Int. J. Human–Computer Studies 161 (2022) 102787

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

PressTapFlick: Exploring a gaze and foot-based multimodal approach to gaze typing

Vijay Rajanna *, Murat Russel, Jeffrey Zhao, Tracy Hammond

Department of Computer Science & Engineering, Sketch Recognition Lab, Texas A&M University, College Station, TX, USA

ARTICLE INFO

Keywords:
Gaze typing
Multimodal interaction
Foot-based interaction
Virtual keyboard
Optikey
3D printing
Microcontroller

ABSTRACT

Text entry is extremely difficult or sometimes impossible in the scenarios of situationally-induced impairments and disabilities, and for individuals with motor impairments (physical impairments and disabilities) by birth or due to an injury. As a remedy, many rely on gaze typing with dwell-based selection as it allows for hands-free text entry. However, dwell-based gaze typing could be limited by usability issues, reduced typing speed, high error rate, steep learning curve, and visual fatigue with prolonged usage. Addressing these issues is crucial for improving the usability and performance of gaze typing.

In our work, we present a dwell-free, multimodal approach to gaze typing where the gaze input is supplemented with a foot input modality. Our combined gaze and foot-based typing system comprises of an enhanced virtual QWERTY keyboard (VKB), and a footwear augmenting wearable device that provides the foot input. In this multi-modal setup, the user points her gaze at the desired character, and selects it with the foot input. We further investigated two approaches to foot-based selection, a foot gesture-based selection and a foot press-based selection, which are compared against the standard dwell-based selection.

We evaluated our gaze typing system through a comparative study involving three experiments (51 participants), where each experiment used one of the three target selection methods, and had 17 participants in it. In the first experiment the participants used dwell-based selection, second, foot gesture-based selection, and third, foot press-based selection for gaze typing. We found that with dwell-based selection the highest mean typing speed of 11.65 WPM (max 14.83 WPM) was achieved when using a dwell time of 400 ms. Similarly, among foot-based selection methods the highest mean typing speed of 14.98 WPM (max 18.18 WPM) was achieved with foot press-based selection. Furthermore, ANOVA tests revealed that the difference in the typing speeds between the three selection methods is significant, however, no significant difference was found in the error rate.

Overall, based on both typing performance and qualitative feedback the results suggest that gaze and foot-based typing is convenient, easy to learn, and addresses the usability issues associated with dwell-based typing. Furthermore, toe tapping is the most preferred foot gesture of all the four gestures (toe tapping, heel tapping, right flick and left flick) we used in the study. Also, we found that when using foot-based selection users quickly develop a rhythm in focusing at a character with gaze and selecting it with the foot, and this familiarity reduces the errors significantly. We believe, our findings would encourage further research in leveraging a supplemental foot input in gaze typing, or in general, would assist in the development of rich foot-based interactions.

1. Introduction

Text entry is one of the basic operations performed on a computer, and is achieved through various input modalities. While text entry through typing on a physical keyboard is primarily used, the keyboard-based text entry has limitations under certain circumstances. These circumstances can be classified into two groups: (1) situationally-induced impairments and disabilities (SIID) (Kane et al., 2008; Qian et al., 2013; Schildbach and Rukzio, 2010), and (2) physical impairments (Majaranta, 2011). In case of the SIIDs, a user's hands are assumed to be engaged in other tasks, and hence unavailable for typing on a physical keyboard. For example, a surgeon performing an operation, a musician playing music, a factory worker wearing thick gloves or with greasy hands, a person driving a vehicle, etc. Similarly, in the case

* Corresponding author.
E-mail address: vijay.drajanna@gmail.com (V. Rajanna).

https://doi.org/10.1016/j.ijhcs.2022.102787
Received 13 May 2021; Received in revised form 29 December 2021; Accepted 27 January 2022
Available online 31 January 2022

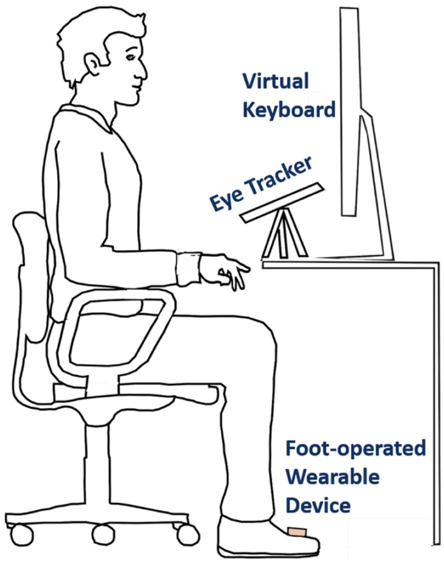


Fig. 1. Gaze and foot-based typing system: an eye tracker is placed in front of a monitor displaying the virtual keyboard, and the user is wearing a footwear augmented with the wearable device.

of physical impairments, either by birth or injury, a user's hands maybe unavailable, or the user may not have enough control over their hands (e.g., shaky hands) to type on a keyboard. In both of these scenarios, gaze typing plays a crucial role in assisting these individuals to enter text on a computer through their eye movements. Alternatively, speech to text input also serves as one of the viable solutions in the above scenarios when using a keyboard is not possible. However, speech to text input has various limitations because of its accuracy and applicability (Deng and Huang, 2004; Hauptmann, 1995). Accuracy of speech recognition is often impacted by noisy environments and accent of the user (Huang et al., 2004; Humphries et al., 1996; Kat and Fung, 1999), also speech to text can not be used to enter confidential information (passwords, unique IDs, etc.) in public places.

Text entry by gaze has gained significant focus because of its robustness, applicability, and the ability to customize the system to be appropriate for different kinds of impairments (accessible technology). Among all the gaze-assisted text entry methods available like gaze gestures (Porta and Turina, 2008; Ward et al., 2000; Wobbrock et al., 2008), gaze switches (Fejtová et al., 2006; Grauman et al., 2003; ten Kate et al., 1980), and so on, one method that has received maximum focus is "Gaze Typing" (Duchowski, 2007; Jacob, 1991; Majaranta, 2011). Gaze typing uses a virtual keyboard on a monitor, and to enter a character a user fixates her gaze on a specific character for a duration of time referred to as the dwell time (e.g., 500 ms). A constant fixation for the duration of dwell time, confirms the user's intent to select the target character (Majaranta, 2011). This method of gaze typing is referred to as dwell-based gaze typing, and duration of the dwell time varies between systems commonly from 400 ms to 1000 ms (Hansen et al., 2003; Majaranta and Räihä, 2007). While the majority of the gaze typing systems use dwell for selection, the other two approaches include dwell-free gaze typing and gaze typing with multimodal input. When using dwell-free gaze typing system the user gazes over the target characters but does not fixate on them (MacKenzie and Zhang, 2008). The system uses an internal dictionary to generate a possible set of words which are then presented to the user for selection with dwell time. Lastly, multimodal gaze typing systems still use gaze to point at the target character, however, the selection is triggered by a supplemental input modality (Rajanna, 2016a; Zhao et al., 2012). We will comment more on dwell-based, dwell-free, and multimodal gaze typing systems in the prior work section (2).

Existing dwell-based and dwell-free gaze typing systems have various limitations that affect gaze typing performance and usability, and these issues have been well discussed (Frey et al., 1990; Isokoski, 2000; Jacob, 1991; Majaranta and Räihä, 2002). To summarize the issues with dwell-based gaze typing, first, they place a high demand on the user's attention, and sometimes results in inadvertent selection of keys due to the MIDAS Touch (Jacob, 1991) issue. The issue of MIDAS Touch states that when eye position on the screen is used as a direct substitute for the mouse, wherever the user fixates during a visual search, the point gets activated. This unintentional or indiscriminate selection is inefficient, and leads to user frustration (Jacob, 1991). Second, text entry can be slow based on the dwell duration used, typical typing speeds are below 10 WPM (Frey et al., 1990; Majaranta and Räihä, 2002). Third, a single dwell time is not suitable for all users, hence it is hard to find an optimal dwell-time and this also contributes to a steep learning curve. If a shorter dwell time is used (150 to 400) to improve the typing speed, the user is constantly forced to perform a visual search for the target key without inadvertently fixating for too long before finding the correct target (Isokoski, 2000; Majaranta and Räihä, 2007). This results in more errors and a higher overproduction rate (Isokoski, 2000; Majaranta et al., 2009; Majaranta and Räihä, 2007). But, a longer dwell-time, though increases accuracy, reduces the typing speed, limits quicker users, and increases visual fatigue (Isokoski, 2000; Majaranta et al., 2009; Majaranta and Räihä, 2007). Lastly, some users simply can not focus at a point for a sufficiently long duration (Huckauf and Urbina, 2007; Urbina and Huckauf, 2010).

Similarly, dwell-free typing systems that use extra saccades for gaze typing only marginally increase the typing speed, with a major downside of increasing the keystrokes required per character (Majaranta, 2011). Dwell-free typing systems mainly rely on language modeling, and word and character prediction to support text entry. However, the systems that use word prediction induce cognitive and perceptual load on the user, and the learning curve is steep. The user constantly switches focus from the keyboard to scanning predicted list of words to see if the desired word is populated. Hence, though word prediction may reduces keystrokes per character (KSPC), the improvement in typing speed achieved is minimal, and in some cases worse than non prediction system (Koester and Levine, 1994; Majaranta, 2011). Also, the prediction system consumes precious screen real estate to populate the word list. Another limitation when using word prediction is the lack of an extensive library. For this reason, these systems will under-perform when typing unknown words like family names or local places. Hence, they are limited for practical use in free communication, but work well under constrained input conditions (Hansen et al., 2004a).

While there are limitations with both dwell-based and dwell free gaze typing systems, there have been significant works done to mitigate these limitations which we will discuss in the prior work section (2). When considering how crucial gaze typing is in the scenarios of SIIDs and physical impairments, it is essential to address the current limitations, and improve the gaze typing experience, usability, and performance. In this research, we present a dwell-free, multimodal, gaze typing system that uses a supplemental foot input. The foot input is achieved through a footware augmented with a wearable device, which communicates with the central system wirelessly. The rational for choosing a supplemental foot input are discussed in the design motivation section (4). Fig. 1 shows the pictorial depiction of the system, and further details are provided in system design and implementation section (5). We also present an enhanced virtual QWERTY keyboard where the key layout is customized to maximize the selection area for a few selected keys, and hence supports ease of interaction. Gaze and foot-based point-and-click interactions on a desktop have been previously explored (Hatscher et al., 2017; Klamka et al., 2015; Rajanna and Hammond, 2016). However, when considering gaze typing, except for a preliminary study by Rajanna (2016a), to our knowledge there exists no study that thoroughly studies gaze typing with a supplemental foot input. Hence, the core contribution of our work is to thoroughly investigate gaze typing using foot input - specifically through foot gestures and subtle foot press-based selection methods. In this regard, we also present the design and implementation details of small form factor wearables for foot input that were developed over multiple design iterations.

Through the system evaluation we wanted to determine the feasibility of a gaze and foot-based typing system, and compare its performance to existing gaze and dwell-based typing systems. The specific goals are discussed in research questions Section 3. Hence, we conducted a comparative user study involving three experiments with a total of 51 participants. Overall our results suggest that an efficient gaze typing system that addresses most of the usability issues can be achieved by incorporating a supplemental foot input modality. We also learned that foot-based selection at least matches, and likely improves, the gaze typing performance compared to dwell-based selection. Furthermore, the users appreciated the greater control over the interface with gaze and foot-based typing as inadvertent key selections were significantly reduced. We found that by dividing the responsibility, i.e., focusing on a key and its selection between two separate input modalities, helps to achieve a more usable and robust gaze typing system. Also, we learned from the user studies that the key to achieving a higher typing speed (WPM) on our system is the ability to synchronize focusing on a character, and selection with foot input.

2. Prior work

Research in gaze-assisted text entry dates back more than 20 years (Majaranta and Räihä, 2002). There have been various gaze typing methods developed and most of them use gaze as the only input modality, but some of them use a supplemental input modality along with gaze. In this section we will discuss the major classifications of gaze-assisted text entry methods, and also discuss how gaze is combined with supplemental input modality for point-and-click interactions.

2.1. Gaze and foot-based interaction

Studies investigating the usability and permanence of gaze and foot-based interactions are limited. Rajanna and Hammond (2016) presented GAWSCHI, a gaze and foot-based interaction framework that enables accurate and quick gaze-driven interactions, The authors demonstrated that the gaze and foot-based interactions are as good (time and precision) as mouse-based interactions as long as the dimensions of the interface element are above a threshold. Klamka et al. (2015) combined gaze input with a foot pedal to perform secondary mouse tasks like panning and zooming. The authors found that gaze-supported foot input allows for user-friendly navigation and is comparable to mouse input. Hatscher et al. (2017), demonstrated the usability of gaze- and foot-based interaction on a large monitor in operation theaters. In this setup, a physician performing minimally-invasive interventions can look and interact with medical image data displayed on the large monitor with the gaze input. These prior works demonstrated the feasibility of using a supplemental foot input with gaze for discrete point-and-click interactions. This motivated us to further explore using foot input in gaze typing.

2.2. Gaze typing

2.2.1. Dwell-based gaze typing

In dwell-based gaze typing, a user fixates her gaze on the target character for the duration of dwell time to select it. In their study of a dwell-based gaze typing system, Majaranta et al. (2003) found that the kind of feedback method influences the text entry speed and error rate, and an auditory feedback is more effective than visual feedback. Furthermore, Majaranta et al. (2009) also studied the effects of adjustable dwell time on the performance of gaze typing. The authors found that using adjustable dwell time novices' text entry rate increased from 6.9 wpm in the first session to 19.9 wpm in the tenth session. Also, the dwell time decreased from an average of 876 ms to 282 ms, and the error rates decreased from 1.28% to 0.36%. Hansen et al. (2003) conducted a comparative study by varying the selection method on a Danish on-screen keyboard "GazeTalk." The authors showed that dwell time selection on keys is a little slower, introduces more errors, and has a higher overproduction rate than click-based selection. They further found that a major problem with dwell-based selection is that the participants cannot just leave the mouse pointer anywhere on the screen as they normally would do with mouse. Also, if they forgot to "park" the mouse in a text field, it would activate the button below it inadvertently and this adds to the overproduction rate. A large group of participants in the experiment found that a dwell time of 500 ms is too short, but 750 ms was comfortable.

