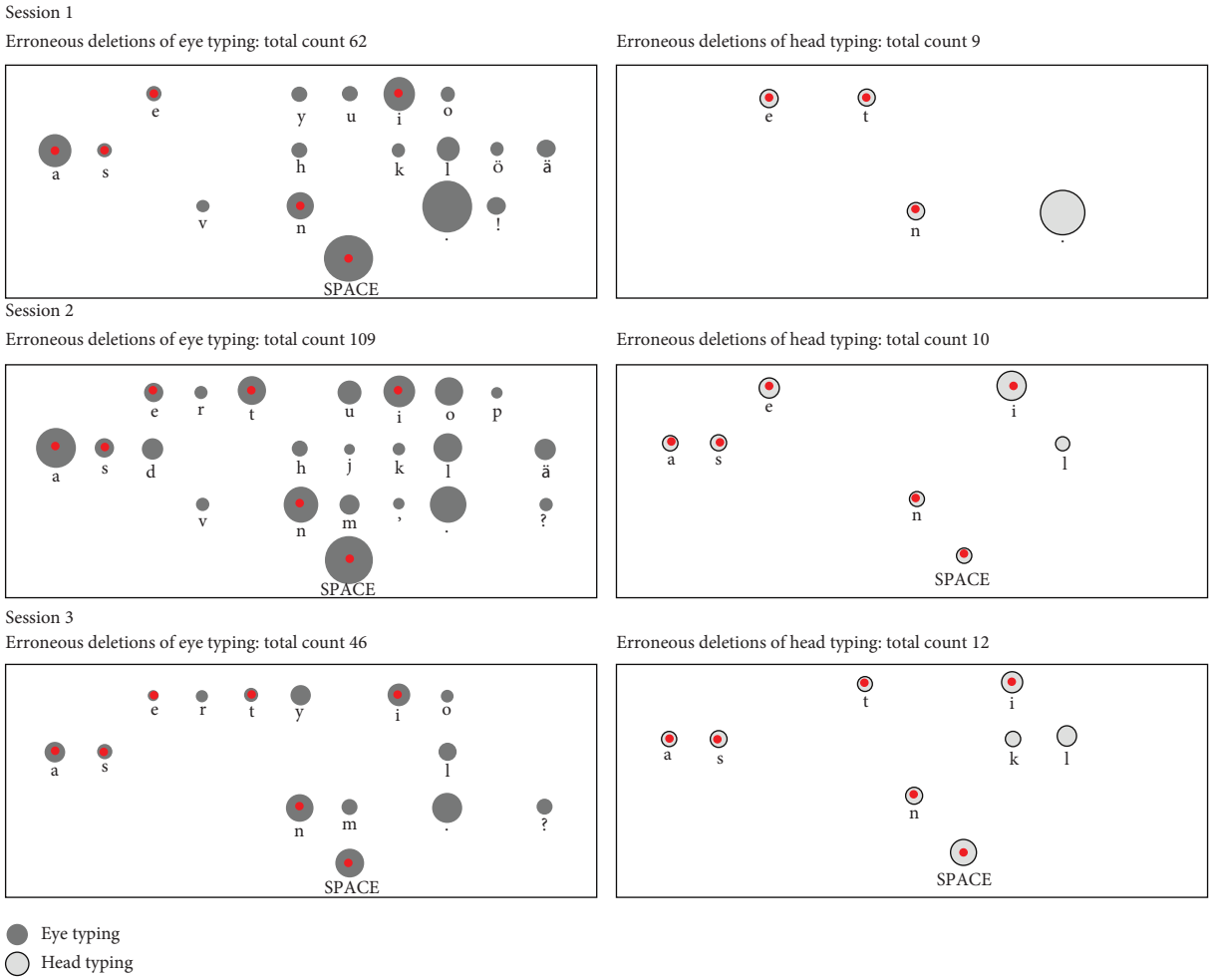


Figure 5: Deletion errors in sessions 1, 2, and 3 (total counts are depicted numerically). The most frequent characters "*a*," "*i*," SPACE, "*t*," "*n*," "*e*," and "*s*" are marked as red. Function keys are excluded from analysis.

tiredness, mostly because of the lack of blinking. Fifth, the participants' opinions about the use of head-based VBI in public spaces were unequal. Some participants mentioned that "it was bothering to open mouth because it looks funny," while others said that "mouth opening felt OK" and would consider using the technique in public spaces.

4.7. Case Study with a Person with Motor Disability. The results revealed that eye typing worked much better than head typing for our participant with motor disability (expert in eye typing). Altogether, the participant typed 20 ± 3.8 phrases (SD = 5.9) by gaze with an average speed of 9.1 ± 0.9 WPM (SD = 1.5). The speed of head typing was 2.3 ± 0.4 WPM (SD = 0.6) that resulted into 5.3 ± 0.9 phrases (SD = 1.5). Error rates of eye typing and head typing were $0.1 \pm 0.05\%$ (SD = 0.1) and $1.8 \pm 1\%$ (SD = 1.8), correspondingly. Notably, the participant was able to output error-free text in two eye-typing sessions and one head-typing session. Consistent with the quantitative results, the subjective evaluations of this participant in all categories were all positive for eye typing and negative for head typing at the end of the experiment.

We interviewed the participant regarding her expected use of the VBIs for text entry. In general, the participant enjoyed the fast speed of eye typing, especially when the calibration of the eye tracker was nearly ideal. It was inconvenient for the participant to type text when the calibration was imperfect. The participant further emphasized tiredness of the neck during head typing. Notably, this participant preferred to rotate her head (and did not move the torso at all) while steering the pointer during head typing. The participant further mentioned that head typing could feel better if technology worked more robustly (the face tracker lost the participant's face, and it was difficult to point at the bottom corners of the keyboard using the head). The participant preferred typing text with gaze-based VBI (or use it as an additional modality for other means of text entry), even if all technical problems were solved for head-based VBI. Therefore, the only anticipated usage of head-based VBI for this participant was in situations where the eye tracker's calibration was not sufficient to support accurate pointing at the keys of the keyboard. Regarding the use of head movements and mouth-opening gestures in public spaces, the participant felt that it would be acceptable for her to use both.

Figure 6: Erroneous selections in sessions 1, 2, and 3 (total counts are depicted numerically). The most frequent characters "*a*," "*i*," SPACE, "*t*," "*n*," "*e*," and "*s*" are marked as red. Function keys are excluded from analysis.

5. Discussion

5.1. Correctness of Text Entry. As instructed, the participants typed the correct text with both VBIs, with error rates of less than 1% (eye typing) and 0.5% (head typing) in the last session (Figure 3). The subjective evaluations reflected this fact, as the average scores for the perceived accuracy of text entry were positive for both interfaces (Figure 10). The short length of corrective actions for both interfaces indicates the participants put effort into frequent verification of the transcribed text, noticing and fixing erroneous characters immediately. However, head typing required significantly fewer corrective actions than gaze typing, as was also observed in the earlier studies [27, 35]. Importantly, head-based VBI supported error-free performance for many novices right from the beginning (Figure 3). Such a small variance in the error rate implies that no special learning is required to achieve high typing accuracy with head-based VBI.

