This is the Accepted Manuscript of the following article published by in the International Journal of Human–Computer Interaction [24. July 2022]:

Zhe Zeng, Elisabeth Sumithra Neuer, Matthias Roetting, & Felix Wilhelm Siebert (2022) A One-Point Calibration Design for Hybrid Eye Typing Interface, International Journal of Human–Computer Interaction, 1-14. DOI: 10.1080/10447318.2022.2101186.

ACCEPTED MANUSCRIPT

A One-Point Calibration Design for Hybrid Eye Typing Interface

Zhe Zeng[a], Elisabeth Sumithra Neuer[a], Matthias Roetting[a] and Felix Wilhelm Siebert[b]

[a]Chair of Human-Machine Systems, Technical University of Berlin, Germany [b]Department of Technology, Management and Economics, Technical University of Denmark, Denmark

ARTICLE HISTORY
Compiled July 29, 2022

ABSTRACT
We present an eye typing interface with one-point calibration, which is a two-stage design. The characters are clustered in groups of four characters. Users select a cluster by gazing at it in the first stage, and then select the desired character by following its movement in the second stage. A user study was conducted to explore the impact of auditory and visual feedback on typing performance and user experience of this novel interface. Results show that participants can quickly learn how to use the system, and an average typing speed of 4.7 WPM can be reached without lengthy training. The subjective data of participants revealed that users preferred visual feedback over auditory feedback while using the interface. The user study indicates that this eye typing interface can be used for walk-up-and-use interactions, as it is easily understood and robust to eye-tracking inaccuracies. Potential areas of application, as well as possibilities for further improvements, are discussed.

KEYWORDS
eye movement; eye tracking; gaze interaction; eye typing; calibration

1. Introduction

With the development of eye tracking technology, gaze interaction has the potential to become a ubiquitous part of assisting our daily interaction. Gaze input makes it possible to combine both visual search and action activation, into one step. Users can perceive relevant information and activate desired commands, e. g., by focusing their gaze on a corresponding item on the interface. The interest in gaze-based interaction design is growing rapidly. The applications such as eye typing (Majaranta & Räihä, 2002), gaze-controlled web browsing (Menges, Kumar, Müller, & Sengupta, 2017), wheelchair control (Eid, Giakoumidis, & El Saddik, 2016), and gaze selection in virtual reality (Blattgerste, Renner, & Pfeiffer, 2018), have been developed and studied.

As an input modality, gaze has a number of attractive features. Firstly, gaze-based input enables a system to be more accessible. It can help people with physical limitations to communicate with each other. Not only people with permanent disabilities, such as amyotrophic lateral sclerosis (ALS) can benefit from this (Hwang, Weng, Wang, Tsai, & Chang, 2014), but also people with a temporary restriction for physical activity, users with an arm injury, or users that are carrying objects. Additionally, there is

CONTACT Zhe Zeng. Email: zengzhefy@gmail.com

increasing concern about hand hygiene since the COVID-19 pandemic. People are encouraged to use contactless interactions to minimize, and if possible to avoid contacts between users' hands and displays. Even though there are many contactless interfaces using gesture or voice, the gaze is a potential complementary input modality to provide a more hygienic interaction when using public displays/interfaces. In addition, without bystanders hearing voice commands, gaze interaction has a high degree of privacy (Katsini, Abdrabou, Raptis, Khamis, & Alt, 2020). As an example, PIN code input using gaze is much more inconspicuous than voice or gesture control in a public setting.

Although there is increasing interest in gaze-based interaction, proposed implementations have suffered from a number of limitations. The so-called Midas touch problem represents one of the main challenges for gaze-based interaction (Jacob, 1990). It describes the difficulty to distinguish between visual search patterns and the intentional selection of actionable items on an interface. Another challenge in the design of gaze-based interfaces is the lack of accuracy of real-time gaze coordinates, stemming from both biological characteristics of the eye (Robinson, 1968) as well as limitations of the eye tracking equipment (Feit et al., 2017). Due to miniature eye movements, such as tremor, drift, and microsaccades, there is hardly a moment that the eye is absolutely still (Duchowski, 2017). Thus, the use of the gaze as a cursor cannot be as accurate as e. g., using a computer mouse. Besides, most eye trackers require calibration before use, which is time-consuming and can be inconvenient for spontaneous interaction. Even with individual calibration, eye trackers are still prone to problems with fluctuating spatial accuracy during use (Feit et al., 2017). They found that eye tracking systems need to be re-calibrated multiple times a day to maintain interactable accuracy for most eye typing systems, and the individual difference is relatively large both in accuracy and precision as well.

To address the Midas touch problem, dwell-time was introduced to separate searching and selecting, i. e., to activate an action, user need to gaze at one item for a predefined time. However, dwell-based gaze interaction requires calibrating the eye tracker before using it. To overcome the calibration challenge, pursuit-based interfaces have been proposed by Vidal, Bulling, and Gellersen (2013). Using these interfaces, users can activate an action by following the corresponding moving target with their gaze. An action is activated based on the fitting of the relative motion trajectory, instead of fixed gaze coordinates on the interface, which is why a personal calibration is not necessary for pursuit-based interaction. Hence, this calibration-free method has been attracting a lot of interest, as it overcomes the shortcomings arising from low spatial accuracy of gaze data. Researchers have developed dynamic gaze interfaces based on pursuit movements and developed different applications, such as entering PIN code (Almoctar, Irani, Peysakhovich, & Hurter, 2018; Cymek et al., 2014), controlling smart watches (Esteves, Velloso, Bulling, & Gellersen, 2015), and selection in VR device (Sidenmark, Clarke, Zhang, Phu, & Gellersen, 2020).