In another study on "GazeTalk," Hansen et al. (2004b) compared gaze input against head, and mouse inputs. Gaze input was found to be the slowest of the three inputs with a typing speed of 6.26 WPM. Mott et al. (2017) presented a cascading dwell gaze typing system that dynamically adjusts the dwell time of keys based on the likelihood that a key will be selected next. In a longitudinal study involving 17 non-disabled participants gaze typing over 8 sessions, they were able to achieve a typing mean speed of 12.39 WPM compared to a typing speed of 10.62 when using static dwell time. Räihä and Ovaska (2012) conducted an extensive dwell-based gaze typing study over four weeks where each participant gaze typed for almost five hours. In the learning phase an average typing speed of 16 WPM was achieved, and the average dwell time reached in the last phase (10th) was 300 ms. In the extended phase, a typing speed of 20–24 WPM was reached.

2.2.2. Dwell-free gaze typing

A user does not fixate, but gazes over the target characters when using dwell-free gaze typing. The system uses the gaze path and an internal dictionary to suggest a possible set of words for selection. MacKenzie and Zhang (2008), implemented a dwell-free gaze typing system with word and letter prediction. The authors showed that letter prediction is as good and in some cases even better than word prediction. Kurauchi et al. (2016) presented "EyeSwipe," a dwell-free gaze typing system that uses gaze paths. First and last characters of the word are selected using reverse crossing technique, and the middle characters are selected by glancing at their vicinity. From a user study involving 10 participants, a gaze typing speed of 11.7 WPM was achieved after 30 min of typing. Pedrosa et al. (2015), presented "Filteryedping," a dwell-free gaze typing system. The interface filters out unintentionally selected keys from the sequence of letters looked at by the user, and a candidate list of words are presented for selection. The results showed that the system allowed a typing speed of 15.95 words per minute after 100 min of typing.

2.2.3. Gaze typing with multimodal input

There have also been a few multimodal gaze typing approaches where a user focuses on the character with their gaze, and a secondary input is used to select the character. Zhao et al. (2012) proposed a multimodal approach for gaze typing to prevent Midas' touch associated with Dwell. The system uses gaze for pointing and tooth-clicks for key selection. In a study involving six subjects, the system resulted in a gaze typing speed of nearly 5.5 WPM, an error rate higher than dwell-based system. Also, the authors shared that tooth-click is not convenient for applications that require very frequent activation commands over long periods of time, as in the case of typing. Kumar et al. (2020) presented TAGSwipe a multimodal approach that combines gaze path with touch gestures for text entry. In a study involving 12 participants, the system achieved a typing speed of 15.46 WPM.

Beelders and Blignaut (2012) implemented a gaze and speech-based multimodal system which achieved an average typing speed of 0.2 to 0.3 characters per second, and the gaze typing speed did not improve even after multiple sessions of typing. Feng et al. (2021) presented a bi-modal typing interface by combining gaze and head gestures. After 8

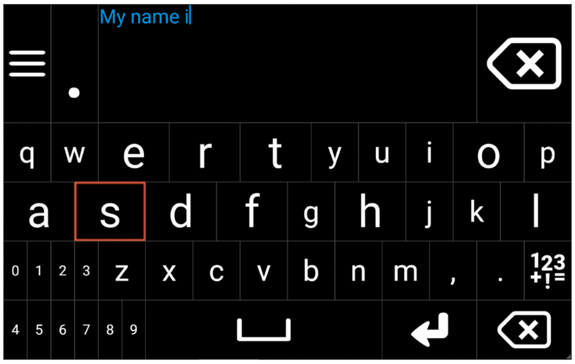


Fig. 2. An enhanced QWERTY keyboard: the keyboard layout is customized such that the frequently used keys have larger dimensions, infrequently used symbolic keys are moved to a secondary screen, and the backspace key is made redundant to help correct errors quickly.

experimental sessions with HGaze typing, participants achieved an average text entry rate of 11.5 WPM. Hedeshy et al. (2021) incorporated humming as a selection method in gaze typing. On a VKB that supported word-level text entry using gaze paths, the participants achieved an average speed of 20.45 WPM in the last session. Other notable works include a system by Cecotti (2016) that combined gaze input with external switch for typing on a menu selection based VKB. Similarly, Meena et al. (2016) also combined gaze with external switch for typing on a Hindi VKB. Lastly, Tuisku et al. (2016) conducted longitudinal typing study by combining gazing and smiling.

2.3. Gaze assisted text entry using gaze gestures

Text entry systems based on discrete gaze gesture leverage on the principles of sketch recognition, where a few semantically associated strokes are interpreted as a shape (Hammond and Davis, 2005; Rajanna et al., 2017). In this method, every character is encoded into a set of strokes such that each set is uniquely identified with a character. To enter a character, the user draws strokes on a canvas in the order specified, and the system recognizes these set of strokes as a character (Majaranta, 2011). Text entry systems that use discrete gaze gestures generally have a lower typing speed and higher learning time. Wobbrock et al. (2008), presented "EyeWrite" a gaze gesture-based text entry system that encodes letter like gesture sets for each character. With "EyeWrite" the users achieved a speed of 5 WPM on average, whereas users achieved a speed of about 7 WMP on a virtual keyboard with dwell-based selection. Some of the other notable gaze gesture-based text entry systems include "QuickWriting," by Bee and André (2008), "Eye-S, " by Porta and Turina (2008), and "StarGazer," by Hansen et al. (2008). While the systems discussed so far use discrete gaze gestures, Ward et al. (Ward et al., 2000) presented "Dasher," a system that uses continuous gaze gestures and language modeling to support efficient text entry. Dasher achieved a typing speed of up to 34 WPM.

In summary, most of the works we have discussed in this section try to address the limitations associated with gaze typing systems. Using variable dwell time (Majaranta et al., 2009) or cascading dwell time (Mott et al., 2017) helps to **improve the typing speed and reduce errors.** Using dwell-free typing (Kurauchi et al., 2016; MacKenzie and Zhang, 2008; Pedrosa et al., 2015), gaze gesture-based typing (Hansen et al., 2008; Ward et al., 2000; Wobbrock et al., 2008), or multimodal approaches (Kumar et al., 2020; Zhao et al., 2012) **either minimize or completely eliminate dwell-based selection**. Implementation like AugKey (Diaz-Tula and Morimoto, 2016) **minimizes eye movements** by augmenting keys suffixes to speed up typing with word prediction. While the majority of improvements in gaze typing focused on optimizing dwell time, improving word and character predictions, or using custom keyboards, in this research we focus on a multimodal, dwell-free approach using the foot input. We believe, a multimodal gaze typing system that addresses the usability issues while also achieving efficiency is critical.

3. Research questions

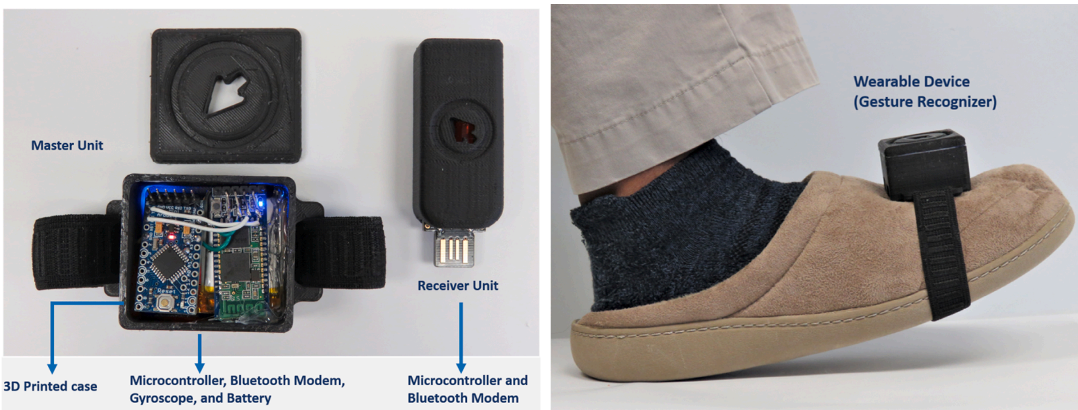
We understand that a gaze typing system that improves typing performance at the cost of increased physical and mental strain, and which requires significant learning would not be acceptable. Hence, with a goal of improving both typing performance (typing speed and reduced errors) and overall gaze typing experience, we formulated the following research questions.

1. Is a multimodal gaze typing system with a supplemental foot input feasible? Is the learning required extensive? Does such a system feel intuitive?
2. Does the foot-based selection at least match or improve the gaze typing performance compared to dwell-based selection?
3. Is foot gesture-based selection better than foot press-based selection or vice versa? How the typing speed and error rate compare between the two methods?
4. What foot gestures do participants find convenient to use and why? Do they switch between different gestures to prevent stress on the foot?
5. Does a supplemental foot input in a gaze typing system induce physical strain since the activations are frequent?

4. Design motivation

To support comfortable gaze typing experience while also improving the typing performance, we strongly believe gaze should only be used for pointing and a secondary input should be used for selection. Kumar et al. (2007) had also shown that combining gaze with key-based activation is a compelling multimodal approach that does not overload visual channel. The elimination of dwell time eliminates related issues like needing to focus until the dwell time elapses or even adjusting an optimal dwell time itself. Importantly, with this design a user does not have to "park" the cursor when not typing which is a notable usability improvement. Hence, in our solution, we replaced dwell-based selection with a direct and instantaneous method - foot-based selection. To achieve foot-based selection, we improved on the design of the foot-operated device presented by Rajanna and Hammond (2016) to create a wearable that has small form factor and easy to operate. We created two wearables, one recognizes foot press actions by sensing the pressure applied and the other recognizes foot gestures by sensing directional motions. We hypothesize that the usage of additional input modality distributes the responsibilities among two input channels, foot and gaze, and this does not strain the user under normal usage conditions.

Next, we elaborate on the rationale behind selecting foot as a supplemental input when other input modalities were available. Velloso et al. (2015) discussed successful applicability of foot input in human-computer interaction across various use cases. As discussed previously, works like (Hatscher et al., 2017; Klamka et al., 2015; Rajanna and Hammond, 2016; Rajanna, 2016b) have already explored gaze and foot-based coarse point and click interactions on a computer, and they found that the foot is one of the promising supplemental inputs to be combined with gaze. While we decided on replacing dwell-based selection with foot input, we had to justify the target user group of our system. First, the system is intended to be used by individuals with physical impairments and disabilities, who have even a little control over their feet. For example, individuals with missing or underdeveloped arms by birth, individuals that experienced paralysis in arms (quadriplegia), individuals with stiff or shaky arms, and for those that lost arms in an accident. The Centers for Disease Control and Prevention estimates about 1500 babies in the United States are born with upper



(a) Foot Gesture Recognition Device (b) Gesture Recognizer attached to Footwear

Fig. 3. (a) Foot Gesture Recognition Device: the master and receiver units. The master unit is attached to user's footware, and the receiver unit is connected to the computer through USB port. (b) Gesture Recognizer attached to Footwear - Master unit: the entire circuitry of the master unit is housed inside a3D-printed container that is attached to the user's footware. The user is executing a toe tap gesture.

limb reductions each year (Parker et al., 2010). Furthermore, according to the statistics provided by amputee coalition[1], 2 million Americans live with limb loss or limb difference, and more than 28 million are at risk of amputation surgery (Ziegler-Graham et al., 2008). 35% of these limb losses are upper limb amputations (fingers or arm) (Ziegler-Graham et al., 2008), and the primary reason is being involved in a traumatic incident (traffic, workplace, firearm, agricultural accidents) (Braza and Martin, 2020). Our foot input devices are designed to be used by individuals with upper limb reduction or loss. Second, the system is also intended to be used in the scenarios of situationally-induced impairments and disabilities. For example, a physician wanting to enter short texts or in general interact with a computer during a medical intervention (Hatscher and Hansen, 2018; Hatscher et al., 2017), will not be able to use a physical keyboard and mouse due to sterilization issues. Similarly a factory worker with greasy hands or wearing thick gloves that wants to enter text on a computer can be benefited by our system. Presently, text entry tasks in the scenarios of situational impairments are explored in the context of interactions on mobile devices (Goel et al., 2012; Sarsenbayeva et al., 2017; Wobbrock, 2006) and virtual reality (Bowman et al., 2002; Rajanna and Hansen, 2018). In summary, whenever the hands are engaged in other taks, unstable, or unavailable, our system can be used for text entry.

5. System design and implementation

Based on our design decisions, we created a gaze and foot-based typing system that comprises of three primary modules: (1) Gaze Interaction Server, (2) Virtual Keyboard (VKB), and (3) Foot-Operated Wearable Device. A pictorial depiction of the system is shown in Fig. 1.

5.1. Gaze interaction server

The gaze interaction server is the central module that coordinates between the VKB and foot-operated input device to achieve gaze typing. It runs on the computer and receives input from the foot-operated device. The foot-operated device connects to the central module on the computer over a Bluetooth connection. The gaze interaction server converts the foot input received as a single byte characters into the key selection commands that are understood by the virtual keyboard.