There are two plausible reasons for eye typing being less accurate than head typing: (1) inherent inaccuracies of gaze pointing [68] and (2) possible limitations of cognitive and visual processing during eye typing. Typing with gaze requires controlled and steady usage of the eyes for the typing task itself; the typist needs to use gaze to guide the pointer to a designated place on a computer screen and hold it there until the dwell accumulation algorithm activates a key. Therefore, as noted earlier, other activities that require visual attention, such as locating the right key in the layout, verifying the typed character, and rereading text, serve as distractors in the typing process, thus contributing to erroneous selections [65].

Practitioners can use the results shown in Figure 9 to approximate the error correction cost for head typing and eye typing. As shown in the figure, with low error rates, the error correction cost is approximately the same for both interfaces. However, with an increase in the error rate, the cost of error correction increases for head typing, which may

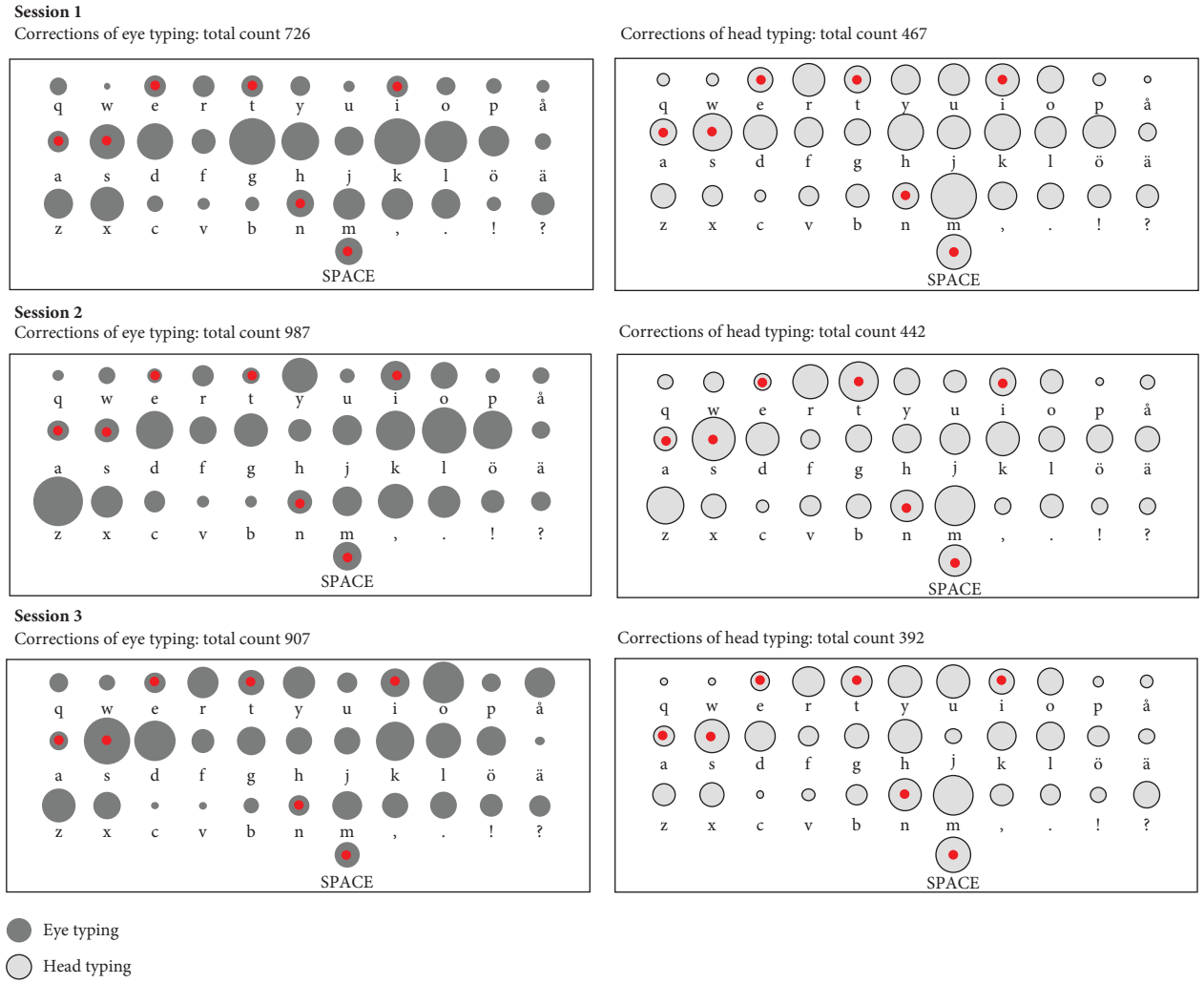


FIGURE 7: Corrective actions in sessions 1, 2, and 3 (total counts are depicted numerically). The most frequent characters "*a*," "*i*," SPACE, "*t*," "*n*," "*e*," and "*s*" are marked as red. Function keys are excluded from analysis.

be explained by the need to perform large and slow gross movements of the head (and possibly the torso) during the error correction process. In our study, head typing had an error rate that was approximately twice as small as that of eye typing and therefore had a smaller error correction cost at the end.

5.2. Spatial Error Analysis

5.2.1. Eye Typing. The clusters of uncorrected errors were quite similar to those reported by Räihä and Ovaska [19] where the same eye tracker, phrase corpus, and keyboard with a similar layout and larger keys were used (e.g., the key "*s*" was also often hit instead of "*a*" and vice versa). They suggested that uncorrected errors of eye typing in many cases resulted from the difficulty of "focusing" on the right key (i.e., inherent inaccuracies of gaze pointing). We hypothesize that exact spatial location of eye-typing errors may partly be hardware-dependent (and thus inherent to the eye tracker used in both studies) and partly own to the layout peculiarities where pairs of frequently used letters are located in close vicinity to each other (such as "*a*" and "*s*").