However, those dynamic interfaces differ greatly from the interfaces we use every day. Users need to search for targets among moving objects. Ludvigh and Miller (1958) reported that visual acuity decreases significantly with the increase in angular velocity of the moving object. The search task is more difficult for the complex interfaces, such as eye typing interface (Lutz, Venjakob, & Ruff, 2015), where many characters move at the same time. Thus, in this study, we propose to distinguish between dynamic and static areas according to the location of attention. Since people generally look at the interactive item before activating it, and the interactive content is generally located in the attention area. In our design, only objects within the attention area are dynamic.

Other objects in the static area stay still to minimize distractions. A novel eye typing interface is designed based on this concept, to create a robust gaze-based text entry system. In this system, instead of a time-consuming personal calibration, the one-point calibration from Lutz et al. (2015) is used, which aligns eye coordinates based on the offset between screen midpoint and collected midpoint of eye tracker. It simplifies the calibration process while ensuring robust interaction.

Further, we are interested in providing feedback to achieve more efficient typing performance and better user experience. In the eye typing system, the information input and output share the visual channel. We would like to address the question, whether visual feedback is advantageous, as it provides more direct feedack, or if it is disadvantageous, as it could potentially lead to a higher mental workload. At the same time, the feedback from other channels, e. g., auditory, could be more efficient, or combined modalities could present a better solution for feedback on a dynamic interface. Thus, to investigate the effect of feedback, different modalities (visual only, auditory only, combined visual and auditory and no feedback) are compared in a user study.

2. Related Work

Eye tracking technology is used in psychology, marketing, and also as an input device for human-computer interaction. For gaze interaction, eye typing is a common application. Using gaze for text entry has several unique benefits, and was initially developed as an important communication tool for people with certain disabilities (Majaranta & Räihä, 2002). Furthermore, systems have been developed to enter text using gaze for public displays Lutz et al. (2015), and recently have been widely studied as an input method in virtual reality (Rajanna & Hansen, 2018). Dwell time, gestures, or pursuit movements are used to replace the click/tap action to activate a character. The layout of common eye typing interfaces can be divided into two main types: traditional QWERTY layouts and other specific layouts.

2.1. QWERTY Typing Layout

The QWERTY keyboard layout is familiar to most users, hence it is easy to learn the interaction process for gaze-based typing on QWERTY interfaces. There have been a number of studies involving the implementation of QWERTY layout in eye typing.

2.1.1. Dwell-based Typing Interface

For most eye typing systems, the input and output space share the same space on the interface and the gaze works as an invisible cursor. Designs based on dwell time are popular, as they allow control over the Midas touch problem (Jacob, 1990). That is, users can enter a letter by fixating their gaze on it for a pre-defined duration, i.e. the *dwell time*. Research has shown the duration of dwell time is a key parameter to avoid Midas touch, and dwell time is varied from 200 to 1000 ms (Majaranta, MacKenzie, Aula, & Räihä, 2006; Penkar, Lutteroth, & Weber, 2012). To further improve typing efficiency, individual differences can be taken into account and more flexible dwell duration were investigated, e.g. users were allowed to adjust the dwell time themselves (Majaranta, Ahola, & Špakov, 2009), systems were capable of automatically adjusting the dwell time based on users' performance (Špakov & Miniotas, 2004), or previously entered text and location of keys (Mott, Williams, Wobbrock, & Morris, 2017; Pi, Koljonen, Hu, & Shi, 2020). However, dwell-time based gaze system depends heavily on the spatial accuracy of the eye tracker, as even slight inaccuracies in tracking can lead to an unintentional input for characters close to the intended target.

2.1.2. Dwell-free Typing Interface

In dwell-free interfaces, users do not need to fixate their gaze on actionable items for a pre-defined dwell time. For example, on the Context-switching interface (Morimoto & Amir, 2010), two QWERTY keyboards are displayed on the screen, and the user swipes between them to enter text. In addition to character-level text entry method, there is a word-level text entry method based on the analysis of letter sequence using a language model. Users can enter text by looking at the letters they need to enter in order with their eyes (Kurauchi, Feng, Joshi, Morimoto, & Betke, 2016; Pedrosa, Pimentel, Wright, & Truong, 2015). When starting and ending a word, an explicit command needs to be given out to distinguish between receiving visual information and selection.

The traditional QWERTY layout has both advantages and disadvantages for gaze-based interfaces. Since it is well known by most users, the QWERTY layout can reduce the cost of learning. However, it was developed for finger typing and is not well adapted for gaze-based typing.

2.2. Specific Typing Interface

In addition to the traditional layout, there are a number of novel keyboard layouts designed specifically for eye typing. Due to the unstable accuracy of gaze data, those layouts try to rearrange the letter positions or cluster letters to facilitate the user experience.