5.2. Virtual keyboard

In our experiment we used a QWERTY virtual keyboard developed from the open-source VKB "OptiKey.[2]" We enhanced the standard QWERTY VKB layout to be suitable for gaze and foot-based typing, and to improve the typing efficiency. The keyboard layout was customized over multiple design iterations. These customizations can be categorized as (a) regrouping, (b) re-sizing, and (c) redundancy. The enhanced keyboard layout is shown in Fig. 2. In the layout regrouping phase, we moved the infrequently used symbolic keys to the secondary screen which can be activated through a menu key, and also moved the numeric keys to the primary screen. In the layout re-sizing phase, we emphasized the most frequently used keys with larger dimensions. Letters with higher relative frequency (as per the dictionary), and some functional keys (space, enter, backspace) were made prominent than the others (Fig. 2). Lastly, as part of introducing redundancy, we added two instances of backspace keys, one at the top row and the other at the bottom row, so that the user can correct errors quickly. The virtual keyboard constantly receives the user's gaze points on the screen as a pair of (X, Y) co-ordinates from the eye tracker (Tobii EyeX[3]). As the user's gaze scans the keys on the keyboard, each key looked at by the user is highlighted along the border of the key with the red color. Once the key is selected either with dwell time or input from the foot, the background of the key is highlighted in blue, an audio feedback is generated ('click' sound), and the character is printed in the writing space.

5.3. Foot gesture recognition device

The foot gesture recognition device consists of two units, a master (sender) and a receiver as shown in Fig. 3 a. We aimed at creating a small form factor foot-operated input device that can be attached to the user's footwear. Hence, the entire circuitry of the master unit is housed inside a 3D-printed container that is attached to the user's footware as shown in Fig. 3 b. The receiver is an USB enabled unit that is connected directly to the computer. The master unit is responsible for recognizing the foot gesture, and sending the appropriate command (e.g., click) to the receiver. The receiver is responsible for executing the command on the computer. The implementation details along with circuit diagrams for both master and receiver units are discussed in Appendices A.1 and A.2 respectively.

[1] www.amputee-coalition.org - Accessed Sept 2021

[2] github.com/OptiKey - Accessed Sept 2021

[3] tobiigaming.com/product/tobii-eyex/ - Accessed Sept 2021

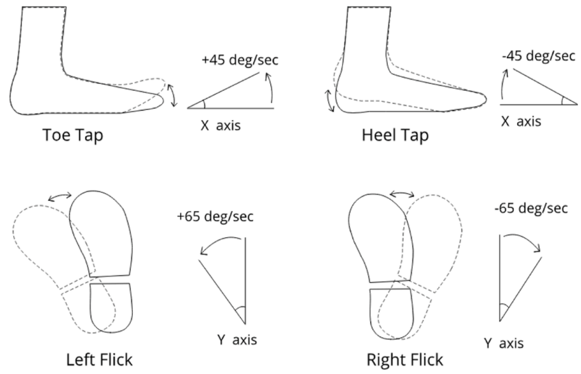


Fig. 4. Foot Gesture Recognition: the list of foot gestures that are recognized by the device. The angular velocities indicate the speed at which the user has to tap or flick the foot to trigger a corresponding gesture.

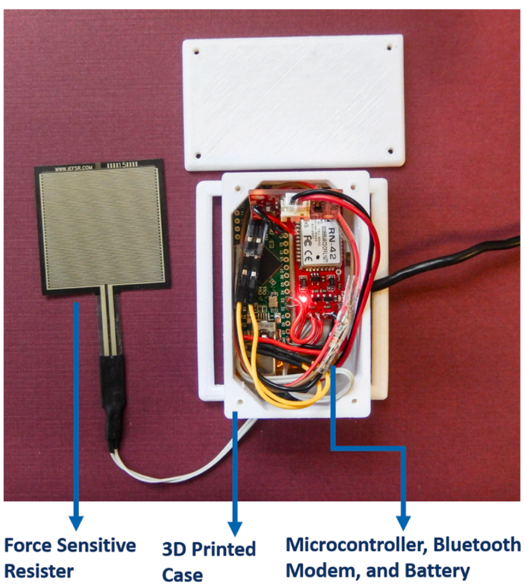


Fig. 5. Foot Press Sensing Device: the entire circuitry is housed inside a 3D-printed container, and the force sensitive resister that senses foot press actions extends from the main circuit and is placed inside the footware.

5.3.1. *Gesture recognition*

The four foot gestures we used are simple and directionally unique in the horizontal and vertical axes. Because of these properties, it is not required to recognize the entire shape of the gesture, but is sufficient enough to check if the angular velocity of the foot has reached a pre-defined threshold in the specific direction to be classified as a specific gesture as shown in the Fig. 4.

Through multiple pilot studies we have identified that from the baseline position of the foot, if its angular velocity changes to +45 deg/s or more along the X axis, the gesture can be classified as toe tap. Similarly, a change in angular velocity of -45 deg/s along the X axis is classified as heel tap. Furthermore, a change of +65 deg/sec along Y axis is classified as left flick, and a change of -65 deg/sec is classified as right flick. Since we found that users can flick much faster than tap with foot, we set velocity threshold for flick slightly higher than tap gestures. The gesture recognition accuracy of the device was nearly 100% as long as the corresponding angular velocity was met.

5.4. *Foot press sensing device*

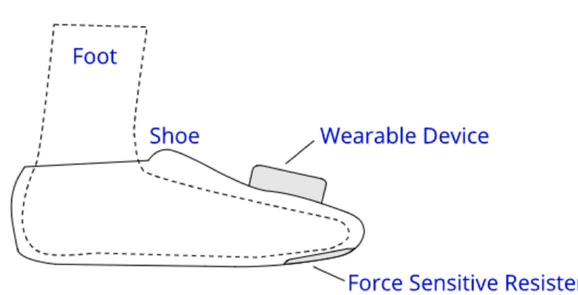
Similar to the foot gesture recognition device, the foot press sensing device has a small form factor and is housed in a 3D printed container, while just exposing the pressure sensor as shown in Fig. 5.

The device is attached to the user's footware as shown in Fig. 6 a, and an outline of how the pressure sensor is placed inside the footware shown in Fig. 6 b. Since the device senses the pressure applied by the user, the system does not require physical movements of the foot, but a gentle press is enough, making it convenient to use. The user input is recognized by measuring the output voltage of a voltage divider circuit as the user applies pressure on the sensor. The implementation details of the pressure sensing device along with a circuit diagram are discussed in Appendix B.

6. Experiment design

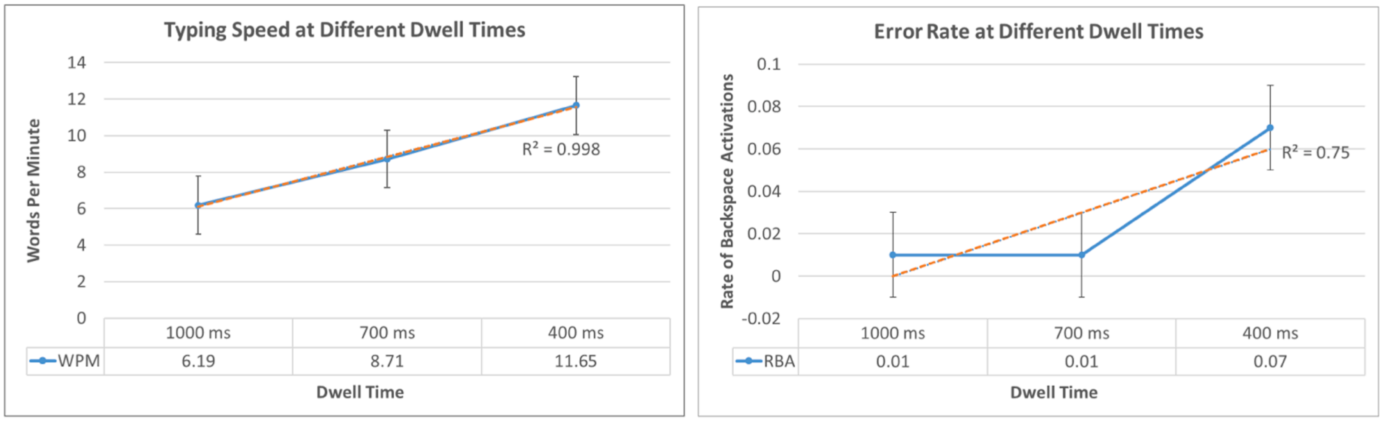
To explore the feasibility of a gaze and foot-based typing system, and compare its performance to existing dwell-based typing systems, we conducted three experiments by involving a total of 51 participants (17 participants in each experiment). The participants were recruited through E-mails and flyers posted on the campus, and the participation was voluntary and they did not receive any rewards. As we received responses, participants were randomly assigned to three experimental groups. In each experiment we used a unique selection method, out of the three selection methods we developed: (1) dwell-based selection, (2) foot gesture-based selection, and (3) foot press-based selection. We also made sure that each subject participated in only one of the three experiments, which means we had a different set of users participating in

(a) Footwear augmented with pressure sensor



(b) Placement of pressure sensor

Fig. 6. Foot Press Sensing Device: (a) Foot Press Sensing device attached to the user's footware. The pressure sensor is placed inside the footware. (b) An outline of how the foot press sensing device is attached to the user's footware, and the placement of the pressure sensor inside the footware.



(a) Typing speed (WPM)

(b) Error rate (RBA)

Fig. 7. Dwell-based Selection: from (a) we observe that the typing speed increases with decreasing dwell time, and the regression line has an R^2 value of 0.99. From (b) we observe that the error rate increases with the decreasing dwell time, and the regression line has an R^2 value of 0.75.

each experiment. We followed this model to avoid any familiarity developed with the gaze typing system by participating in one experiment influencing the user's performance in a subsequent experiment. As our goal was to establish baseline performance of our system, similar to prior studies (Kurauchi et al., 2016; MacKenzie and Zhang, 2008; Majaranta et al., 2003; Wobbrock et al., 2008), no participant in our study had any physical impairment or disability. The enhanced keyboard used in our study did not have word or character suggestion features, and the participants were asked to correct all the errors in the entered text. For eye tracking we used Tobii EyeX tracker[4] which has an accuracy of around 1° of visual angle (Feit et al., 2017). At the beginning of the study each participant was calibrated through the tracker's standard calibration procedure, and was allowed to move their head freely during the study. The eye tracker was attached to a 23″ monitor, with a 1900 × 1200 resolution (19.5″ × 12.19″ screen size). The details of each phase, specifically the task performed, and the results are discussed in the results and discussion Section 7.

7. Results

The efficiency of our gaze typing system was evaluated based on a text-focused and key-selection-focused metrics. The two gaze typing metrics we considered were Words Per Minute (WPM), and Rate of Backspace Activation (RBA), shown in Eq. (2) and 3 respectively (Majaranta, 2011). In our experiments, the participants entered phrases from MacKenzie and Soukoreff (2003), a collection 500 phrases commonly used for evaluation of text entry techniques.

$$Words\ Per\ Minute\ (WPM) = \frac{Number\ of\ Characters}{Time\ Spent\ for\ Typing\ (min) \times 5} \quad (1)$$

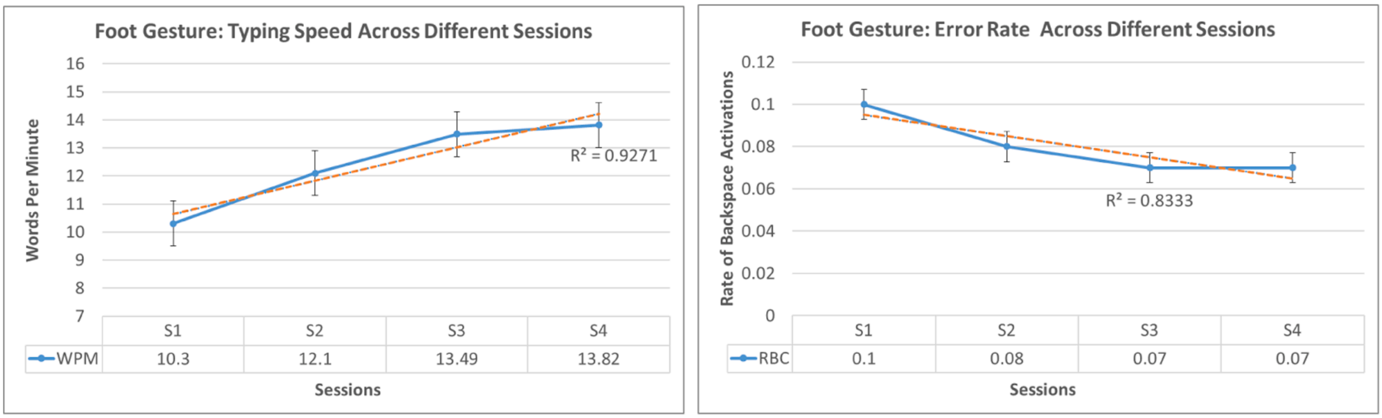
$$Rate\ of\ Backspace\ Activation\ (RBA) = \frac{Number\ of\ Keystrokes\ for\ Backspace\ or\ Delete}{Number\ of\ Characters\ Typed} \quad (2)$$

7.1. Experiment 1: Gaze and dwell-based typing

In this experiment, the participants gaze typed using dwell-based selection. 17 participants (9 male, 8 female) with their ages ranging from 21 to 32 ($\mu_{age} = 23.5$) participated in this experiment. We choose three different dwell times: 400 ms, 700 ms, and 1000 ms. The three different dwell times were chosen based on the most common least, average, and maximum dwell times used in the prior studies (Hansen et al., 2001; 2003; 2004a; Majaranta et al., 2009; Majaranta and Räihä, 2002; 2007). While we tested dwell times lower than 400 ms (300 ms, 350 ms) with a few pilot participants, we learned that the system does not support a comfortable gaze typing experience as is it forced the user to keep up with the pace of the system, and also resulted in a high error rate. At the end, a few participants did mention that even 400 ms of dwell time was challenging, but none left the study uncompleted. Each participant typed 10 phrases with each of the three dwell times, first starting with 1000 ms, next 700 ms, and lastly with a dwell time of 400 ms. Participants did type a few practice phrases with each dwell time. Also, they rested for five minutes before switching to a different dwell time. In total, each participant typed a total of 30 phrases, and overall 510 phrases were entered by 17 participants (17 × 30) across three dwell times. Fig. 7a and b show WPM and RBA respectively across the three different dwell times.