The participants initially made more frequent corrections when trying to select keys in the middle row than with keys located in the other two rows. The pointing-sensitive areas of the middle keys were smaller and had more neighboring keys than those located on the periphery of the keyboard (Figure 7, upper-left). After practice, the participants started making fewer errors in the middle of the layout (Figure 7, bottom-left). We hypothesize that novices developed strategies for dealing with inaccuracies in gaze pointer control, as experienced typists do [28]. However, they began to make more frequent and lengthy corrections on the periphery of the keyboard. Considering that the most prominent errors that penetrated the final text were also primarily localized on the periphery of the keyboard (Figures 5 and 6), we hypothesize that peripheral locations are difficult to inspect by gaze (i.e., after typing a character, the gaze immediately shifts away from that key in searching for the next key); therefore, the participants could simply

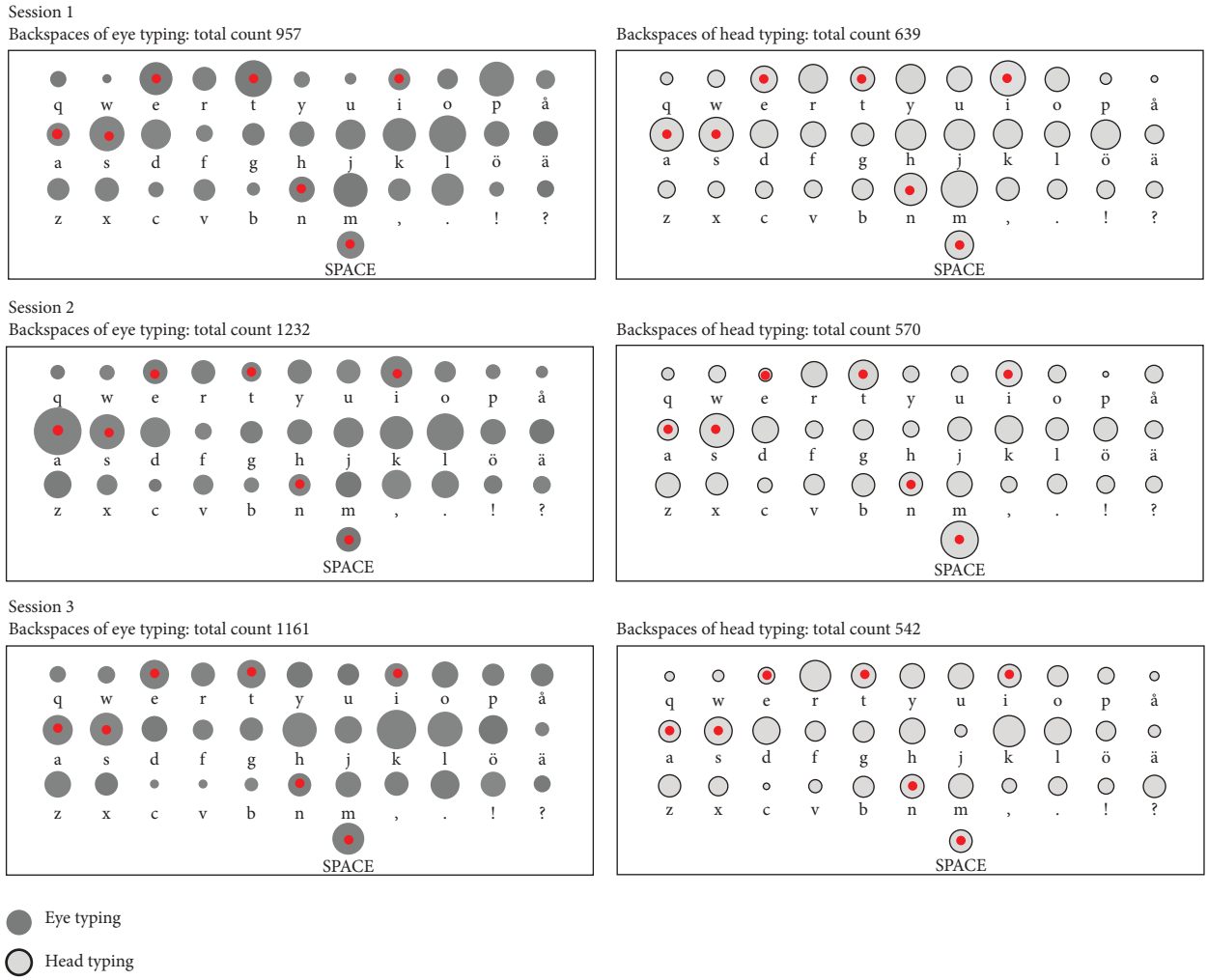


FIGURE 8: Average lengths of corrective actions in sessions 1, 2 and 3. The most frequent characters "*a*," "*i*," SPACE, "*t*," "*n*," "*e*" and "*s*" are marked as red. Function keys are excluded from analysis.

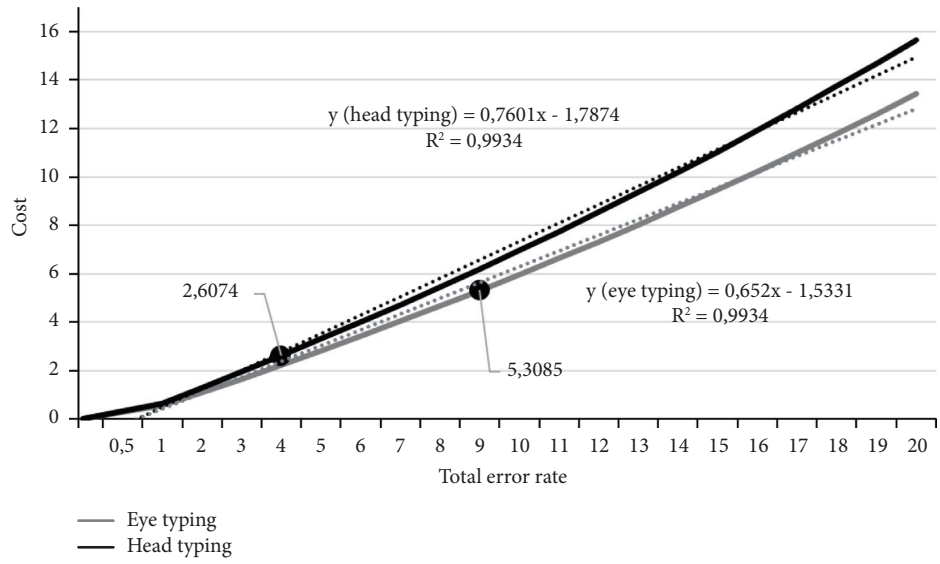


FIGURE 9: The increase in error correction cost with an increase in probability to commit an error. Data points define the error correction cost for eye typing (5.3 s) and head typing (2.6 s) in this study.