2.2.1. Static-graphic Gaze Interface

J. P. Hansen, Johansen, Hansen, Itoh, and Mashino (2003) developed *GazeTalk*, which has 12 active buttons on the screen. The interface is updated in real-time and what is most likely to be entered is shown. The characters or words displayed in each button are based on the predictions of a language model. To overcome the limitations of inaccurate gaze data, the size of actionable items is increased, which helps with item distinction, but reduces the number of displayable actionable items or results in the need for bigger displays. Huckauf and Urbina (2008) proposed the concept of two-stage selection. Their *pEYEs* application takes the form of popup menus, this pie layout is available for menu selection and text entry. Users first select a cluster and then select the intended character from the selected cluster. In this interface, the displayed characters and the area for activation are separated. The characters are located near the center of the circle and the activation area is located at the edge of the circle.

Additionally, research has been carried out on gaze gestures to overcome space limitations. Gaze gesture systems such as *Quikwriting* (Perlin, 1998) or *EyeWrite* (Wobbrock, Rubinstein, Sawyer, & Duchowski, 2008) require users to spend a certain amount of time learning and remember the gestures. Hence they are not suitable for walk-up-and-use scenarios.

2.2.2. Dynamic-graphic Gaze Interface

A number of studies have investigated the use of a dynamic interface for gaze typing. In those eye typing interface, the items will move during selection.

One of the well-known eye typing systems is *Dasher* (Ward & MacKay, 2002). In this interface, the characters with high probability are placed in a column based on the probability of the next possible character. The characters in a column are continuously moving from right to left. The user needs to keep looking at the moving desired character. When the character being viewed crosses the midpoint of the screen, the character is selected. As an efficient typing system, *Dasher*'s text entry rate can be up to 25 words per minute (WPM). There was one more dynamic eye typing interface, called *StarGazer* (D. W. Hansen, Skovsgaard, Hansen, & Møllenbach, 2008), where user could select characters with zooming and panning actions.

However, all the typing interfaces mentioned above need to be individually calibrated before they can be used. More recently, studies are dedicated to designing interactive interfaces with reduced calibration time or without calibration. Lutz et al. (2015) developed the text entry system (*SMOOVS*), a pursuit-based gaze interface with one-point calibration. To avoid the Midas touch problem, the interface is divided into interactable areas and a central area that is not interactable. All elements in the interface are designed to stay static when the gaze cursor is located in the non-interactable middle area of the interface. When the gaze position is moving out of the deactivated central area, the first moving stage starts. In this stage, all clusters move outward simultaneously. In a second moving stage, the tiles within the selected cluster move apart from each other around the center of the cluster. Although users can quickly understand the interface and learn how to use it, there are multiple interruptions when completing a selection. Lutz et al. (2015) reported when entering a given sentence of 59 characters in length, users had to discontinue at least 31.33 times on average at the condition of 300 px/s. In addition, the text entry rate for *SMOOVS* ranged from 2.9-3.34 WPM, which is relatively low in comparison to other typing systems with personal calibration. In order to reduce the amount of keystrokes for entering a word or sentence, a word prediction function was added to the *SMOOVS* interface to improve the typing speed (Zeng & Roetting, 2018).

In addition to eye typing interface based on one-point calibration, there are three more calibration-free eye typing interfaces (Abdrabou, Mostafa, Khamis, & Elmougy, 2019; Bafna, Bækgaard, & Paulin Hansen, 2021; Porta, Dondi, Pianetta, & Cantoni, 2021). Abdrabou et al. (2019) attempt to redesign the SMOOVS interface in a circular motion. In *EyeTell*, there are two circles, an inner circle and an outer circle, they move in a clockwise or counter-clockwise circular motion (Bafna et al., 2021). Unlike the above interface, in *SPEye*, the characters/character clusters are stationary, but each character/character cluster has small corresponding dot moving between the midpoint and the character/character cluster. The user makes a selection by following the corresponding dots. However, the typing performance and user experience still remain to be improved to meet the needs of daily life.

2.3. Feedback in Gaze Interaction

In addition to differences in the basic workings of gaze-based spellers, the design of appropriate feedback can enhance users' experience (Nielsen, 1994). First, feedback can provide information about the state of the interaction, thereby allowing the user to adjust any erroneous activation before it is registered. Second, feedback can be used to confirm that a selection is registered. Generally, users are relatively more tolerant of interfaces with appropriate feedback. However, inappropriate feedback can also be misleading or distracting to users.

2.3.1. Feedback in Dwell-based Gaze Interface

Majaranta, MacKenzie, Aula, and Räihä (2003) found an effect of feedback on typing performance of the dwell-based system. Participants typed faster when receiving a combined click and visual feedback than other forms of feedback such as speech only or visual only. Short no-speech sound, like a 'click', was preferred by participants over synthetic speech. The speech feedback is limited in some cases, as it takes more time to pronounce a selected letter than just giving a short, no-speech sound as feedback. In addition, Majaranta et al. (2006) found that the feedback needs to be set according to the duration of dwell time.

2.3.2. Feedback in Gaze Gestures

Since the interaction via gaze gestures is usually facilitated without a graphical user interface, the related studies focus more on vibrotactile feedback (Rantala et al., 2020). It was found that the implementation of vibrotactile feedback can reduce response time as well as improve the user's subjective evaluation (Kangas et al., 2014). Köpsel, Majaranta, Isokoski, and Huckauf (2016) compared visual, haptic, and auditory feedback modalities. They evaluated feedback given both during and after the input of a gaze gesture. The results show that for gaze gesture interactions, the evaluation of the feedback modalities was comparable in terms of task completion time, error rate, and user experience.