The highest mean typing speed of 11.65 Words Per Minute (WPM) was achieved with a dwell time of 400 ms, and the corresponding mean Rate of Backspace Activation (RBA) was 7% (highest error). The Least mean error rate, i.e., a 1% RBA was achieved with a dwell time of 1000 ms, the corresponding mean typing speed was 6.19 WPM (lowest typing speed). From Fig. 7b, we observe that the error rate increases with the decreasing dwell time. The reason for this observation is two fold: first, with a higher dwell time such as 1000 ms, the user gets enough time to search for the character, and also quickly recover from inadvertent selections by looking away from the character before the dwell time threshold elapses. This helps in achieving a significantly lower error rate. Second, with a shorter dwell time like 400 ms, the user is expected to shift their gaze quickly between desired characters, without inadvertently selecting others during visual search. Since it is challenging to

[4] help.tobii.com/hc/en-us/articles/212818309-Specifications-for-EyeX



(a) Typing speed (WPM)
(b) Error rate (RBA)

Fig. 8. Foot gesture-based Selection: from (a) we observe that the typing speed increases with subsequent sessions, and the regression line has an R^2 value of 0.92. From (b) we observe that the error rate decreases with subsequent sessions, and the regression line has an R^2 value of 0.83.

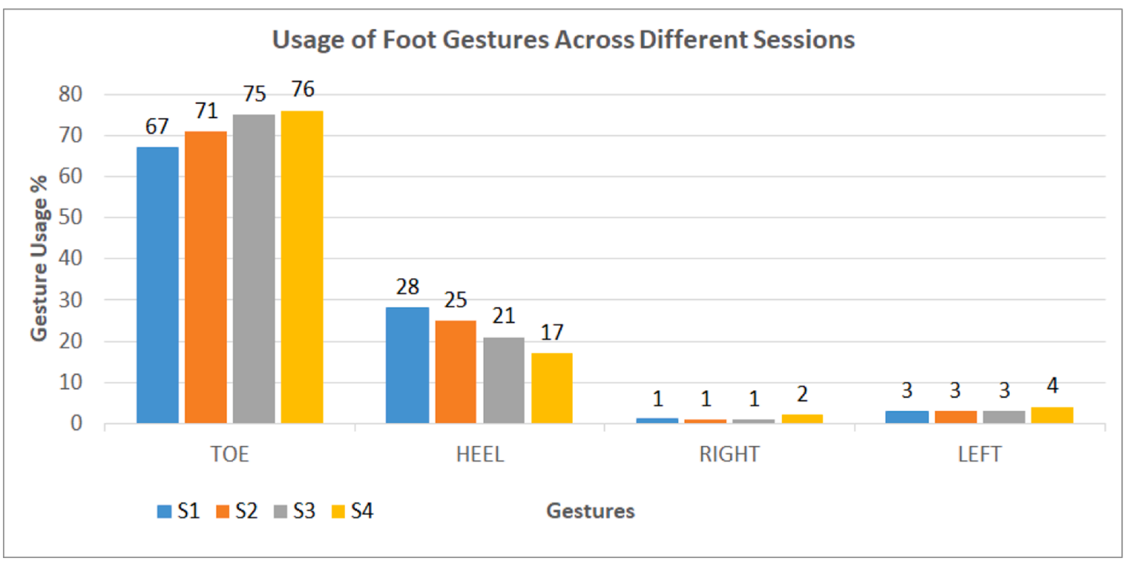


Fig. 9. Percentage of use of each gesture across four sessions (mean values).

avoid inadvertent selections, shorter dwell time results in increased errors.

7.2. Experiment 2: Gaze and foot gesture-based typing

In this experiment, the participants gaze typed using foot gesture-based selection. 17 participants (9 male, 8 female) with their ages ranging from 20 to 27 ($\mu_{age} = 22.3$) participated in this experiment. The experiment was further divided into four sessions, and in each session a participant typed 10 phrases. Hence, each participant typed a total of 40 phrases, and overall 680 phrases were entered by the 17 participants (17 $\times$ 40) across four sessions. All sessions were conducted in one sitting with a gap of five minutes of resting between each session. Based on the prior studies, where foot input was used for interacting with computers (Velloso et al., 2015), we considered four gestures: (1) toe tap, (2) heel tap, (3) right flick, and (4) left flick for key selection. Irrespective of the gesture performed, the character focused on by the user gets selected following the completion of a gesture. During the study, the user could use any gesture to select characters, therefore they had the freedom to switch gestures as desired. For example, a user could type the entire phrase with toe tap, or switch between gestures to type each word or character. At the beginning of the study, the user was asked to type a few practice phrases using different gestures to develop familiarization with the four gestures. Fig. 8a and b show WPM and RBA respectively across the four typing sessions.

The highest mean typing speed of 13.82 WPM, with the lowest error rate of 7% RBA were achieved at the end of the fourth session. The lowest mean typing speed of 10.3 WPM, with the highest error rate of 10% RBA were observed at the end of the first session. From Fig. 8a, we observe that typing speed increases with subsequent sessions and the participants reach the highest typing speed at the end of the fourth session. As learned from the post study interviews there are three reasons for this observation: (1) participants get familiar with the foot gestures, i.e., they learn how high the toe or heel needs to be raised, or at what speed the right or left flicks to be performed to achieve key selection, (2) generally, the participants choose a convenient gesture and use it throughout the study. This behavior contradicts our hypothesis that the participants switch to different gestures during the study to reduce strain on a single part of the feet, and 3) the participants achieve better synchronization of focusing their gaze on the character and selecting it with a foot gesture with increased exposure to the system. From Fig. 8b, we observe that error rate decreases with subsequent sessions with the participants reaching the lowest error rate at the end of the fourth session. During the initial sessions, generally a user switches

Table 1
Foot Gesture-based Selection: ANOVA tests to understand if a gesture is used equally across the sessions.

	Sessions [S1, S2, S3, S4]	Std. error	Post hoc analysis
Toe	$F(3,48) = 1.376$ $p \approx 0.261$	S1 = 8.450 S2 = 8.691 S3 = 8.828 S4 = 8.424	(S1, S2) $p \approx 1.000$ (S1, S3) $p \approx 0.907$ (S1, S4) $p \approx 1.000$ (S2, S3) $p \approx 1.000$ (S2, S4) $p \approx 1.000$ (S3, S4) $p \approx 1.000$
Heel	$F(3,48) = 2.180$ $p \approx 0.103$	S1 = 8.276 S2 = 8.007 S3 = 7.648 S4 = 7.477	(S1, S2) $p \approx 1.000$ (S1, S3) $p \approx 0.704$ (S1, S4) $p \approx 0.593$ (S2, S3) $p \approx 1.000$ (S2, S4) $p \approx 0.773$ (S3, S4) $p \approx 1.000$
Right	$F(3,48) = 0.765$ $p \approx 0.519$	S1 = 1.023 S2 = 1.374 S3 = 1.455 S4 = 1.417	(S1, S2) $p \approx 1.000$ *(S1, S3)* $p \approx 1.000$ *(S1, S4)* $p \approx 1.000$ *(S2, S3)* $p \approx 1.000$ (S2, S4) $p \approx 1.000$ (S3, S4) $p \approx 1.000$
Left	$F(3,48) = 0.524$ $p \approx 0.668$	S1 = 2.240 S2 = 2.692 S3 = 2.744 S4 = 3.118	(S1, S2) $p \approx 1.000$ (S1, S3) $p \approx 1.000$ (S1, S4) $p \approx 1.000$ (S2, S3) $p \approx 1.000$ (S2, S4) $p \approx 1.000$ (S3, S4) $p \approx 1.000$

Table 2
Foot Gesture-based Selection: ANOVA tests to understand the learning effects across the sessions.

	Sessions [S1, S2, S3, S4]	Std. error	Post hoc analysis
WPM	$F(3,48) = 47.372$ $p < 0.001$	S1 = 0.601 S2 = 0.556 S3 = 0.642 S4 = 0.471	(S1, S2) $p < 0.01$ (S1, S3) $p < 0.01$ (S1, S4) $p < 0.01$ (S2, S3) $p < 0.01$ (S2, S4) $p < 0.01$ (S3, S4) $p \approx 1.000$
Error	$F(3,48) = 9.793$ $p < 0.001$	S1 = 0.010 S2 = 0.010 S3 = 0.008 S4 = 0.008	(S1, S2) $p < 0.01$ (S1, S3) $p < 0.01$ (S1, S4) $p < 0.01$ (S2, S3) $p \approx 0.528$ (S2, S4) $p \approx 0.960$ (S3, S4) $p \approx 1.000$

their gaze to the next character in the word before selecting the current character, i.e., the eyes move faster than the foot gesture completion. However, as shared in the post study interviews, the participants learn a repeated pattern of gaze pointing and foot gesturing and this improved synchronization leads to reduced error rate with subsequent

Furthermore, we were interested in learning the most and least used gestures, and how the usage of each gesture varies across sessions. We fond that toe tap was the most used gesture and right flick was the least used gesture. The percentage of usage of all gestures throughout the study are - toe tap 72.25%, heel tap 22.75%, right flick 1.25%, and left flick 3.25%. To explore the statistical significance, we performed a one-factor ANOVA with replication. The independent factor was "Gestures" which had four levels: toe tap, heel tap, right flick, and left flick. The dependent variable was "Gesture Usage" metric which was the percentage of each gesture used by each participant. The results show that there is a significant difference in the usage of different types of gestures with an $F(3, 201) = 92.955$, $p < 0.001$. Also, post-hoc analysis with Bonferroni correction indicates that the gesture usage between any pair of gestures is significant ($p < 0.05$), except for left and right flicks ($p > 0.05$). Overall, these results indicate that most participants strongly preferred one gesture which was toe tap. Fig. 9 shows the percentage of usage of each gesture across four sessions, and it is apparent that participants increasingly use toe tap with subsequent sessions.

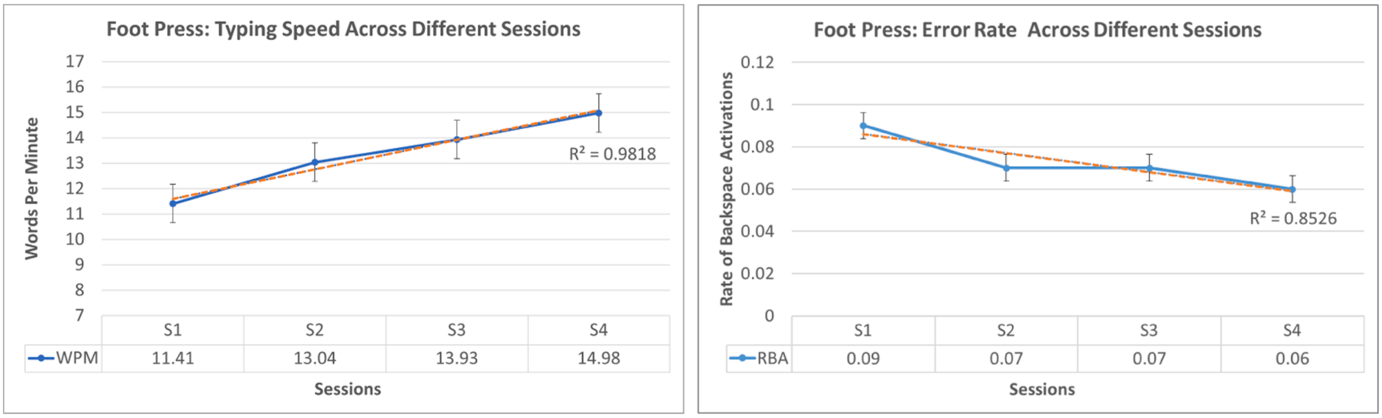
Next, to answer the question, do users choose a single gesture initially and use it throughout the study, or do they change gestures as the study progresses, we conducted one-factor ANOVA with replication. The independent factor was "sessions" which had four levels S1, S2, S3, and S4 which indicate sessions one through four. The four dependent variables were "gesture usage" metric for toe tap, heel tap, right flick, and left flick. A total of four ANOVA tests were performed, and for each ANOVA test we considered one dependent variable of the four dependent variables. The "gesture usage" metric was measured as the percentage of each gesture used, by each participant, in each session. The results of ANOVA tests are presented in Table 1, and we observe that "sessions" is not a significant factor for dependent variables: toe tap, heel tap, right flick, and left flick. This indicates that the difference in the usage of a gesture, e.g., toe tap is not significant across four sessions. In addition, post-hoc analysis with Bonferroni correction indicates that for a given gesture, the difference in its usage is not significant across any two sessions as well. All these observations strongly suggest that though the system supports multiple gestures, a user selects a single, convenient gesture, and uses it throughout the study.

Lastly, we discuss the learning effects of gaze typing when using foot gesture-based selection. We performed a one-factor ANOVA with replication. The independent factor was "sessions" which had four levels: S1, S2, S3, and S4. The dependent variables were typing speed (WPM) and error rate (RBA). From Table 2 we observe that "sessions" is a significant factor for typing speed (WPM). The differences in typing speed across sessions is significant ($p < 0.001$), and the speed increases with subsequent sessions. Also, post-hoc analysis with Bonferroni correction indicates that the difference in typing speed between any pair of sessions is significant ($p < 0.01$), except for sessions 3 and 4. This indicates that the typing speed reaches a plateau at nearly 13.8 WPM. Similarly, the difference in error rate across sessions is significant ($p < 0.001$), and the error rate decreases with subsequent sessions. Post-hoc analysis with Bonferroni correction indicates that the difference in error rate between sessions 1 and 2, 1 and 3, and 1 and 4 are significant ($p < 0.01$). Since the difference in error between sessions 2 and 3, and 3 and 4 is not significant, these results suggest that the participants quickly learn (by session 1) to use the foot gesture-based selection and start making fewer errors, and this behavior continues throughout the study.