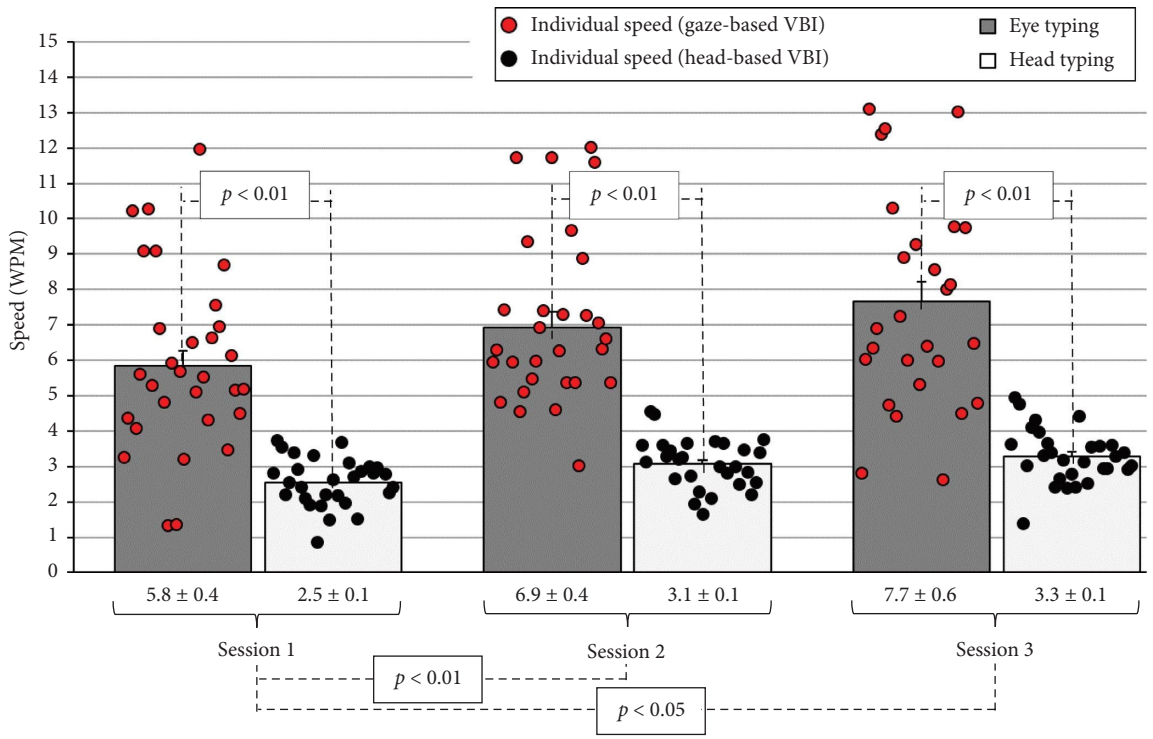


FIGURE 10: Columns show speed of text entry (error bars define 1 SEM from the mean values) during sessions 1–3. Scatter plots illustrate speed of individual typists.

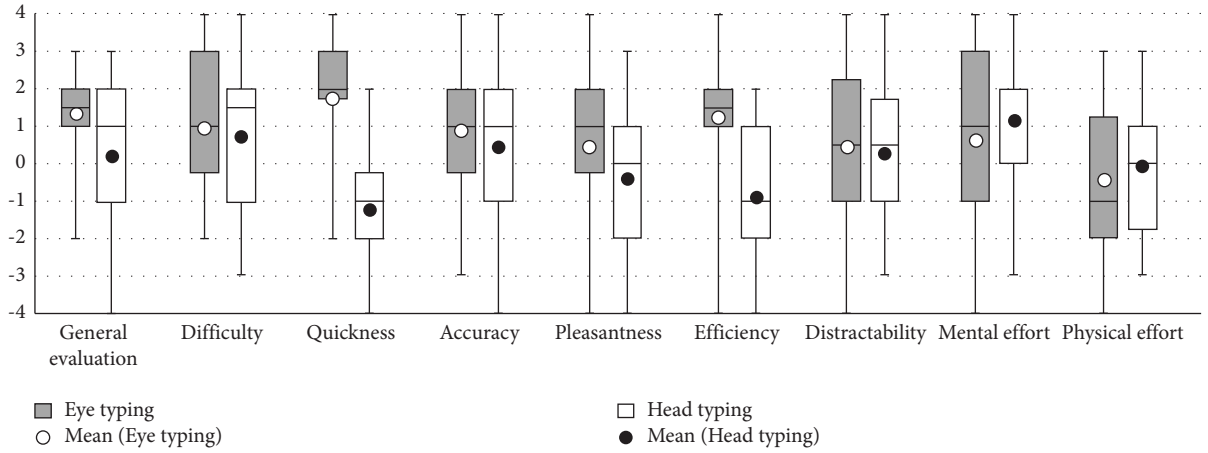


FIGURE 11: Final bipolar ratings of eye typing and head typing after 45 minutes of text entry.

overlook those errors that were not located in their immediate focus of visual attention (i.e., the central part of the keyboard).

5.2.2. Head Typing. In head typing, the eyes are free for visual inspection and verification, which may explain the generally smaller number of mistakes and corrective actions steadily decreasing towards the end of the experiment (Figures 5 and 6). We are unaware of other studies in which we could compare our results with the spatial distribution of uncorrected errors using this technique. There appears to be no correspondence between character deletions and erroneous selections for head typing as observed for eye typing. The figures reveal that the uncorrected errors in head typing were also located on the periphery of the keyboard, primarily in the upper part. We hypothesize that moving the head pointer to the top row might be more difficult for novices to perform than other movements. These results confirm the earlier consideration that selecting keys from extreme locations on the vertical axis is more difficult than selecting keys from extreme locations on the horizontal axis [18, 27]. After practice, errors and corrective actions became less frequent, suggesting that the participants learned to move in optimal ways to enter the text.

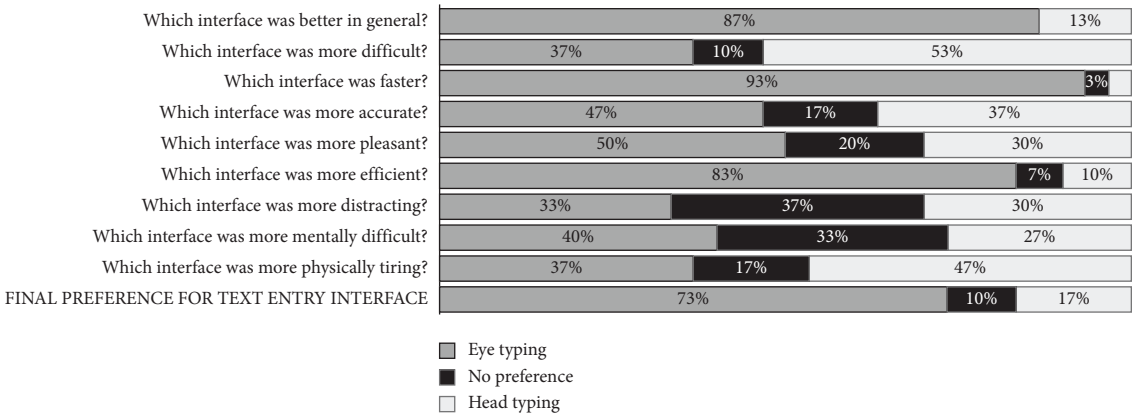


Figure 12: Pairwise preferences about text entry VBIs after 45 minutes of typing.