2.3.3. Feedback in Pursuit-based Gaze Interface

The influence of feedback has also been considered in research on pursuit-based gaze interaction. Špakov, Isokoski, Kangas, Akkil, and Majaranta (2016) added a "tick" tone in smooth pursuit-based widgets. The comparison of feedback modalities (no feedback, visual, auditory, and haptics) showed that feedback conditions do not significantly affect the performance of pursuits-based gaze interaction. However, most users preferred tactile and auditory feedback (Kangas et al., 2016). When interacting with dynamic interfaces, the cognitive workload is higher than for interactions with static interfaces, and suitable feedback could be considered in the design of dynamic interfaces. Even though pursuit-based gaze interaction has been studied for years, there is less research on how to design the feedback of the interaction to users.

3. Hybrid Gaze Interface Design

In this study, a hybrid gaze interface, combining dwell-time and pursuit movement, is designed with the goal to facilitate an interaction that is perceived as natural as possible by users.

3.1. Layout

The text entry interface is based on an octagon-like layout with eight clusters that allow users to input different characters (see Figure 1, left). As Lutz et al. (2015)

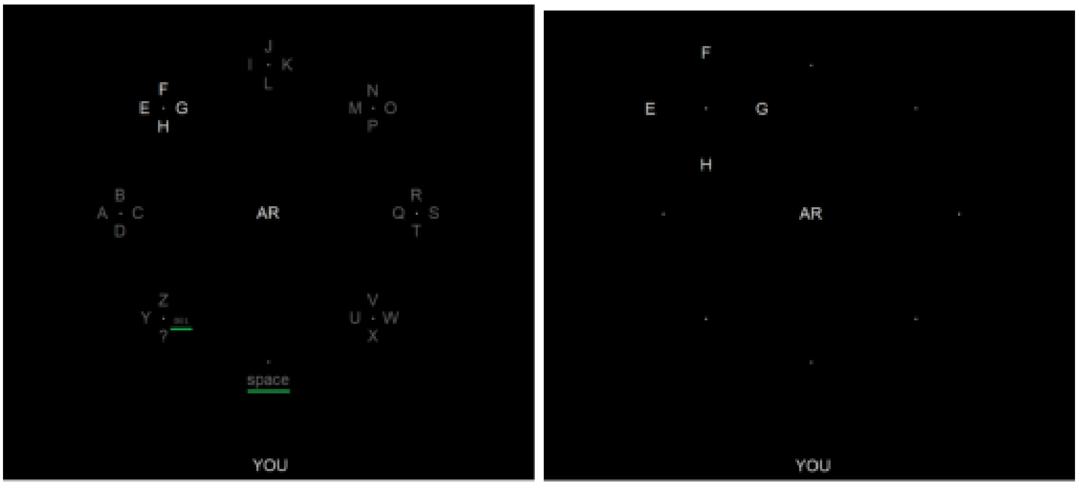

Figure 1.: The hybrid eye typing interface. When user try to enter the phrase "you are a wonderful example". The typed word ("YOU") is located at the bottom of the screen. The current typed characters ("AR") are displayed in the center of the deactivated area. To Enter E, the eyes first start looking at the EFGH cluster, and this cluster is highlighted (left). The characters in the selected cluster moved outward in pursuing stage. Eyes need to follow the movement of letter E to enter it (right).

reported, the accuracy of gaze data decreased significantly from the middle to the edge of the screen after one-point calibration, i. e., the interactions in the central area of the interface are registered with higher accuracy. Hence, a larger number of objects can be distinguished in the first stage of detection which is placed more central in the interface, than in the second stage which is placed on the outside of the interface. In addition, Zeng, Siebert, Venjakob, and Roetting (2020) found that the circle-layout interfaces containing 6 and 8 linear moving objects achieved good detection rate. Therefore, in this study, eight clusters are designed in the first stage. There are four character tiles in each cluster. Regarding the direction of movement, studies have shown that the detection accuracy for smooth pursuit eye movements is influenced by the movement direction. Horizontal and vertical directions are more robustly detected than diagonal directions (Ke, Lam, Pai, & Spering, 2013; Krukowski & Stone, 2005). Thus, in order to achieve a more robust identification of gaze directions, characters in each cluster are designed to move along the horizontal and vertical axis, e. g., in cluster "ABCD", A moves left, B moves up, C moves right, and D moves down. In this study, 1° visual angle corresponds to 39 pixels at a distance of 60 cm from the user to the screen.

The alphabetical A-Z layout was used in this study, which contains 26 letters, as well as delete, space key, and question mark. The space and delete keys are underlined in green only for visualization (see Figure 1, left). As the space key is the most frequently used key (ETA, 2020), it is located in a vertically downward direction of the interface as a single actionable item (not in a cluster). While the key is mainly used to enter a space between words, it also serves as a way to confirm the currently entered word and indicate the start of an entry for a new word. Once the space key is used, typed words are moved from the center to the bottom of the screen (where the word "YOU" is located in Figure 1). The delete key is used for correcting typed characters.