7.3. Experiment 3: Gaze and foot press-based typing

Experiment 2 demonstrated that the majority of participants find toe tapping more convenient than other gestures. Also, we observed that the participants generally do not switch to different gestures, however, they pick a gesture initially and use the same gesture throughout the study. These observations motivated us to develop a third selection method: foot press-based selection. The reason for considering foot press-based selection was to achieve higher typing speed than foot gesture-based selection. In this method a Force Sensitive Resistor (FSR) was placed inside the shoe and under the toe area of the foot To select a character, the user needs to perform a subtle press action on the sensor. Unlike the foot gestures, the foot press-based selection does not require any movement of the foot except a subtle press. Hence, we hypothesized that the ease and convenience of foot press-based selection would result in higher typing speed than the foot gesture-based selection.

To test our hypothesis, we conducted an experiment where participants gaze typed using foot press-based selection method. 17 participants (14 male, 3 female) with their ages ranging from 20 to 26 ($\mu_{age} = 22.3$) participated in this experiment. At the beginning of the study each



(a) Typing speed (WPM)

(b) Error rate (RBA)

Fig. 10. Foot press-based Selection: from (a) we observe that the typing speed increases with subsequent sessions, and the regression line has an R^2 value of 0.98. From (b) we observe that the error rate decreases with subsequent sessions, and the regression line has an R^2 value of 0.85.

Table 3
Foot press-based Selection: ANOVA tests to understand the learning effects across the sessions.

	Sessions [S1, S2, S3, S4]	Std. error	Post hoc analysis
WPM	$F(3,48) = 44.324$ $p < 0.01$	S1 = 0.473 S2 = 0.389 S3 = 0.373 S4 = 0.406	(S1, S2) $p < 0.01$ (S1, S3) $p < 0.01$ (S1, S4) $p < 0.01$ (S2, S3) $p \approx 0.155$ (S2, S4) $p < 0.01$ (S3, S4) $p < 0.05$
Error	$F(3,48) = 5.743$ $p < 0.01$	S1 = 0.011 S2 = 0.008 S3 = 0.010 S4 = 0.008	(S1, S2) $p < 0.05$ (S1, S3) $p \approx 0.150$ (S1, S4) $p < 0.01$ (S2, S3) $p \approx 1.000$ (S2, S4) $p \approx 1.000$ (S3, S4) $p \approx 0.985$

Table 4
Top typing speed: ANOVA for WPM and error.

	Selection method [Dwell (DW), foot press (FP), foot gesture (FG)]	Mean	Std. error	Post hoc analysis
WPM	$F(2,50) = 14.844$	DW = 11.648	DW = 0.437	(DW, FP) $p < 0.01$
	$p < 0.01$	FP = 14.98	FP = 0.406	(DW, FG) $p < 0.01$
		FG = 13.814	FG = 0.470	(FP, FG) $p \approx 0.199$
Error	$F(2,50) = 0.208$	DW = 0.069	DW = 0.014	(DW, FP) $p \approx 1.000$
	$p \approx 0.813$	FP = 0.060	FP = 0.008	(DW, FG) $p \approx 1.000$
		FG = 0.068	FG = 0.007	(FP, FG) $p \approx 1.000$

participant typed a few practice phrases to develop familiarity with foot press-based selection. Similar to experiment 2, each participant completed four typing sessions, and 10 phrases were typed in each session. Therefore, a total of 680 phrases were entered by the 17 participants (17 × 40) with each participant typing 40 phrases from four sessions. Fig. 10a and b show WPM and RBA respectively across the four typing sessions.

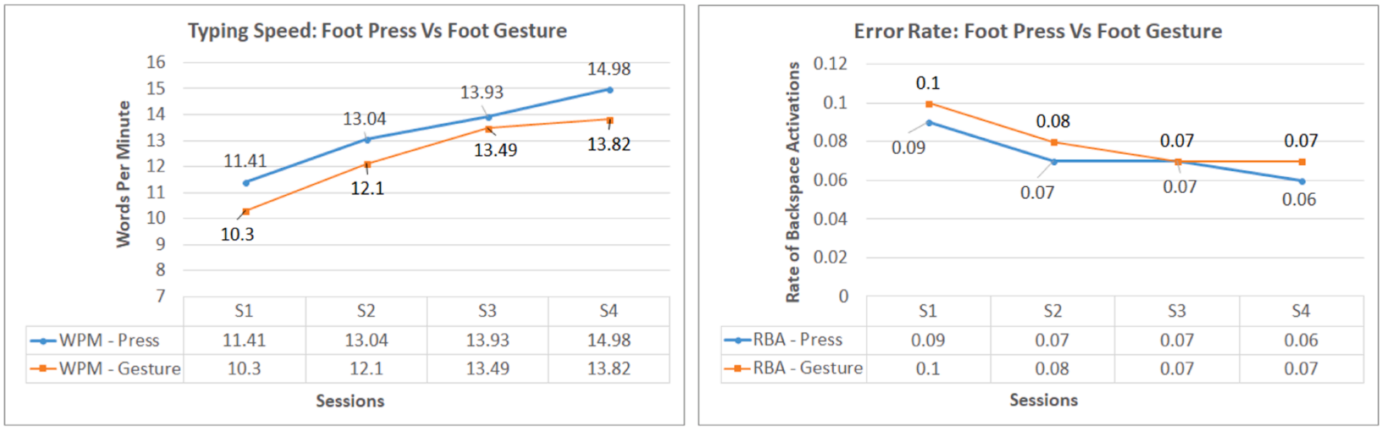
The highest mean typing speed of 14.98 WPM, the lowest error rate of 6% RBA were achieved at the end of the fourth session. The lowest mean typing speed of 11.41 WPM, the highest error rate of 9% RBA were observed at the end of the first session. As observed in experiment 2, Fig. 10a shows that the typing speed increases with successive sessions, and Fig. 10b shows that the error rate decreases with subsequent sessions. This increased typing speed and decreased error rate is attributed to the increased familiarity and synchronization between pointing with gaze and selecting with foot-press.

To understand the learning effects of foot press-based gaze typing, we conducted one-factor ANOVA with replication. Similar to experiment 2, "sessions" was an independent factor with four levels S1, S2, S3, and S4. The two dependent variables were WPM and error rate (RBA). From Table 3 we observe that "sessions" is a significant factor for both WPM ($p < 0.01$) and error rate ($p < 0.01$). From post-hoc analysis with Bonferroni correction on the typing speed between pairs of sessions, we found that the difference in typing speed between a pair of any two sessions is significant ($p < 0.05$), except for sessions 2 and 3. Similarly, post-hoc analysis with Bonferroni correction on the error rate between a pair of any two sessions indicates that the difference between sessions S1 and S2, and S1 and S4 are significant ($p < 0.05$). Similar to foot gesture-based activation, since there is no significant difference in the error between session 2 and 3, and 3 and 4, these results suggest that the participants quickly learn to use the foot press-based selection.

7.4. *Gaze typing performance: dwell vs. gesture vs. press*

In the previous sections, we analyzed the typing performance of each selection method individually. In this section, we will analyze how the typing performance of each method compares against other methods. First, we analyzed the top typing speeds and associated error rates of each selection method with one-factor ANOVA without replication. The independent factor was the "Selection Method" which had three levels Dwell (DW), Foot Press (FP), and Foot Gesture (FG). The two dependent variables we considered were the typing speed (WPM) and error rate. The top typing speed and associated error rate for dwell-based selection was considered from the experiment where dwell time was set to 400 ms. Similarly, the top typing speed and associated error rate for foot gesture-based and foot press-based selections were considered from session 4 of the experiment.

Table 4 lists the results of ANOVA tests. We observe that the difference in typing speed between selection methods is significant ($p < 0.01$). Post-hoc analysis with Bonferroni correction indicates that the difference in typing speed is observed mainly due to the significant difference in typing speed between dwell and foot press ($p < 0.01$), and dwell and foot gesture ($p < 0.01$), but no significant difference was observed



(a) Typing speed (WPM) (b) Error rate (RBA)

Fig. 11. Comparison of typing speed (WPM) and error rate (RBA) between foot press and foot gesture-based selection methods across four sessions.

Table 5
Mixed factor anova: WPM and error.

	Selection method [Foot press, foot gesture]	Sessions [S1, S2, S3, S4]	Interactions
WPM	$F(1,32) = 2.008$	$F(3,96) = 90.179$	$F(3,96) = 1.060$
	$p \approx 0.166$	$p < 0.001$	$p \approx 0.370$
Error	$F(1,32) = 0.229$	$F(3,96) = 14.227$	$F(3,96) = 0.733$
	$p \approx 0.635$	$p < 0.001$	$p \approx 0.535$

Table 6
Mixed factor ANOVA: post hoc analysis for sessions.

	Mean [S1, S2, S3, S4]	Std. error	Post hoc analysis
			(S1, S2) $p < 0.001$
	S1 = 10.858	S1 = 0.382	(S1, S3) $p < 0.001$
	S2 = 12.570	S2 = 0.339	(S1, S4) $p < 0.001$
WPM	S3 = 13.712	S3 = 0.371	(S2, S3) $p < 0.001$
	S4 = 14.398	S4 = 0.311	(S2, S4) $p < 0.001$
			(S3, S4) $p < 0.05$
			(S1, S2) $p < 0.001$
	S1 = 0.095	S1 = 0.007	(S1, S3) $p \approx 0.907$
	S2 = 0.073	S2 = 0.006	(S1, S4) $p < 0.001$
Error	S3 = 0.070	S3 = 0.007	(S2, S3) $p \approx 1.000$
	S4 = 0.064	S4 = 0.006	(S2, S4) $p \approx 0.325$
			(S3, S4) $p \approx 1.000$

between foot gesture and foot press ($p \approx 0.199$). Furthermore, the difference in the error rate between the selection methods is not significant ($p > 0.05$). These results suggest that though users make the same amount of errors across selection methods, the performance differs by how fast they type using each method.

7.5. Gesture- vs. press-based selection

Since we focused primarily on foot-based selection, we further analyzed the gaze typing performance by only considering the foot gesture-based and foot press-based selection methods. Fig. 11a and b compares the gaze typing speed and error rate between the two foot-based selection methods.

We performed two-way mixed model ANOVA with replication on dependent variables: WPM, error. The two factors (independent variables) we considered were: (1) selection method, and (2) sessions. The factor "selection method" is a between-subjects factor and it has two levels: (1) foot gesture-based selection, and (2) foot press-based selection. "Selection method" is a between subjects factor since the participants who gaze typed using foot gestures were not involved in the evaluation of foot press-based selection. "Sessions" is a within-subjects factor and has four levels: S1, S2, S3, and S4.

Table 5 lists results of ANOVA. We observe that the difference in neither the typing speed (WPM) nor error rate is significant between the two foot-based selection methods ($p > 0.05$). However, consistent with previous analysis, the difference is significant between sessions for both typing speed ($p < 0.001$) and error rate ($p < 0.001$). Lastly, we observe no interaction effects between the $Selectionmethod \times Sessions$ for both WPM and error rate.

Since 'sessions' was the only factor significant for both the dependent variables, the post-hoc analysis with Bonferroni correction for factor 'Sessions' is shown in Table 6. We observe that the difference in typing speed between any pair of sessions is significant ($p < 0.05$), and the typing speed generally increases with subsequent sessions. We also observe that, difference in error between sessions S1 and S2, and S1 and S4 are significant ($p < 0.001$), but the difference is not significant between sessions S2 and S3, and S3 and S4. This observation with error is similar to what was observed in experiment 2 and 3, which indicates that irrespective of foot gesture-based or foot press-based selection, the participants reduce the error they make from session 1 to 2, from session 2 onward the changes in the errors observed are not significant.

8. Qualitative feedback - gaze typing usability

One of the main goals of our work was to understand the advantages of using a supplemental foot input for gaze typing over just using dwell-based selection. While we initially considered collecting subjective rating using System Usability Scale (Brooke et al., 1996), NASA-TLX (Hart and Staveland, 1988), or use custom Likert-scale questions, we did not use any of these methods because of two reasons. First, since our experiments followed a between subjects design, comparing subjective ratings may not be accurate. Second, subjective ratings will not capture the details of the overall user experience. Hence, we performed a qualitative study by conducting semi-structured interviews where each participant shared their experience gaze typing on our system. We transcribed all interviews and analyzed them using grounded theory methods (Charmaz and Belgrave, 2007). From our analysis we tried to explore the strategies the participants adopted, aspects that assisted in typing faster, aspects that resulted in errors, aspects that influenced their overall experience, and limitations faced. Based on the selection method participants used, following are the central themes we identified that were repeated across the participants.

8.1. Gaze and dwell-based typing

8.1.1. Longer dwell time

Many participants found that with a longer dwell time (e.g., 1000 ms) it is much convenient to type as it allows enough time to search for the character without inadvertently selecting the wrong characters. Even when the user has focused on a wrong character, there is enough time to quickly switch the focus to the correct character. Both these aspects significantly reduces the error rate, and this is also observed in our results (Section 7.1). However, a down side of using a longer dwell time is that the participants found it strenuous to type long phrases, and in general typing for extended time when using a longer dwell time is exhausting on the users. Some users even had complications like eyes filled with water or being turned red once they completed the typing session. Hence, users with sensitive eyes are more likely to experience issues when gaze typing using a longer dwell time.

8.1.2. Shorter dwell time

We found that a shorter dwell time like 400 ms enables fast typing, and it is most suitable for users with sensitive eyes who cannot focus on a key for a longer time. Despite the advantages, a shorter dwell time often results in increased error rate as we have observed in our results (Section 7.1). Also, it is difficult to correctly enter the same character consecutively (e.g., too, see, lottery), and as shared by most participants it is highly cognitively demanding. The main problem users face is that they are forced to look away from the keys or onto the text input area when not typing, and this is an undesirable user experience in an everyday scenario.

8.2. Foot gesture-based selection

8.2.1. Initial learning

All participants felt that using foot gestures required initial learning. P27 shared that “in the first couple of rounds, I was like having slow it down a little bit, and make sure that my eyes were adjusting, I was getting each specific letter, making sure I was not going too fast picking wrong letter, eventually I was able to go faster, because my eyes would kind of like catch as I was going.” Also this learning phase involved processing the feedback and adjusting the foot gesture as P22 shared “I had some difficulty with the consistency of the feedback, trying to figure out just how high up or how much force to go down, but towards the end it was a lot easier to have that kind of measure.” Hence, it is evident that the initial learning is crucial in using foot gestures for gaze typing.