5.3. Speed of Text Entry. As Figure 9 shows, an average eye typist is expected to perform faster than the quickest head typist under the given conditions. Overall, 73% of the participants preferred gaze-based VBI for text entry (Figure 11). Final interviews revealed that many participants made their final preference in favor of gaze-based VBI based on its high speed of text production, as the ability to type fast appears to be a highly desirable property of the text entry interface (Figures 10 and 11). However, some participants (17%) typed faster using their heads and therefore preferred head-based VBI.

5.3.1. Speed versus Accuracy. The superiority of the eye-typing speed compared to head typing was reported earlier for static unambiguous keyboards that prohibit error correction [27]. In the current study, we hypothesized that the error correction demand may negatively affect eye-typing speed; the participants would naturally prefer typing slowly, ensuring that no errors occur in the final text. The results showed, however, that the participants still tended to increase their typing speed, presumably at the expense of the resulting text quality.

Earlier work [35] also reported that irrespective of test instructions, their participants were biased towards the speed of typing rather than its correctness (the instructions were to type as fast and as accurate as possible). The authors also mentioned that their participants did not invest greater effort in correcting errors caused by gaze input compared to head (or hand) input. In this respect, we note that our able-bodied participants, all of whom were experienced writers of electronic text, may have had a mindset of easy error correction that they could perform afterwards. This mindset and typing behavior could change if, for example, the experimental conditions did not allow the participants to proceed with the next phrase until the current phrase was written without a single mistake.

5.3.2. Eye-Typing Speed. As shown in Figure 9, there is a steady increase in the average speed of eye typing up to 8 WPM in the last session, which is lower than reported 10–18 WPM after 45 minutes of text entry [52]. Only the four best eye typists achieved speeds greater than 12 WPM during the last session. This is likely because our participants did not receive intensive training every day but rather obtained an experience of casual and infrequent text typing. Instruction with stress on text correctness was another factor that perhaps limited eye-typing speed.

Similar to an earlier study [27], eye typing resulted in a greater variation in text entry speed among the participants (from 3 to 13 WPM) than head typing (from 1 to 5 WPM), especially at the end of the experiment (Figure 9). In eye typing, the difference occurred because many participants decreased their dwell time, while a few increased it. Nevertheless, both categories of the participants are predicted to improve their performance in case of extensive and prolonged practice of eye typing, as it was observed in the studies with comparable user groups [16, 52].

5.3.3. Head-Typing Speed. Head-typing speed barely changed throughout typing practice and remained at approximately 3 WPM during all sessions, similar to a previous study with similar experimental conditions by Gizatdinova et al. [27], but slower than 7.8 WPM reported by Shin et al. [26] for their interface that combined head pointing with a mouth-opening gesture. The speed of the participant with a motor disability was approximately 2 WPM, which is comparable to the speed of users with disabilities with static keyboards [24, 26].

The between-user variability in the speed of head typing was much smaller than that of eye typing (Figure 9). None of the head typists distinctly outperformed others. This may be explained by the fact that in contrast to gaze-based VBI, head-based VBI did not offer the possibility of adjusting the typing parameters according to the preferences of the typist. The mouth-opening gesture had a fixed duration and required relatively wide mouth opening. Several participants mentioned that they could type quicker if the mouth-opening gesture recognition worked faster. Indeed, the speed of head typing with a key press used for selection previously was reported as 4.4 WPM [18].

Table 4: Characteristics of gaze- and head-based VBIs for hands-free text entry.

Eye typing	Head typing
Pros	Pros
Fast speed of text production (efficiency)	Possibility to freely observe text and keyboard layout
Easiness-of-use/less physical effort if large (1.5–2°) keys are used (efficiency)	Small number of errors (effectiveness) and small number of error corrections (efficiency)
	Does not require extensive learning for error-free performance
Text entry is possible even for severely disabled users	Can be used with small (0.4°) keys [28]
	Lower error correction cost (when the error rate is small)
Cons	Cons
Gaze is tied to the typing task itself	Slow speed of text entry
Requires large keys (and a keyboard takes large screen space)	Requires that control over head (and may be even torso) motion is preserved
Eye fatigue in case of small keys [28]	Physical fatigue
Requires learning for error-free performance	
Higher error correction cost (when the error rate is small)	

5.4. Applicability of VBIs for Hands-Free Text Entry

5.4.1. Gaze versus Head. The interfaces demonstrated advantages and limitations under the test conditions (Table 4). Head typing in general satisfies the required minimum rate of interactive conversation (defined as 3 WPM by Darragh and Witten [69]). However, as it was noted by De Vries et al. [70], even 5–7 WPM may not be functional in most work situations. When fast text entry is required, gaze-based VBI is undoubtedly preferred over head-based VBI. It can be argued that fast text entry is important, for instance, for text communication through messengers and phones. In these applications, shortening of words and typographical errors are commonplace, and gaze-based VBI can be widely utilized in fast messaging. Dwell-based eye typing can also be a preferred choice for monotonous text entry, such as transcribing phrases in this study. However, short dwells will lead to unintentional activation of the interface if a user switches to actions that require active visual search, such as filling web forms, emailing, or navigating menus.

Our results indicate that head-based VBI, despite its slow speed, has the potential to be useful in both typing-only and text-editing applications, owing to its two main advantages. First, pointer control is clearly separated from the focus of visual attention, which supports spotting incorrect key selections. Second, the stable control of the pointing action minimizes the risk of selecting the wrong key. Therefore, we anticipate that head-based VBIs may be beneficial, especially for text entry tasks that are often interrupted by other tasks, such as visual investigation. Facial gestures used in head-based VBI explicitly activate keyboard keys and eliminate unintentional errors. Moreover, different facial gestures can offer a rich set of activation commands that are not limited to 'select' command only. We also hypothesized that head-based VBIs are more suitable than gaze-based VBIs for interaction with computers that do not involve text typing. Dwell time is not a convenient selection technique when nonregular activation tasks are required, as eyes have to move constantly without long stops to avoid unintentional activations.

5.4.2. Limitations. Similar to other studies in the field of head typing (see Tables 1 and 2), text entry was tested with able-bodied participants. Therefore, the results reported in this study for head-based VBI primarily apply text entry for those users who have good control over their neck and face movements. This includes those individuals who can use head pointing or mouth sticks as an input method. It is difficult to predict whether the results can be generalized to a wider range of users with motor disabilities, especially those who have difficulties in controlling their neck and torso movements. A single user with motor disability (but preserved control over her neck and torso motion), an expert in eye tracking, showed much better performance in terms of text correctness and speed during eye typing than during head typing. The misalignment in typing errors between this participant and the others was presumably due to her high eye-typing skills.

5.4.3. Future Work. As neither of gaze- or head-based VBI alone hardly supports both fast and error-free text entry across a range of conditions, we suggest that users who preserve good control over their eyes, face, and neck (but not hands) can use both interfaces for writing electronic texts, switching between them whenever a text-typing scenario or condition changes. Even more, both VBIs could be merged into a single interface for text entry since it has been already shown that head movements can facilitate precise pointing, while gaze is used for fast (although sometimes inaccurate) cursor control [71–73].