3.2. Interaction Design

Stage 0 Inactive. When a user looks at the central area of the screen, e. g., when reading the typed characters, the interface remains static. When a user's gaze is registered outside of the inactive central area, the detection of cluster selection is activated.
Stage 1 Searching and cluster selection. In this stage, all clusters remain static. The user needs to find the cluster which includes the desired character. The cluster in which the gaze is registered is highlighted (see example "EFGH" cluster in Figure 1, left). However, if the user's gaze moves out of the area of this cluster, the highlighting is reversed. If the gaze moves to another cluster, that new cluster is highlighted, allowing users to adjust their selection before the activation of an input.
Stage 2 Character selection. After a cluster is selected, the second stage begins (for more details on the classification algorithm, see 3.4). In this stage, the characters in the selected cluster move outward. Other clusters that were not selected fade out gradually at the same time to avoid visual distraction. That is, the active cluster is highlighted and the entailed characters move outward, while the rest of the clusters fades out. The typing interface at the end of stage 2 is shown in Figure 1 (right). The moving speed of characters is 433 pixels/s (approximately 10°/s). After the characters in the selected cluster have moved 75 pixels, they fade out and then reappear in their original position (i. e., where they were before they moved). If a character is successfully detected, feedback is given simultaneously. Detailed information can be found in the subsection 3.3.
Stage 3 Return to initial position. During stages 1 and 2, if the user looks at another cluster, the characters that are not watched will automatically return to their original position.

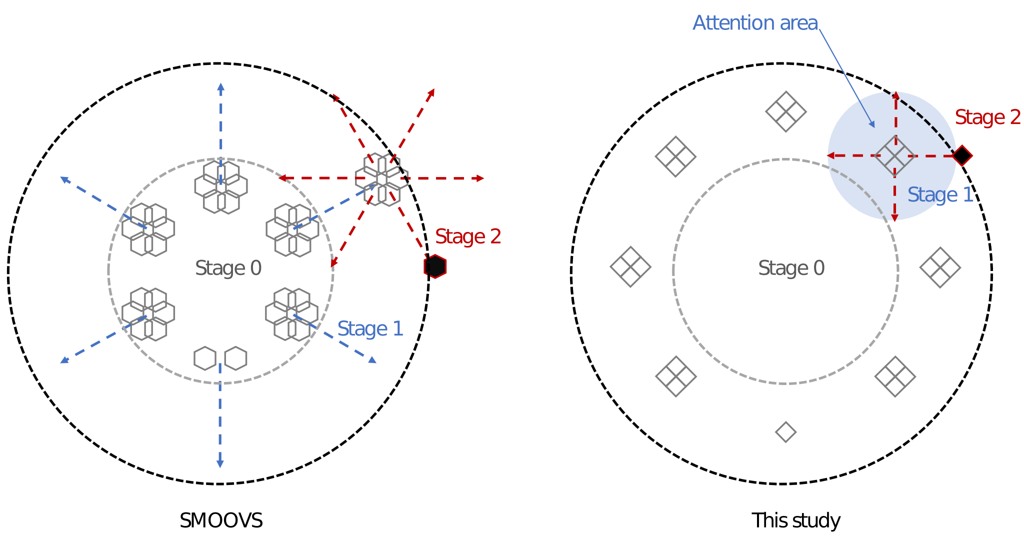

Figure 2.: Interaction stages of SMOOVS and this study. The blue dashed arrows represent the stage 1 of the moving trajectory, and the red dashed arrows represent the stage 2 of the moving movement. The blue translucent circular area represents the position of the eye's attention.

In SMOOVS, when the gaze position moves out of the center idle region, all clusters

Table 1.: Comparison of SMOOVS and this study.

	SMOOVS	This study
Stage 0	Inactive	Inactive
Stage 1	All clusters move outward.	Interface remains static, highlights the cluster being looked at.
	No feedback	Real-time feedback on the cluster being looked at.
Stage 2	All cluster move outward. Selected clusters are clearly visible, other clusters are hardly visible.	Only the characters of the selected cluster move outward, other clusters fade out.
Stage 3	After completing stage 1 and 2, all character return to their original position while guiding user's eyes to the middle idle area to start the next input.	User can change the selected cluster at any time, and the clusters that is not in the attention area automatically returns to its original position.

move at the same time. This study constrains the moving items in the interface. Only items in the attention area move, and the user follows the movement of one desired object in the attention area for selection (see Figure 2). This further reduces the distractions associated with dynamic interfaces. Moreover, it also brings real-time feedback on which cluster has been watched/selected. Compared to *SMOOVS*, the user can change the cluster to follow at any time without returning to the center point to start a new round of selection in this study, potentially leading to a smoother user experience. In addition to the difference in the selection procedure, the interface is also reduced from six components to four in each cluster, making it less difficult for visual search in each group. A more detailed comparison for each stage can be found in Table 1.

3.3. Feedback Design

Feedback was given to confirm that the system is responding to the input, visualize reactions, and confirm an activated action. The feedback factor consists of four levels, namely no feedback, visual feedback, auditory feedback, and both visual and auditory feedback.

Visual feedback appears behind the selected character in a gray circle (the diameter of the circle is 44 pixels, approximately 1°). Auditory feedback was given via computer speakers. The auditory feedback sound was a short continuous beep at 300 Hz. The visual and auditory feedback is synchronized and lasts approximately 200 milliseconds (ms). To avoid effects of sequence caused by fatigue and practice, the order of conditions was randomized.