8.2.2. Developing a rhythm

Through the initial learning once the participants learned to synchronize focusing on the character and executing the foot gesture all of them felt that they had developed a rhythm. As it is evident from the following comments, developing a rhythm is the key to leveraging on foot gestures.

- P26: “Once I got the rhythm of it, it seemed like a pretty natural combination to me, but getting started, each sentence I had to remind myself to tap.”
- P30: “I thought it worked pretty well, I felt like you get into a rhythm after a while where you don’t even really think about typing, you just kind of do it every time you look at a new character.”

8.2.3. Intuitiveness and physical strain

Majority of the participants felt that foot gesture-based selection becomes intuitive and the interaction is not strenuous for writing a few phrases. P19 shared “Unless you are writing an essay it’s not strenuous, for what we did it’s not strenuous.” Also P29 shared “After some practice, I definitely got faster and more accurate.” Regarding how natural the foot gestures felt, P31 shared “Tapping the floor feels like tapping the keyboard, I am pressing something, it gives feeling of completion.”

8.2.4. Switching gestures

Contrary to our hypothesis, users hardly switched between foot gestures. Most users picked one convenient gesture and used it throughout the typing session. However, a few users did like the idea of switching between gestures to reduce strain, and they shared that:

- P30: “I used the toe until my foot got tired and then I would use the heel a little bit to give my toe a break.”
- P29: “I initially started using the heel tap the most, but overtime I found that the front foot tap was easier (toe tap) and less strenuous. I rarely used the left and right.”

8.2.5. Sources of errors

As observed during the studies and also as reported by the participants most of the errors are due to participants trying to type faster and losing the synchronization between gaze and foot.

- P24: “The error was not due to hitting the wrong letter, but having to stop at the right letter, your eyes are gone before it’s processed.”
- P26: “Sometimes I try to move my gaze fast before my foot tap that caused a lot of errors, that’s a coordination problem.”

8.3. Foot press-based selection

When analyzing transcripts from participants that used foot press-based selection, we came across a few similar themes like initial learning, intuitiveness, source of error, and so on as found with the analysis of gesture-based action. However, we did find some new themes related to interface design and the design of the foot press sensing device itself.

8.3.1. Initial learning

When instructed about the system, initially participants felt it might be difficult to synchronize gaze with subtle foot press actions. However, within typing a few phrases they felt it is easy to type, foot press became an automatic response, and many even picked up speed. Following are some of comments shared by the participants:

- P48: “It came by instinct, after the first sentence, it was just like an automatic response for the foot to keep on doing that, that was pretty easy.”
- P45: “I was getting faster as I kept typing. It felt pretty good.”

8.3.2. Intuitiveness and physical strain

Participants felt that it was easy to use the system and selecting with foot-press became natural after typing a few phrases. They typed for approximately one hour and most felt typing was not strenuous.

- P38: “The whole gaze typing was pretty neat and easy to use. After typing a few phrases, clicking with the foot became natural.”
- P46: “The system was intuitive and pretty good.”

However, a few participants (nearly 17%) did mention that if the system is expected to be used for prolonged time, it would result in foot strain.

- P50: “For short phrases it is not straining, if I was googling something, or like having a short conversation it is fine.”
- P40: “The system becomes a bit tiring after prolonged use (mostly my toe, my eyes were fine).”

8.3.3. Sources of errors

It was quite interesting to observe that the source of errors when using foot press-based selection was different from foot gesture-based selection. Most of the errors were caused by double selecting the same character as the participants were adjusting the pressure exerted on the sensor while they typed initial few phrases.

- P41: “The foot sensor could have more feedback on it when pressed. It was too easy to press multiple times, and better feedback like a click or vibration when pressed would help that.”
- P42: “Once you know how much pressure you need to put, it is ok. At first I was double typing.”

9. Discussion

In this section we will revisit the research questions discussed in Section 3. First, we wanted to test the feasibility of using supplemental foot input in gaze typing. Results from our experiments indicate that with a short learning curve, users do conveniently coordinate their gaze and foot input to enter text on a computer. Also, typing at a comfortable speed with foot-based selection, the typing speed appears to reach a plateau at round 15 WPM. This is likely due to the physiological limitation of how quickly users can move their foot. Regarding the error rate, from both experiment 2 and 3, we observe that the error steeply reduces from session 1 to 2, and there onward no significant change is observed. This indicates that users quickly learn to coordinate their gaze and foot, and make significantly less errors.

Second, we wanted to compare the gaze typing performance of foot-based selection with a widely used method like dwell-based selection. While we did not use variable dwell time, we did use a dwell time of 400 ms which is in the range of the lowest dwell times used in the previous studies (Hansen et al., 2001; 2003; 2004a; Majaranta et al., 2009; Majaranta and Räihä, 2007), and many participants also reported that dwell time of 400 ms was demanding and unnatural to use. With a dwell time of 400 ms the average typing speed achieved was 11.65 WPM in contrast to 13.82 WPM with foot gestures and 14.98 with foot press-based selection. ANOVA test discussed in Section 7.4 showed that the difference in typing speed was significant between dwell and foot based selection methods ($p < 0.05$), however, there was no significant difference in the error rate ($p > 0.05$). These results indicate that foot-based selection at least matches, and likely improves, the gaze typing performance compared to dwell-based selection.

Third, we wanted to compare the performance of foot gesture-based selection to foot press-based selection. ANOVA tests in Section 7.5 revealed that there is no significant difference in typing speed as well as error rate between the two selection methods ($p > 0.05$). Since foot press requires almost no foot movement compared to a foot gesture, we expected that foot-press based selection would achieve a higher typing speed. Though foot press-based selection generally achieved a speed of 7% higher than foot gesture-based selection, the difference was not significant. Also, there was no significant difference in the error rate ($p > 0.05$). Hence, we infer that subtle foot press-based selection or distinctive foot gesture-based selection achieve the same performance.

Fourth, we wanted to understand the pattern of usage of foot gestures and user strategies. ANOVA tests from experiment 2 (Section 7.2) demonstrated that though users were provided with four gestures to be used as selection methods, users generally choose a single gesture and use the same gesture throughout the study. We also observed that users hardly switched between foot gestures, but whenever they did, they switched between toe and heel tapping or right and left flicks - the switch was always between symmetric gestures. Also, though the user may switch to a different gesture, she quickly returns back to the primary gesture. This observation again contradicts our hypothesis that the availability of multiple gestures encourages users to switch between gestures when they get tired with one gesture. This behavior is likely due to the fact that a user might have already developed a rhythm with the primary gesture, and switching gestures has an additional cost of learning time and hence reduced performance. Furthermore, since toe tapping was used significantly higher than other gestures, also as most shared that they preferred toe tapping out of all gestures, we infer that toe tapping is the most efficient and convenient gesture for foot gesture-based interactions. Also, we suggest, unless multiple gestures are required to achieve different interactions, it is best to incorporate only toe tapping gesture. Redundancy of gestures may not improve performance.

Fifth, since foot input is going to be frequent in a gaze typing task, we wanted to understand the physical strain associated with foot-based selection. Though participants typed for one hour with intermittent breaks between the sessions, except for a few participants the majority of participants (nearly 88%) reported that foot interaction, either gestures or press, was not strenuous when inquired during the post-study interview. Some participants did express that typing for nearly one hour was quite strenuous on their eyes but not as much on their foot. These results suggest that the majority of participants would experience minimal to no physical strain with foot-based selection when typing for short duration like an hour.

To summarize, our motivation was to explore a foot input based multimodal approach to gaze typing. We expected such an approach to be convenient to use, require minimal learning, at least match or improve the performance compared to dwell-based selection, and importantly address the usability issues associated with dwell-based gaze typing. Comparing foot-based selection with various prior works that used dwell-based selection, we find our solution to be promising. While our solution may not achieve performance like Dasher (Ward et al., 2000) (34 WPM after significant training), its mean typing performance of nearly 15 WMP (max 18.18 WPM) is similar to, and in some cases better than, the majority of dwell-based typing systems. Error rate steeply reducing from session 1 to 2 and staying low from there onward combined with qualitative feedback indicate minimal learning.

Using a supplemental foot input eliminates the need to dwell on the keys, which in turn helps not to overload the visual channel for selection tasks. Importantly, since the MIDAS touch issue is avoided, there is no need to park the cursor when not typing. Furthermore, from our studies we had a better understanding of using a supplemental foot input for gaze typing. From the interviews we found that the performance of gaze and foot-based typing is primarily dependent on coordination of pointing with gaze and selecting with foot (press or gestures). Often, participants reported that the major reason for errors was not because they typed the wrong letters, but that their gaze shifted before the current letter had been selected. The knowledge gained regarding using foot-based selection methods can be applied beyond gaze typing to any gaze assisted interactions. Overall, the participants seemed excited with the possibility of gaze and foot-based typing, and highly liked the usability and applicability of our system specifically in the scenarios of situationally induced impairments and disabilities.

10. Limitations

10.1. Target user group

While dwell-based selection caters to a wide range of users with physical impairments, our solution, that uses foot input, is limited to users that have at least some control over their foot. Though our target group appears to be limited, the wearable sensors can be modified (3D printed) to be placed anywhere on the body so that the user can perform a gesture or press action for selection. Furthermore, the current study neither included participants with disability nor scenarios of situational impairments were simulated. All experiments were conducted in a lab setting as it helped to capture the baseline performance of the three selection methods we considered. In future studies, we plan to include users with disability, and also simulate scenarios of situational impairment.

10.2. Composition of participants

While both Experiment 1 and 2 had a comparable male and female participants distributions of 53% and 47% respectively. However, Experiment 3 had a skewed male and female participants distributions of 82% and 18% respectively. In future longitudinal experiments, we will avoid such imbalance in the gender distribution of participants.

10.3. Strain with prolonged usage

Most participants expressed that using foot-based activation for a short duration like typing emails, notes, commenting, exploring web, and so on is engaging and would not result in foot strain. However, the foot will strain if typing for longer duration like writing an essay, documentation and so on.

10.4. Design of the foot input device

we received suggestions for improvements related to press sensing area, gestures and feedback. We will be incorporating these feedback in future iterations of the foot input devices.

11. Conclusion

Gaze typing is becoming one of the crucial input modalities for text entry in two scenarios: (1) situationally-induced impairments and disabilities (SIID), and (2) physical impairments. To address performance and usability issues associated with gaze typing, we present a dwell-free, multimodal, gaze typing system that uses a supplemental foot input to select characters. With this design, a user focuses on the target character with their gaze and selects it with the foot input. We implemented two methods of foot-based selection: (1) foot gestures, and (2) foot press. The foot gestures we supported are toe tap, heel tap, right flick, and left flick, and any gesture could be used to select the character. Additionally, we enhanced the standard QWERTY keyboard by modifying the layout, and the dimension of the keys to improve gaze typing performance. We tested the performance and usability of all three selection methods–dwell, foot gestures, and foot press–through three experiments. Each experiment had 17 participants, and a total of 51 participants took part in the study.

From the three experiments, we found multiple observations. First, users can comfortably coordinate their gaze and foot input to enter text on a computer, and the learning required is minimal. Overall, foot-based selection at least matches, and likely improves, the gaze typing performance compared to dwell-based selection. Second, while subtle foot press-based selection may appear to be less straining and a faster selection method, we found no difference in the performance between foot gesture and foot press-based selection methods. Third, though the system supported multiple foot gestures with an intent that users would switch between gestures to reduce strain, users preferred to select a single gesture and use it throughout the study. Users do not prefer to switch between using different gestures to avoid the time and effort involved in familiarizing a new gesture. Fourth, toe tapping is the most preferred foot gesture for gaze typing, and we believe this also translates to point-and-click interactions on a computer. Lastly, when using foot-based selection, either foot gestures or foot press, users quickly develop a rhythm between pointing on the character with their gaze and performing foot-based selection. In summary, our results suggest that using a supplemental foot input with gaze typing, or in general gaze-assisted interactions, would improve both performance and user experience.

12. Future work

We will be further focusing on addressing the current limitations, and adding new features to the gaze and foot-based typing system. First, one of the concerns expressed by the participates was that gaze and foot-based typing might become strenuous after typing continuously for prolonged time. We will try to address this issue by testing multiple designs of the wearable device, the amount of pressure to be applied, and the placement and dimensions of the pressure sensor. Second, a longitudinal study (4 weeks) will be conducted by including both participants with and without any motor impairments. The current study included only participants without any disability as the goal was to establish the baseline performance of the system. The goal of the longitudinal study is to understand if the participants would consistently achieve the maximum typing speed in the first session itself and observe if foot-based target selection becomes natural. Third, we will be enhancing the wearable device to encode text editing commands as simple movements of the feet. For example, a left flick would delete the last typed word, or two left flicks would delete an entire line of text. Fourth, we will develop multiple designs of the 3D printed casing to house the circuit so that the design is appropriate for specific kinds of disability or situational impairment. Lastly, we will focus on improving the feedback on the foot-operated device. Currently, only auditory feedback is provided for a click action, however, we want to further enhance this by incorporating tack-tile feedback.

CRediT authorship contribution statement

Vijay Rajanna: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Murat Russel:** Conceptualization, Validation, Data curation, Writing – original draft, Writing – review & editing. **Jeffrey Zhao:** Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Tracy Hammond:** Conceptualization, Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank the members of the Sketch Recognition Lab for their feedback and support during this work. We thank the staff at Engineering Design Center for allowing us to use the facility to develop the prototypes. Lastly, we thank all the participants for their time and feedback. Please note that this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Foot gesture recognition device - implementation

A1. Master unit

The circuit diagram of the master unit is shown in Fig. A.12. The unit is built using four main modules: (1) a motion processing unit (MPU-6050) with gyroscope and accelerometer, (2) an Arduino Pro-Mini Microcontroller, (3) a Bluetooth Module (HC-05), and (4) a Battery Recharging Unit (Adafruit Powerboost 1000C). The device is powered by a rechargeable battery and can be turned on and off with a switch. The gyroscope provides foot orientation data, and the microcontroller constantly reads the changes in foot orientation and recognizes various foot gestures. Once a foot gesture is identified, this information is sent to the receiver unit connected to the computer. Fig. 4 shows the list of foot gestures that are recognized by the master unit.