Research on gaze-based VBIs can focus on the correctness of text production by developing aids that ease cognitive and visual information processing (e.g., [74]. Concerning head-based VBIs, more research is required on how the synchronization and optimization of the head (and perhaps the torso) movements are performed in particular typing tasks or key layouts. The newly introduced metric of corrective actions provides insights into error correction behavior and can help drive the development of optimal key layouts for head-based VBIs.

Regarding computer vision methods, the use of face detectors that build 3D head models or head rotation trackers, such as EyeTwig (https://www.eyetwig.com, accessed in March 2023), Enable Viacam (https://eviacam.crea-si.com, accessed in March 2023), or CameraMouse (https://www.cameramouse.org, accessed in March 2023), would allow the replacement of torso movements with head rotations, making pointing notably easier and, therefore, increasing the number of users living with motor disabilities who could use this VBI efficiently. The pointing speed can further be improved if the pointer-controlling algorithm discriminates between the speeds of head movements in the same manner as it is implemented for mouse cursor control (https://kinesicmouse.xcessity.at, accessed in March 2023) (e.g., [75]). The final interviews revealed several possible improvements of the mouth-opening gesture detector. Thus, the users wished to obtain an option of speeding up the gesture for making fast key activations by mouth opening. It would be interesting to optimize the gesture detector and perform a user study that would compare head pointing coupled with adjustable dwell versus head pointing coupled with an adjustable mouth-opening gesture.

It is noteworthy that some authors have implemented head typing using techniques other than camera-based head/face analysis. Thus, head motion for text entry has been computed not from video input, but using inertial sensors of VR/AR head-mounted displays (HMDs) by Yu et al. [76] and Xu et al. [55, 58]. The results are promising for both static and dynamic layouts, with 6–19 WPM recorded for novices (24 WPM for experienced users), which indicates potential for speed improvement in camera-based head typing when fast and robust video processing methods are used.

6. Conclusions

In this study, we empirically and systematically investigated the ability of gaze- and head-based VBIs to support error-free text entry. We proposed a new text entry metric, called

corrective actions per character (CAPC), which measures the efficiency of text production and serves as an indicator of error correction strategies of text typists. We analyzed the errors and error corrections relative to the spatial layout of the virtual keyboard and estimated the error correction costs for both interfaces. The results showed that head-based VBI allowed typing of electronic text without mistakes, which was notably better than gaze-based VBI. Most participants wrote error-free text with head-based VBI in the first session, infrequently making mistakes and taking corrective actions. Gaze-based VBI was more prone to errors in text entry and required multiple corrective actions but supported faster speed of text production compared to head-based VBI. Subjective results reflected these findings. In future development of VBIs for hands-free text entry, we suggest combining both gaze and head modalities to improve typing performance and user satisfaction.

Data Availability

The data used to support the findings of this study are available from the first author Dr. Julia Kuosmanen (publishes as Yulia Gizatdinova) at julia.kuosmanen@tuni.fi and julia.f.kuosmanen@gmail.com upon request.

Consent

Informed consent was obtained via the Open Select publishing program.

Disclosure

The funding sources defined in Acknowledgments had no involvement in the study design; collection, analysis and interpretation of data; writing of the report; and decision to submit the article for publication.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We thank the Post Doc Pool/Jenny and Antti Wihuri Foundation, Academy of Finland (grant 308929), Tampere Universities, Tampere Institute for Advanced Study, and University of California, Santa Barbara for support. We thank James Gribble and Editage (https://www.editage.com) for English language editing and our study participants for their valuable participation. Open Access funding was enabled and organized by FinELib 2023.

References

[1] M. Porta, "Vision-based user interfaces: methods and applications," *International Journal of Human-Computer Studies*, vol. 57, no. 1, pp. 27–73, 2002.

[2] M. Turk and M. Kölsch, "Perceptual interfaces," *Emerging Topics in Computer Vision*, Prentice Hall, Hoboken, NJ, USA, 2004.

[3] O. Tuisku, V. Surakka, T. Vanhala, V. Rantanen, and J. Lekkala, "Wireless Face Interface: using voluntary gaze direction and facial muscle activations for human–computer interaction," *Interacting with Computers*, vol. 24, no. 1, pp. 1–9, 2012.

[4] A. Sears, M. Young, and J. Feng, "Physical disabilities and computing technologies: an analysis of impairments," *Human-Computer Interaction: Designing for Diverse Users and Domains*, pp. 87–110, 2009.

[5] E. LoPresti, D. M. Brienza, J. Angelo, L. Gilbertson, and J. Sakai, "Neck range of motion and use of computer head controls," in *Proceedings of the fourth international ACM conference on Assistive technologies (Assets '00)*, pp. 121–128, Association for Computing Machinery, New York, NY, USA, July 2000.

[6] W. Feng, M. Sameki, and M. Betke, "Exploration of assistive technologies used by people with quadriplegia caused by degenerative neurological diseases," *International Journal of Human-Computer Interaction*, vol. 34, no. 9, pp. 834–844, 2018.

[7] H. H. Koester and S. Arthanat, "Text entry rate of access interfaces used by people with physical disabilities: a systematic review," *Assistive Technology*, vol. 30, no. 3, pp. 151–163, 2018.

[8] R. C. Simpson, *Computer Access for People with Disabilities: A Human Factors Approach*, CRC Press, FL, USA, 2013.

[9] P. Majaranta and K.-J. Räihä, "Twenty years of eye typing: systems and design issues," *Proceedings of the symposium on Eye tracking research and applications-ETRA '02*, vol. 15, 2002.

[10] K. Grauman, M. Betke, J. Lombardi, J. Gips, and G. R. Bradski, "Communication via eye blinks and eyebrow raises: video-based human-computer interfaces," *Universal Access in the Information Society*, vol. 2, no. 4, pp. 359–373, 2003.

[11] M. H. Urbina and A. Huckauf, "Dwell-time free eye typing approaches," in *Proceedings of the 3rd Conference on Communication by Gaze Interaction (COGAIN 2007)*, pp. 65–70, Leicester, UK, September 2007.

[12] O. Tuisku, V. Surakka, V. Rantanen, T. Vanhala, and J. Lekkala, "Text entry by gazing and smiling," *Advances in Human-Computer Interaction*, vol. 2013, Article ID 218084, 13 pages, 2013.

[13] H. Venesvirta, O. Špakov, Y. Gizatdinova et al., "Smile to save it-facial expressions for lifelogging," in *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*, pp. 441–448, Stuttgart Germany, November 2017.

[14] M. Yildiz and H. Ö. Ülkütaş, "A new PC-based text entry system based on EOG coding," *Advances in Human-Computer Interaction*, vol. 2018, Article ID 8528176, 8 pages, 2018.

[15] P. Majaranta and K.-J. Räihä, "Text entry by gaze: utilizing eye tracking," in *Text Entry Systems*, pp. 175–187, University of Tampere, Tampere, Finland, 2007.