3.4. Classification Algorithm

In the first stage, the detection of a cluster is based on the midpoint of screen (M), gaze point (G) and the positive direction of the x-axis (as shown in Figure 3). The distance from screen midpoint M to each group midpoint is 300 pixels (7.69° visual angle). The angular distance between the cluster centroids is 5.89° visual angle. The angle θ between the vector between two points and the positive x-direction is given by Function 1 and converted from radian to degree .

$$\theta_g = \arctan 2\,(y_g - y_m, x_g - x_m) \tag{1}$$

where y_g and x_g are the coordinates of the eye position, y_m and x_m are the midpoint coordinates of clusters. The detectable range for each cluster is 45° in first stage. When the angle meets the angular criterion of one cluster $\theta_l < \theta_g < \theta_u$, it is recognized as the corresponding cluster. For example, if $22.5° < \theta_g < 67°$, the cluster "UVWX" is detected as the looked cluster.

When there are more than two successive measurements detected as the same cluster, i. e., $\theta_{g,i} = \theta_{g,i+1}$, this cluster is highlighted to inform the user about the current selection. The general fixation typically lasts 200-600 ms (Jacob, 1995). The dwell time for cluster selection was set as 400 ms, i. e., when this cluster is further detected for more than 400 ms, then this cluster is activated and the second stage (dynamic stage) starts.

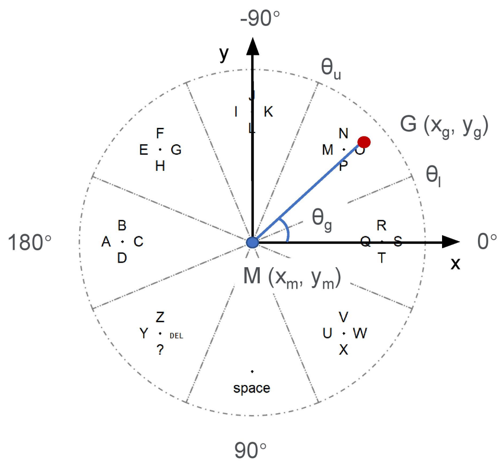

Figure 3.: Visualize the angular criterion in first stage. The red dot is gaze point (G), and blue dot is the midpoint of screen (M). The gray dotted line presents the identifiable range of each cluster.

In the second stage, only characters in the selected cluster begin to move. Users can select the desired character by following the movement of this character with their eyes. Similar to the first stage, character detection is also based on angular range.

After the characters of the selected group have finished moving, 20 gaze points $[(x_{g,1}, y_{g,1}), ..., (x_{g,20}, y_{g,20})]$ will continue to be collected. A set of angles $[\theta_{g,1}, ..., \theta_{g,20}]$ is calculated according to those 20 gaze points and the middle point of this cluster. The *mode* of those angles is calculated and the character corresponding to this angle is selected.

The detectable range of a character for this stage is 85 degrees. In order to reduce false detections, an area of five degrees is defined between the adjacent characters. If the measurement falls into this area, the system skips this input. But the following action itself does not affect the detection. When for two consecutive gaze points the distance between gaze points and the midpoint of this group exceeds 250 pixels in this stage, the characters in this cluster will move back to the original position, the detection process returns to the first stage.

3.5. One-point calibration

The one-point calibration method developed by Lutz et al. (2015) was used in this study to minimize the offset, which is a posterior improvement to the manufacturer's calibration. A number counting down from three appears in the center of the screen and this process lasts three seconds. Users were asked to look at the countdown number in the center of the screen.

Figure 4 visualizes gaze data from the pilot study before and after one-point calibration, where the participant gazes at a stationary point that appeared in sequence at five positions for three seconds on the screen.

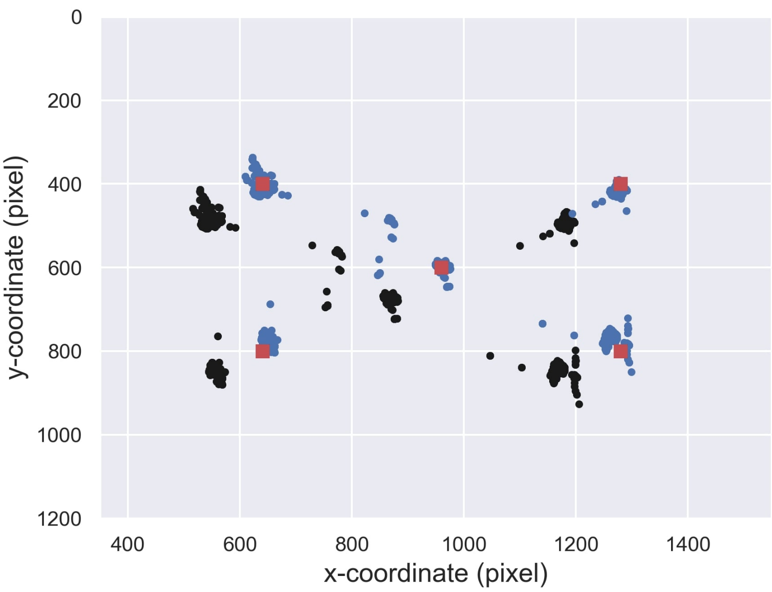

Figure 4.: Visualisation the gaze data before and after one-point calibration. Red points are target points, the black points visualize the raw gaze data, and blue points visualize adjusted gaze data after one-point calibration.

During the one-point calibration, gaze coordinates were recorded and outliers were removed using DBSCAN (Ester, Kriegel, Sander, & Xu, 1996). The average means of distance for both x-and y-axis between the midpoint of the screen and gaze positions were calculated. Then, gaze coordinates were translated with the calculated offsets. In similar experiment settings, Zeng et al. (2020) reported the average gaze estimation error was 4°. Thus if the offset is greater than 4°, the system reminds the participant to adjust their position according to the distance from the screen, and perform the calibration again. All participants were able to successfully complete the one-point calibration without re-calibration. On average across all participants, the offset was 55.42 pixels ($SD = 35.79$), i. e., 1.26° visual angle.