A2. Receiver unit

The receiver is a USB enabled, plug-and-play unit (Fig. 3a), and it consists of two modules: (1) Arduino Leonardo USB Microcontroller, and (2) a Bluetooth Module (HC-05). The circuit diagram of the receiver unit is shown in Fig. A.13. While the receiver unit can execute commands like single click, double click, right click, etc., in our system, irrespective of the gesture identified, a click action is performed at the cursor's position.

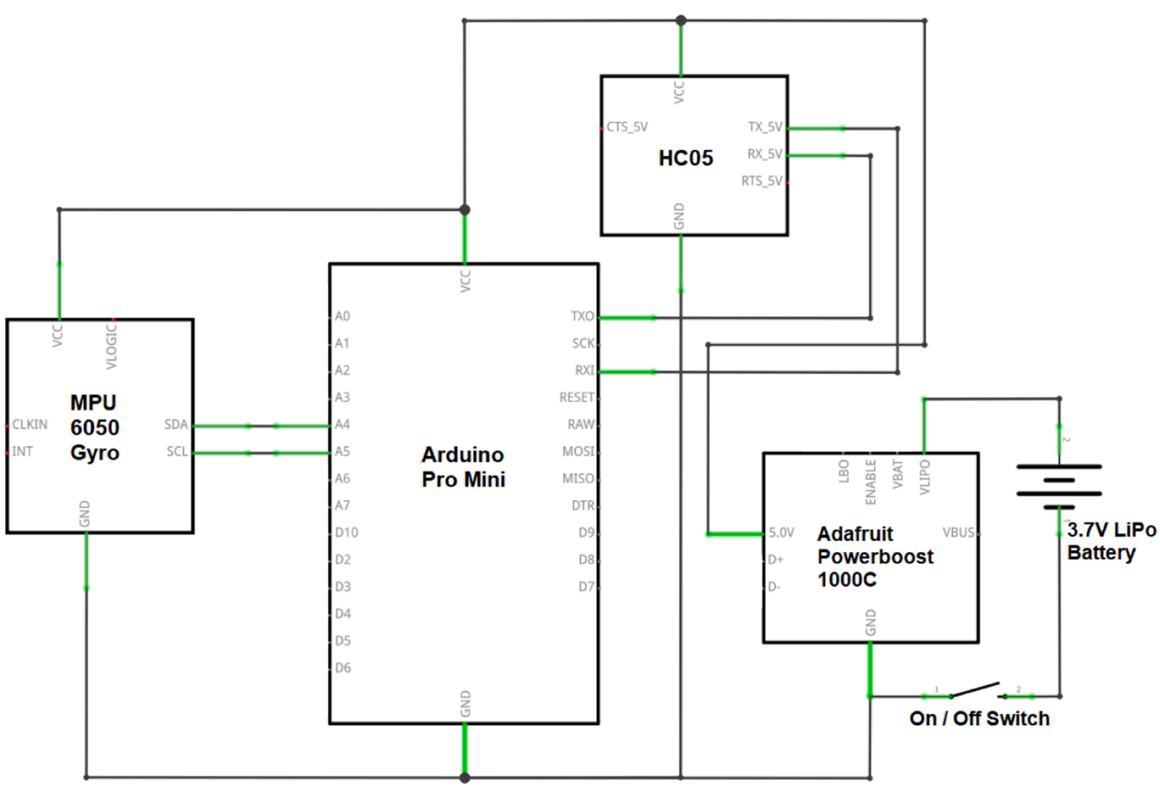


Fig. A1. Foot Gesture Recognition Device - Master unit: circuit diagram showing the four main modules (1) a motion processing unit (MPU-6050) with gyroscope and accelerometer, (2) an Arduino Pro Mini Microcontroller, (3) a Bluetooth Module (HC-05), and (4) a Battery Recharging Unit (Adafruit Powerboost 1000C).

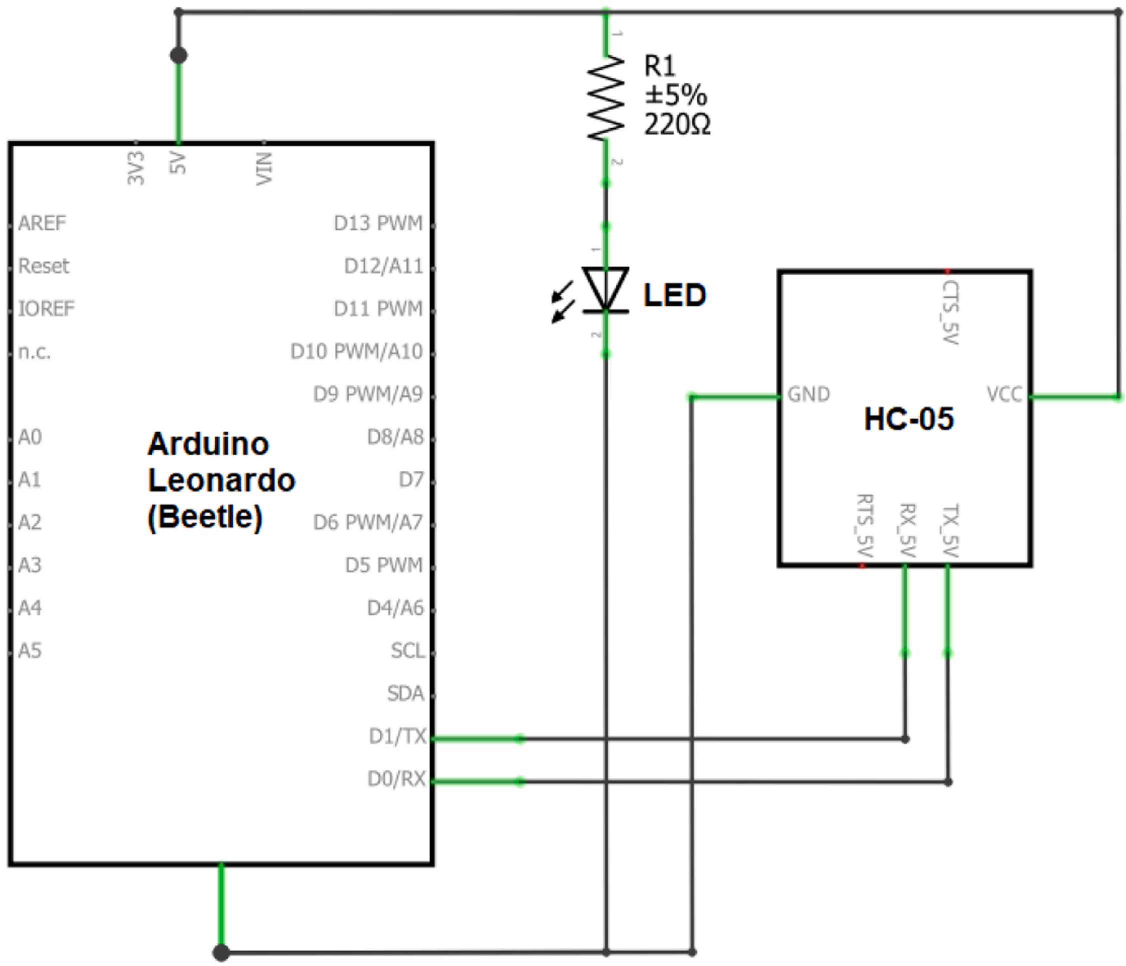


Fig. A2. Foot Gesture Recognition Device - Receiver: circuit diagram showing the two primary modules (1) Arduino Leonardo USB Microcontroller, and (2) a Bluetooth Module (HC-05).

Appendix B. Foot press sensing device - implementation

The circuitry consists of three modules: (1) Teensy 2.0 Microcontroller,[5] (2) Bluetooth Modem (BlueSMiRF),[6] and (3) Force Sensitive Resistor.[7] The circuit diagram is shown in Fig. B.14.

$$V_{out} = V_{in} \cdot \frac{R1}{R1 + R2} \tag{B.1}$$

The minimum amount of pressure to be applied, to be registered as an input action, can be adjusted by setting the output voltage threshold (V_{out}) in Eq. (B.1). In Eq. (B.1), $R1$ and $R2$ are the resistance values, and V_{in} is the input voltage. The user input, based on the pressure thresholds, is encoded as a single byte characters and transmitted to the Gaze Interaction Server via the Bluetooth Modem. The Gazer Interaction Server then decodes the message received into key selection commands.

[5] http://www.pjrc.com
[6] http://www.sparkfun.com/products/12577
[7] http://www.sparkfun.com/products/9376

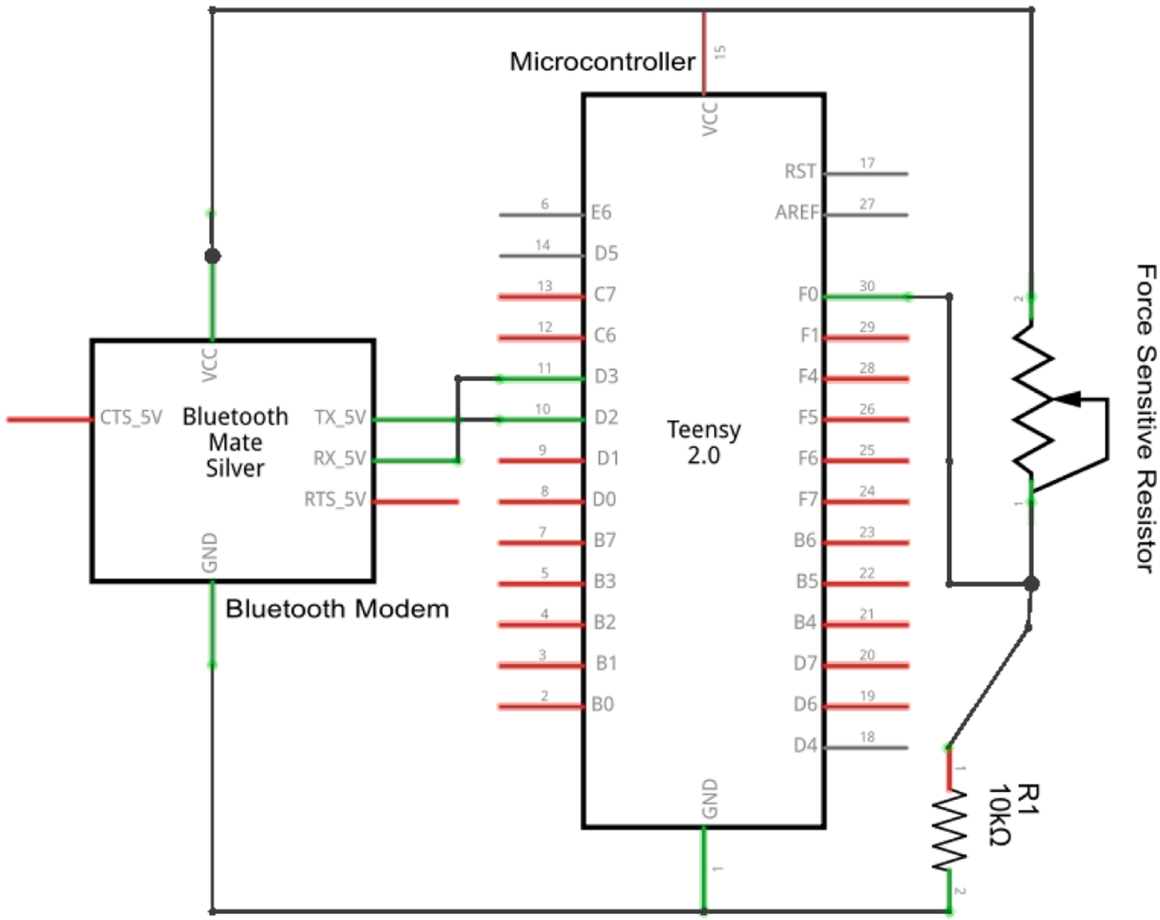


Fig. B1. Foot Press Sensing Device: a circuit diagram showing the three primary modules (1) Teensy 2.0 Microcontroller, (2) Bluetooth Modem (BlueSMiRF), and (3) Force Sensitive Resistor.

References

Bee, N., André, E., 2008. Writing with your eye: a dwell time free writing system adapted to the nature of human eye gaze. Perception in Multimodal Dialogue Systems. Springer, pp. 111–122.

Beelders, T.R., Blignaut, P.J., 2012. Measuring the performance of gaze and speech for text input. Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12. ACM Press, New York, New York, USA, p. 337. https://doi.org/10.1145/2168556.2168631.

Bowman, D.A., Rhoton, C.J., Pinho, M.S., 2002. Text input techniques for immersive virtual environments: an empirical comparison. Proc. Hum. Fact. Ergon.Soc. Annu. Meet. 46 (26), 2154–2158. https://doi.org/10.1177/154193120204602611.

Braza, D.W., Martin, J.N.Y., 2020. Chapter 119 - upper limb amputations. In: Frontera, W.R., Silver, J.K., Rizzo, T.D. (Eds.), Essentials of Physical Medicine and Rehabilitation (Fourth Edition), fourth ed. Elsevier, Philadelphia, pp. 651–657. https://doi.org/10.1016/B978-0-323-54947-9.00119-X.

Brooke, J., et al., 1996. SUS—A quick and dirty usability scale. Usability Eval. Ind. 189 (194), 4–7.

Cecotti, H., 2016. A multimodal gaze-controlled virtual keyboard. IEEE Trans. Hum. Mach. Syst. 46 (4), 601–606. https://doi.org/10.1109/THMS.2016.2537749.

Charmaz, K., Belgrave, L.L., 2007. Grounded Theory. The Blackwell Encyclopedia of Sociology.

Deng, L., Huang, X., 2004. Challenges in adopting speech recognition. Commun. ACM 47 (1), 69–75.