[16] K. J. Räihä, "Life in the fast lane: effect of language and calibration accuracy on the speed of text entry by gaze," in *Proceedings of the Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference*, Bamberg, Germany, September 2015.

[17] C. Kumar, R. Hedeshy, I. MacKenzie, and S. Staab, "TAGSwipe: touch assisted gaze swipe for text entry," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu, HI, USA, April 2020.

[18] Y. Gizatdinova, O. Spakov, and V. Surakka, "Face typing: vision-based perceptual interface for hands-free text entry with a scrollable virtual keyboard," in *Proceedings of the 2012*

IEEE Workshop on the Applications of Computer Vision (WACV), Breckenridge, CO, USA, January 2012.

[19] K.-J. Räihä and S. Ovaska, "An exploratory study of eye typing fundamentals: dwell time, text entry rate, errors, and workload," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 3001–3010, Austin TX USA, May 2012.

[20] M. Betke, J. Gips, and P. Fleming, "The Camera Mouse: visual tracking of body features to provide computer access for people with severe disabilities," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 1, pp. 1–10, 2002.

[21] G. J. Capilouto, "Movement variability and speed of performance using a head-operated device and expanded membrane cursor keys," *Lecture Notes in Computer Science*, vol. 3118, pp. 820–826, 2004.

[22] D. K. Anson, M. Glodek, R. M. Peiffer, C. G. Rubino, and P. T. Schwartz, "Long-term speed and accuracy of Morse code vs. head-pointer interface for text generation," in *Proceedings of the RESNA 27th International Annual Conference*, Orlando, Florida, June 2023.

[23] M. C. Su, S. Y. Su, and G. D. Chen, "A low-cost vision-based human-computer interface for people with severe disabilities," *Biomedical Engineering: Applications, Basis, and Communications*, vol. 17, no. 6, pp. 284–292, 2005.

[24] R. Kjeldsen, "Improvements in vision-based pointer control," in *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 189–196, Portland Oregon USA, October 2006.

[25] R. Kjeldsen, "An on-screen keyboard for users with poor pointer control," *Lecture Notes in Computer Science*, vol. 4556, no. 3, pp. 339–348, 2007.

[26] Y. Shin, J. S. Ju, and E. Y. Kim, "Welfare interface implementation using multiple facial features tracking for the disabled people," *Pattern Recognition Letters*, vol. 29, no. 13, pp. 1784–1796, 2008.

[27] Y. Gizatdinova, O. Špakov, and V. Surakka, "Comparison of video-based pointing and selection techniques for hands-free text entry," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 132–139, Capri Island Italy, May 2012.

[28] Y. Gizatdinova, O. Špakov, O. Tuisku, M. Turk, and V. Surakka, "Gaze and head pointing for hands-free text entry: applicability to ultra-small virtual keyboards," in *Proceedings of the 2018 ACM Symposium on Eye Tracking Research and Applications*, pp. 1–9, Warsaw Poland, June 2018.

[29] D. Sawicki and P. Kowalczyk, "Head movement based interaction in mobility," *International Journal of Human-Computer Interaction*, vol. 34, no. 7, pp. 653–665, 2018.

[30] A. Nowosielski, "Text entry by rotary head movements," *Image Processing and Communications Challenges 10*, vol. 10, pp. 71–78, 2018.

[31] W. Feng, J. Zou, A. Kurauchi, C. H. Morimoto, and M. Betke, "HGaze typing: head-gesture assisted gaze typing," in *Proceedings of the Eye Tracking Research and Applications Symposium (ETRA)*, Germany, May 2021.

[32] R. L. Cloud, M. Betke, and J. Gips, "Experiments with a camera-based human-computer interface system," in *Proceedings of the 7th ERCIM Workshop User Interfaces for All*, pp. 103–110, UI4ALL, Paris, France, October 2002.

[33] G. C. D. Silva, M. J. Lyons, S. Kawato, and N. Tetsutani, "Human factors evaluation of a vision-based facial gesture interface," in *Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition Workshop*, Madison, WI, USA, June 2003.

[34] M. Lyons, C.-H. Chan, and N. Tetsutani, "MouthType: text entry by hand and mouth," *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, vol. 1383, 2004.

[35] J. Hansen, K. Tørning, A. Johansen, K. Itoh, and H. Aoki, "Gaze typing compared with input by head and hand," *Eye Tracking Research and Application: Proceedings of the 2004 Symposium on Eye Tracking Research and Applications*, vol. 22, 2004.

[36] E. Perini, S. Soria, A. Prati, and R. Cucchiara, "FaceMouse: a human-computer interface for tetraplegic people," *Computer Vision in Human-Computer Interaction*, vol. 3979, pp. 99–108, 2006.

[37] M. C. Su, C. Yeh, S. Lin, P. Wang, and S. Hou, "An implementation of an eye-blink-based communication aid for people with severe disabilities," in *Proceedings of the 2008 International Conference on Audio, Language and Image Processing*, pp. 351–356, Shanghai, China, July 2008.

[38] B. Ashtiani and I. MacKenzie, "BlinkWrite2: an improved text entry method using eye blinks," in *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications-ETRA '10*, pp. 339–345, Austin Texas, March 2010.

[39] L. R. Sapaico and M. Sato, "Analysis of vision-based Text Entry using morse code generated by tongue gestures," in *Proceedings of the 2011 4th International Conference on Human System Interactions, HSI 2011*, Yokohama, Japan, May 2011.

[40] A. Królak and P. Strumillo, "Eye-blink detection system for human–computer interaction," *Universal Access in the Information Society*, vol. 11, no. 4, pp. 409–419, 2012.

[41] R. Das and B. ShivaKumar, "Headspeak: morse code based head gesture to speech conversion using intel Realsense™ technology," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 2866–2874, 2019.

[42] A. Nowosielski and P. Forczmański, "Touchless typing with head movements captured in thermal spectrum," *Pattern Analysis and Applications*, vol. 22, no. 3, pp. 841–855, 2019.

[43] P. Kar, K. Mishra, S. Ghosh, S. Chakraborty, and S. Chattopadhyay, "Exploratory analysis of nose-gesture for smartphone aided typing for users with clinical conditions," in *Proceedings of the 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events (PerCom Workshops)*, pp. 380–383, Kassel, Germany, March 2021.

[44] M. O. Taş and H. S. Yavuz, "A human-computer interaction system based on eye, eyebrow and head movements," *Pamukkale University Journal of Engineering Sciences*, vol. 28, no. 5, pp. 632–642, 2022.

[45] O. Poláček, A. Sporka, and P. Slavík, "Text input for motor-impaired people," *Universal Access in the Information Society*, vol. 16, pp. 51–72, 2017.

[46] A. S. Arif and W. Stuerzlinger, "Predicting the cost of error correction in character-based text entry technologies," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, vol. 1, pp. 5–14, 2010.