4. Empirical Evaluation

To investigate how the users enter text with the hybrid eye typing interface, a user study was conducted. In addition, four forms of feedback were compared for their influence on user performance and experience.

4.1. Evaluation metrics

The words per minute (WPM), keystrokes per character (KSPC), minimum string distance (MSD), and subjective feedback are used as dependent variables, explained in the following.

4.1.1. Text enter rate

Typing speed is measured in words per minute, where a word is regarded as five characters, including letters, spaces, punctuation, etc. (MacKenzie & Tanaka-Ishii, 2010). WPM is computed according to the following equation:

$$WPM = \frac{|T| - 1}{S} \times 60 \times \frac{1}{5} \tag{2}$$

where T refers to the total typed characters in the transcribed string. S represents run time in seconds.

4.1.2. Error rate

Keystrokes per character measures how frequently users correct the entered characters and the ratio of the sum of letters and including spaces to the final number of characters in the entered string (MacKenzie & Tanaka-Ishii, 2010). The formula for computing KSPC is:

$$KSPC = \frac{|IS|}{|T|} \tag{3}$$

In Eq.(3), $|IS|$ stands for the length of all characters, including backspaces. As mentioned before, $|T|$ describes the length of total typed characters in the transcribed string. If the user typed perfectly without using backspace and the KSPC value is 1.

Minimum string distance is a string metric for measuring the minimum number of error correction operations required to convert a string to another string (Levenshtein, 1966; MacKenzie & Tanaka-Ishii, 2010). The formula for calculating MSD error rate is:

$$MSD\ error\ rate = \frac{MSD(P,T)}{MAX(|P|,|T|)} \tag{4}$$

In Eq.(4), P refers the given string. It is the uncorrected error and reflects how many errors remain in transcribed string.

4.1.3. Subjective Measures

In addition to the evaluation of typing performance, the participants' opinions were collected via a short questionnaire. After completing the typing tasks for each experimental condition, participants were asked to report their perceptions of the feedback mechanisms. After finishing all typing tasks, we asked participants to rank the four feedback mechanisms and give reasons for their ranking. In addition, open-ended questions were presented to find out whether the moving was appropriate and allowed the participants to follow successfully, whether the size of the characters was clearly visible, and whether there was enough time to search for the desired object in each cluster.

4.2. Participants

In this study, 29 participants were recruited (13 female, 16 male) from an online recruitment system. Their age arranged from 19 to 38 years, with a mean of 26.7 years. All participants had German as their native language. About half of the participants (57%) reported that they had normal vision and 43% of them wore vision aids during the study (6 of them removed vision aids during the experiment). Most participants (83%) had no previous experience with gaze interaction. 72% of participants had no experience with eye tracking. Participants were rewarded with ten Euros per visit or alternatively a certification of student experimental hours for attendance.

4.3. Apparatus and Material

In this experiment, the Tobii EyeX eye-tracker was used to register the gaze location in screen coordinates with a sampling rate of 60 Hz. The eye tracker was mounted beneath a 24" Dell monitor with a resolution of 1920 * 1200 pixels. All data were collected without personal calibration from users. The eye tracker was calibrated once by the experiment supervisor through the eye tracker manufacturer's software, and this manufacturer's procedure was not performed by the participants themselves. The average distance between the participants' eyes to the display was 62 cm ($SD = 7.27$).

4.4. Procedure

After reading the informed consent form, and hygiene concept, the experiment started with an introduction of the gaze-based text entry system. The experiment consisted of one training and one test phase.

In the training phase, the interface without feedback was presented, and participants entered five phrases to learn how to use the text entry system. If the participants had no further questions about the experimental task, the test phase began. In the test phase, four different types of feedback were tested. Each participants went through all four feedback conditions (within-subject design). Each feedback condition was tested with five phrases. Entering a phrase was considered as a trial, i. e., participants typed 5 phrases for each feedback condition and in all, a total of 20 phrases were enter per participant. The order of feedback conditions and phrases was randomized. All phrases used in this experiment were selected from a set of 500 phrases (MacKenzie & Soukoreff, 2003) and translated into German. The participants were instructed to enter the given phrases as fast and as accurately as possible.

At the beginning of each trial, there was a short one-point calibration, which lasted three seconds, then a phrase was displayed in capital letters in the center of the screen. The participants then pressed the space bar to start typing after they memorized the phrase. After finishing the tasks for one condition, participants were asked to take short breaks. When participants completed all the text entry tasks, they were asked to fill out the questionnaire consisting of demographic information and subjective questions. The whole experiment duration was between 40 and 60 minutes per participant.

5. Results

This study featured a one-factor within-subjects design and the feedback factor was tested. The results are based on a total of 580 trials (29 participants * 4 feedbacks * 5 phrases). A trial was removed as an outlier because the completion time of the typing task increased significantly due to a system failure. The typing speed, accuracy and subjective feedback were analyzed.