Diaz-Tula, A., Morimoto, C.H., 2016. Augkey: increasing foveal throughput in eye typing with augmented keys. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, p. 35333544. https://doi.org/10.1145/2858036.2858517.

Duchowski, A.T., 2007. Eye tracking methodology. Theory Pract. 328, 614.

Feit, A.M., Williams, S., Toledo, A., Paradiso, A., Kulkarni, H., Kane, S., Morris, M.R., 2017. Toward everyday gaze input: accuracy and precision of eye tracking and implications for design. Proceedings of the 2017 Chi Conference on Human Factors in Computing Systems, pp. 1118–1130.

Fejtová, M., Fejt, J., Štěpánková, O., 2006. Eye as an actuator. Computers Helping People with Special Needs. Springer, pp. 954–961.

Feng, W., Zou, J., Kurauchi, A., Morimoto, C.H., Betke, M., 2021. HGaze Typing: Head-Gesture Assisted Gaze Typing. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3448017.3457379

Frey, L.A., White, K.P., Hutchison, T.E., 1990. Eye-gaze word processing. IEEE Trans. Syst., Man, Cybern. 20 (4), 944–950.

Goel, M., Findlater, L., Wobbrock, J., 2012. Walktype: using accelerometer data to accomodate situational impairments in mobile touch screen text entry. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, p. 26872696. https://doi.org/10.1145/2207676.2208662.

Grauman, K., Betke, M., Lombardi, J., Gips, J., Bradski, G.R., 2003. Communication via eye blinks and eyebrow raises: video-based human-computer interfaces. Univers. Access Inf. Soc. 2 (4), 359–373.

Hammond, T., Davis, R., 2005. Ladder, a sketching language for user interface developers. Comput. Graph. 29 (4), 518–532. https://doi.org/10.1016/j.cag.2005.05.005.

Hansen, D.W., Skovsgaard, H.H.T., Hansen, J.P., Møllenbach, E., 2008. Noise tolerant selection by gaze-controlled pan and zoom in 3D. Proceedings of the 2008 Symposium on Eye Tracking Research & Applications. ACM, New York, NY, USA, pp. 205–212. https://doi.org/10.1145/1344471.1344521.

Hansen, J.P., Hansen, D.W., Johansen, A.S., 2001. Bringing gaze-based interaction back to basics. Universal Access In HCI, pp. 325–328.

Hansen, J.P., Johansen, A.S., Hansen, D.W., Itoh, K., Mashino, S., 2003. Command without a click: dwell time typing by mouse and gaze selections. Proceedings of Human-Computer Interaction–INTERACT, pp. 121–128.

Hansen, J.P., Tørning, K., Johansen, A.S., Itoh, K., Aoki, H., 2004. Gaze typing compared with input by head and hand. Proceedings of the Eye Tracking Research & Applications Symposium on Eye Tracking Research & applications - ETRA'2004. ACM Press, New York, New York, USA, pp. 131–138. https://doi.org/10.1145/968363.968389.

Hansen, J.P., Tørning, K., Johansen, A.S., Itoh, K., Aoki, H., 2004. Gaze typing compared with input by head and hand. Proceedings of the 2004 Symposium on Eye Tracking Research & Applications. ACM, New York, NY, USA, pp. 131–138. https://doi.org/10.1145/968363.968389.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload, Advances in Psychology, vol. 52. North-Holland, pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Hatscher, B., Hansen, C., 2018. Hand, foot or voice: alternative input modalities for touchless interaction in the medical domain. Proceedings of the 20th ACM International Conference on Multimodal Interaction. Association for Computing Machinery, New York, NY, USA, p. 145153. https://doi.org/10.1145/3242969.3242971.
Hatscher, B., Luz, M., Nacke, L.E., Elkmann, N., Müller, V., Hansen, C., 2017. Gazetap: towards hands-free interaction in the operating room. Proceedings of the 19th ACM International Conference on Multimodal Interaction. ACM, New York, NY, USA, pp. 243–251. https://doi.org/10.1145/3136755.3136759.
Hauptmann, A.G., 1995. Speech recognition in the informedia digital video library: uses and limitations. Proceedings of 7th IEEE International Conference on Tools with Artificial Intelligence. IEEE, pp. 288–294.
Hedeshy, R., Kumar, C., Menges, R., Staab, S., 2021. Hummer: Text Entry by Gaze and Hum. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445501
Huang, C., Chen, T., Chang, E., 2004. Accent issues in large vocabulary continuous speech recognition. Int. J. Speech Technol. 7 (2–3), 141–153.
Huckauf, A., Urbina, M., 2007. Gazing with pEYE. Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization - APGV '07. ACM Press, New York, New York, USA, p. 141. https://doi.org/10.1145/1272582.1272618.
Humphries, J.J., Woodland, P.C., Pearce, D., 1996. Using accent-specific pronunciation modelling for robust speech recognition. Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96, vol. 4. IEEE, pp. 2324–2327.
Isokoski, P., 2000. Text input methods for eye trackers using off-screen targets. Proceedings of the 2000 Symposium on Eye Tracking Research & Applications. ACM, pp. 15–21.
Jacob, R.J.K., 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. ACM Trans. Inf. Syst. (TOIS) 9 (2), 152–169.
Kane, S.K., Wobbrock, J.O., Smith, I.E., 2008. Getting off the treadmill: evaluating walking user interfaces for mobile devices in public spaces. Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services. ACM, New York, NY, USA, pp. 109–118. https://doi.org/10.1145/1409240.1409253.
Kat, L.W., Fung, P., 1999. Fast accent identification and accented speech recognition. 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258), vol. 1. IEEE, pp. 221–224.
ten Kate, J.H., Frietman, E.E., Stoel, F.J., Willems, W., 1980. Eye-controlled communication aids. Med. Prog. Technol. 8 (1), 1–21.
Klamka, K., Siegel, A., Vogt, S., Göbel, F., Stellmach, S., Dachselt, R., 2015. Look & pedal: hands-free navigation in zoomable information spaces through gaze-supported foot input. Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, New York, NY, USA, pp. 123–130. https://doi.org/10.1145/2818346.2820751.
Koester, H.H., Levine, S.P., 1994. Modeling the speed of text entry with a word prediction interface. IEEE Trans. Rehabil. Eng. 2 (3), 177–187.
Kumar, C., Hedeshy, R., MacKenzie, I.S., Staab, S., 2020. Tagswipe: touch assisted gaze swipe for text entry. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, p. 112. https://doi.org/10.1145/3313831.3376317.
Kumar, M., Paepcke, A., Winograd, T., 2007. Eyepoint: practical pointing and selection using gaze and keyboard. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, p. 421430. https://doi.org/10.1145/1240624.1240692.
Kurauchi, A., Feng, W., Joshi, A., Morimoto, C., Betke, M., 2016. Eyeswipe: dwell-free text entry using gaze paths. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, p. 19521956. https://doi.org/10.1145/2858036.2858335.
MacKenzie, I.S., Soukoreff, R.W., 2003. Phrase sets for evaluating text entry techniques. CHI'03 Extended Abstracts on Human Factors in Computing Systems. ACM, pp. 754–755.
MacKenzie, I.S., Zhang, X., 2008. Eye typing using word and letter prediction and a fixation algorithm. Proceedings of the 2008 Symposium on Eye Tracking Research & Applications - ETRA '08. ACM Press, New York, New York, USA, p. 55. https://doi.org/10.1145/1344471.1344484.
Majaranta, P., 2011. Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies: Advances in Assistive Technologies. IGI Global.
Majaranta, P., Ahola, U.-K., Špakov, O., 2009. Fast gaze typing with an adjustable dwell time. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 357–360. https://doi.org/10.1145/1518701.1518758.
Majaranta, P., MacKenzie, I.S., Aula, A., Räihä, K.-J., 2003. Auditory and visual feedback during eye typing. CHI '03 Extended Abstracts on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 766–767. https://doi.org/10.1145/765891.765979.
Majaranta, P., Räihä, K.-J., 2002. Twenty years of eye typing: systems and design issues. Proceedings of the 2002 Symposium on Eye Tracking Research & Applications. ACM, New York, NY, USA, pp. 15–22. https://doi.org/10.1145/507072.507076.
Majaranta, P., Räihä, K.-J., 2007. Text entry by gaze: utilizing eye-tracking. Text Entry Systems: Mobility, Accessibility, Universality, pp. 175–187.
Meena, Y.K., Cecotti, H., Wong-Lin, K., Prasad, G., 2016. A novel multimodal gaze-controlled hindi virtual keyboard for disabled users. 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 003688–003693. https://doi.org/10.1109/SMC.2016.7844807.
Mott, M.E., Williams, S., Wobbrock, J.O., Morris, M.R., 2017. Improving dwell-based gaze typing with dynamic, cascading dwell times. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, p. 25582570. https://doi.org/10.1145/3025453.3025517.
Parker, S.E., Mai, C.T., Canfield, M.A., Rickard, R., Wang, Y., Meyer, R.E., Anderson, P., Mason, C.A., Collins, J.S., Kirby, R.S., et al., 2010. Updated national birth prevalence estimates for selected birth defects in the United States, 2004–2006. Birth Defects Res. Part A 88 (12), 1008–1016.
Pedrosa, D., Pimentel, M.D.G., Wright, A., Truong, K.N., 2015. Filteryedping: design challenges and user performance of dwell-free eye typing. ACM Trans. Access Comput. 6 (1), 1–37. https://doi.org/10.1145/2724728.
Porta, M., Turina, M., 2008. Eye-S: a full-screen input modality for pure eye-based communication. Proceedings of the 2008 Symposium on Eye Tracking Research & Applications. ACM, New York, NY, USA, pp. 27–34. https://doi.org/10.1145/1344471.1344477.
Qian, H., Kuber, R., Sears, A., 2013. Tactile notifications for ambulatory users. CHI '13 Extended Abstracts on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 1569–1574. https://doi.org/10.1145/2468356.2468637.
Räihä, K.-J., Ovaska, S., 2012. An exploratory study of eye typing fundamentals: dwell time, text entry rate, errors, and workload. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, p. 30013010. https://doi.org/10.1145/2207676.2208711.
Rajanna, V., 2016. Gaze typing through foot-operated wearable device. Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility. ACM, New York, NY, USA, pp. 345–346. https://doi.org/10.1145/2982142.2982145.
Rajanna, V., Hammond, T., 2016. Gawschi: gaze-augmented, wearable-supplemented computer-human interaction. Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. ACM, New York, NY, USA, pp. 233–236. https://doi.org/10.1145/2857491.2857499.
Rajanna, V., Hansen, J.P., 2018. Gaze typing in virtual reality: impact of keyboard design, selection method, and motion. Proceedings of the Tenth Biennial ACM Symposium on Eye Tracking Research and Applications. ACM, New York, NY, USA. https://doi.org/10.1145/3204493.3204541.
Rajanna, V., Polsley, S., Taele, P., Hammond, T., 2017. A gaze gesture-based user authentication system to counter shoulder-surfing attacks. Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, New York, NY, USA, pp. 1978–1986. https://doi.org/10.1145/3027063.3053070.
Rajanna, V.D., 2016. Gaze and foot input: toward a rich and assistive interaction modality. Companion Publication of the 21st International Conference on Intelligent User Interfaces. ACM, New York, NY, USA, pp. 126–129. https://doi.org/10.1145/2876456.2876462.
Sarsenbayeva, Z., van Berkel, N., Luo, C., Kostakos, V., Goncalves, J., 2017. Challenges of situational impairments during interaction with mobile devices. Proceedings of the 29th Australian Conference on Computer-Human Interaction. Association for Computing Machinery, New York, NY, USA, p. 477481. https://doi.org/10.1145/3152771.3156161.
Schildbach, B., Rukzio, E., 2010. Investigating selection and reading performance on a mobile phone while walking. Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services. ACM, New York, NY, USA, pp. 93–102. https://doi.org/10.1145/1851600.1851619.
Tuisku, O., Rantanen, V., Surakka, V., 2016. Longitudinal study on text entry by gazing and smiling. Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. Association for Computing Machinery, New York, NY, USA, p. 253256. https://doi.org/10.1145/2857491.2857501.
Urbina, M.H., Huckauf, A., 2010. Alternatives to single character entry and dwell time selection on eye typing. Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10. ACM Press, New York, New York, USA, p. 315. https://doi.org/10.1145/1743666.1743738.
Velloso, E., Schmidt, D., Alexander, J., Gellersen, H., Bulling, A., 2015. The feet in human–computer interaction: a survey of foot-based interaction. ACM Comput. Surv. 48 (2), 21:1–21:35. https://doi.org/10.1145/2816455.
Ward, D.J., Blackwell, A.F., MacKay, D.J.C., 2000. Dashera data entry interface using continuous gestures and language models. Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology. Association for Computing Machinery, New York, NY, USA, p. 129137. https://doi.org/10.1145/354401.354427.
Wobbrock, J.O., 2006. A robust design for accessible text entry. SIGACCESS Access. Comput. (84), 4851. https://doi.org/10.1145/1127564.1127575.
Wobbrock, J.O., Rubinstein, J., Sawyer, M.W., Duchowski, A.T., 2008. Longitudinal evaluation of discrete consecutive gaze gestures for text entry. Proceedings of the 2008 Symposium on Eye Tracking Research & Applications. ACM, New York, NY, USA, pp. 11–18. https://doi.org/10.1145/1344471.1344475.
Zhao, X.A., Guestrin, E.D., Sayenko, D., Simpson, T., Gauthier, M., Popovic, M.R., 2012. Typing with eye-gaze and tooth-clicks. Proceedings of the Symposium on Eye Tracking Research and Applications. ACM, pp. 341–344.
Ziegler-Graham, K., MacKenzie, E.J., Ephraim, P.L., Travison, T.G., Brookmeyer, R., 2008. Estimating the prevalence of limb loss in the United States: 2005 to 2050. Arch. Phys. Med. Rehabil. 89 (3), 422–429. https://doi.org/10.1016/j.apmr.2007.11.005.