[47] P. Majaranta, N. Majaranta, G. Daunys, and O. Špakov, "Text editing by gaze: static vs. dynamic menus," in *Proceedings of*

the 5th Conference on Communication by Gaze Interaction (COGAIN), Lyngby, Denmark, May 2009.

[48] R. J. Jagacinski and D. L. Monk, "Fitts' law in two dimensions with hand and head movements movements," *Journal of Motor Behavior*, vol. 17, no. 1, pp. 77–95, 1985.

[49] R. G. Radwin, G. C. Vanderheiden, and M.-L. Lin, "A method for evaluating head-controlled computer input devices using fitts' law," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 32, no. 4, pp. 423–438, 1990.

[50] S. Zhai, J. Kong, and X. Ren, "Speed–accuracy tradeoff in Fitts' law tasks—on the equivalency of actual and nominal pointing precision," *International Journal of Human-Computer Studies*, vol. 61, no. 6, pp. 823–856, 2004.

[51] P. Kristensson and K. Vertanen, "The potential of dwell-free eye-typing for fast assistive gaze communication," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, p. 241, Santa Barbara CA, USA, March 2012.

[52] P. Majaranta, U.-K. Ahola, and O. Špakov, "Fast gaze typing with an adjustable dwell time," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 357–360, Boston MA USA, April 2009.

[53] O. Tuisku, P. Majaranta, P. Isokoski, and K.-J. Räihä, "Now Dasher! Dash away: longitudinal study of fast text entry by Eye Gaze," in *Proceedings of the 2008 Symposium on Eye Tracking Research and Applications*, Savannah Georgia, March 2008.

[54] M. Ilves, Y. Gizatdinova, V. Surakka, and E. Vankka, "Head movement and facial expressions as game input," *Entertainment Computing*, vol. 5, no. 3, pp. 147–156, 2014.

[55] Y. Gizatdinova and V. Surakka, "Automatic edge-based localization of facial features from images with complex facial expressions," *Pattern Recognition Letters*, vol. 31, no. 15, pp. 2436–2446, 2010.

[56] W. K. Purves, D. E. Sadava, G. H. Orians, and H. C. Heller, *Life, the Science of Biology*, W. H. Freeman, Madison Ave, NY, USA, 2003.

[57] I. S. MacKenzie and R. W. Soukoreff, "Phrase sets for evaluating text entry techniques," in *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*, pp. 754-755, ACM, New York, NY, USA, 2003.

[58] I. S. MacKenzie and W. R. Soukoreff, "Phrase sets for evaluating text entry techniques," in *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*, pp. 754-755, Association for Computing Machinery, New York, NY, USA, 2003.

[59] P. Isokoski and T. Linden, "Effect of foreign language on text transcription performance: Finns writing English," *Proceedings of the third Nordic conference on Human-computer interaction*, vol. 82, pp. 109–112, 2004.

[60] R. Raissi, E. Dimara, J. H. Berry, W. D. Gray, and G. Bailly, "Retroactive transfer phenomena in alternating user interfaces," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, Association for Computing Machinery, New York, NY, USA, April 2020.

[61] D. Miniotas, O. Špakov, and I. S. MacKenzie, "Eye gaze interaction with expanding targets," in *Proceedings of the CHI'04 Extended Abstracts on Human Factors in Computing Systems*, pp. 1255–1258, Vienna Austria, April 2004.

[62] P. Majaranta, I. S. MacKenzie, A. Aula, and K.-J. Räihä, "Effects of feedback and dwell time on eye typing speed and accuracy," *Universal Access in the Information Society*, vol. 5, no. 2, pp. 199–208, 2006.

[63] R. W. Soukoreff and I. S. MacKenzie, "Measuring errors in text entry tasks: an application of the Levenshtein string distance statistic," in *Proceedings of the CHI '01 Extended Abstracts on Human Factors in Computing Systems*, pp. 319-320, Seattle, Washington, DC, USA, March 2001.

[64] M. Zhang and J. Wobbrock, "Beyond the input stream: making text entry evaluations more flexible with transcription sequences," in *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 831–842, New Orleans LA USA, October 2019.

[65] I. S. Mackenzie, "KSPC (Keystrokes per character) as a characteristic of text entry techniques," *Lecture Notes in Computer Science*, vol. 2411, pp. 195–210, 2002.

[66] F. E. Grubbs, "Sample criteria for testing outlying observations," *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 27–58, 1950.

[67] J. Cohen, "Statistical power analysis for the behavioral sciences," *L. Erlbaum Associates*, Routledge, NY, USA, 1988.

[68] R. J. K. Jacob, "Hot topics-eye-gaze computer interfaces: what you look at is what you get," *Computer*, vol. 26, no. 7, pp. 65-66, 1993.

[69] J. J. Darragh and I. H. Witten, *The Reactive Keyboard*, Cambridge University Press, Cambridge, UK, 1992.

[70] R. C. De Vries, J. Deitz, and D. Anson, "A comparison of two computer access systems for functional text entry," *American Journal of Occupational Therapy*, vol. 52, no. 8, pp. 656–665, 1998.

[71] O. Špakov, P. Isokoski, and P. Majaranta, "Look and lean: accurate head-assisted eye pointing," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 35–42, Safety Harbor FL, USA, March 2014.

[72] A. Kurauchi, W. Feng, C. Morimoto, and M. Betke, "HMAGIC: head movement and gaze input cascaded pointing," in *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA 2015*, Corfu Greece, July 2015.

[73] L. Sidenmark, D. Mardanbegi, A. R. Gomez, C. Clarke, and H. Gellersen, "BimodalGaze: seamlessly refined pointing with gaze and filtered gestural head movement," in *ACM Symposium on Eye Tracking Research and Applications (ETRA '20 Full Papers)*, Association for Computing Machinery, New York, NY, USA, 2020.

[74] A. Diaz-Tula and C. H. Morimoto, "Augkey: increasing foveal throughput in eye typing with augmented keys," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3533–3544, San Jose CA, USA, May 2016.

[75] A. Nowosielski, "Minimal interaction touchless text input with head movements and stereo vision," *Computer Vision and Graphics*, Springer International Publishing, Berlin, Germany, pp. 233–243, 2016.

[76] C. Yu, Y. Gu, Z. Yang, X. Yi, H. Luo, and Y. Shi, "Tap, dwell or gesture?: exploring head-based text entry techniques for HMDs," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 4479–4488, Denver Colorado USA, May 2017.

[77] W. Xu, H.-N. Liang, A. He, and Z. Wang, "Pointing and selection methods for text entry in augmented reality head mounted displays," in *Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Beijing, China, October 2019.

[78] W. Xu, H.-N. Liang, Y. Zhao, T. Zhang, D. Yu, and D. Monteiro, "RingText: dwell-free and hands-free text entry for mobile head-mounted displays using head motions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1991–2001, 2019b.