5.1. Typing speed

The grand mean of typing speed was 4.7 WPM. The fastest typing speed was registered with 8.38 WPM. The descriptive statistics with regard to typing speed for each feedback condition are shown in Figure 5. A repeated-measures ANOVA at a significance level of $\alpha = 0.05$ was applied to analyze data. Mauchly's test of sphericity indicated that no correction was needed. The different feedback conditions were found to differ significantly for the WPM variable ($F(3, 84) = 4.61, p < .01$). A set of Bonferroni corrected t-tests were conducted for pairwise multiple comparisons. The post-hoc comparisons revealed that the WPM of the combined visual-auditory feedback was significantly greater than no feedback ($p < .001$).

5.2. Typing accuracy

5.2.1. Keystrokes per character (KSPC)

The grand mean of KSPC was 1.67. The descriptive statistics of KSPC for each feedback condition are presented in Figure 6. The Kolmogorov-Smirnov test showed that the data did not conform to a normal distribution ($p < .001$), hence the Friedman test was used to assess differences in the KSPC error between feedback conditions. No significant difference was found between the different feedback conditions ($\chi^2(3) = 6.19, p = .10$).

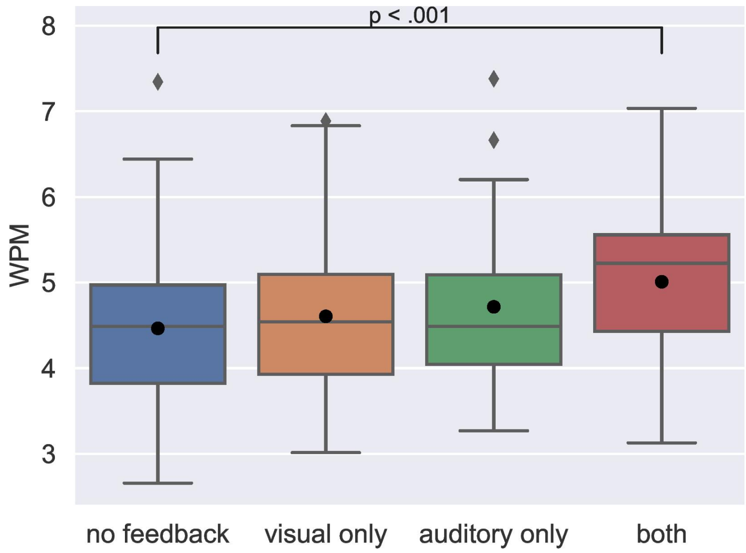

Figure 5.: Word per minute for each feedback, black dots visualize the average values, diamonds are outlier observations.

5.2.2. *Minimun string distance (MSD)*

The grand mean of MSD was 0.1 and the result of the descriptive statistics of MSD was shown in Figure 7. The Kolmogorov-Smirnov test showed that the data did not conform to a normal distribution ($p < .001$), and the Friedman test was conducted to assess the MSD error. No significant difference was found between the feedback conditions ($\chi^2(3) = 2.83, p = .42$).

5.3. Subjective evaluation

All participants reported that they noticed differences in feedback, and almost all of them correctly described these differences. However, the presence of visual feedback was not perceived by two of the participants. More than half (55%) of the participants thought that the combined visual-auditory feedback is helpful. Approximately one-third (31%) considered that visual feedback is helpful, and 17% thought that auditory feedback is helpful. Additionally, participants were asked to rank all four feedback conditions. The most preferred feedback was ranked first and the least preferred was ranked fourth. Figure 8 shows the results of this ranking.

More than half of participants ranked the sole-visual feedback condition first and nearly half of participants ranked the combined auditory-visual feedback condition first. A Friedman's test revealed that there was a significant difference between the ranking of the four feedback conditions ($\chi^2(3) = 45.68, p < .001$). Post-hoc pairwise comparisons with the Wilcoxon signed-rank test showed that the ranking for both visual-auditory feedback was significantly higher than no and sole-auditory feedback conditions ($p < .001$). Also, the ranking for the combined visual-auditory feedback condition was significantly higher than the sole-visual ($p < .001$) and no feedback condition ($p < .001$).

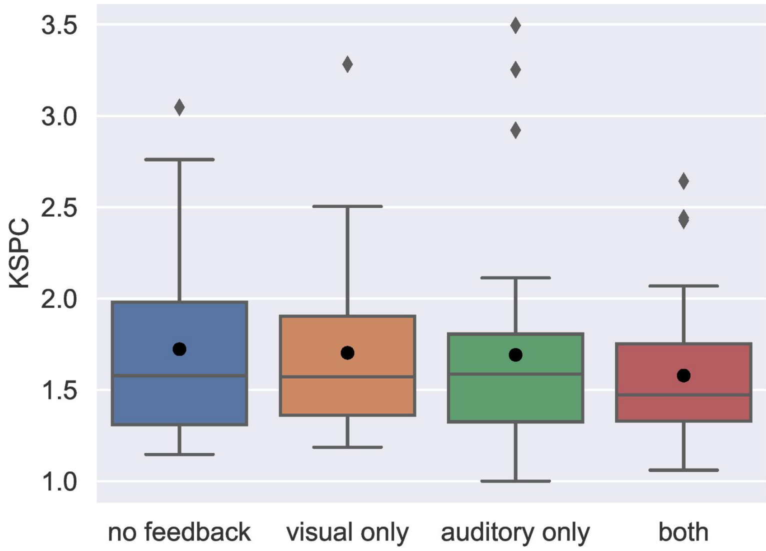

Figure 6.: Keystrokes per character for each feedback, black dots visualize the average values, diamonds are outlier observations.

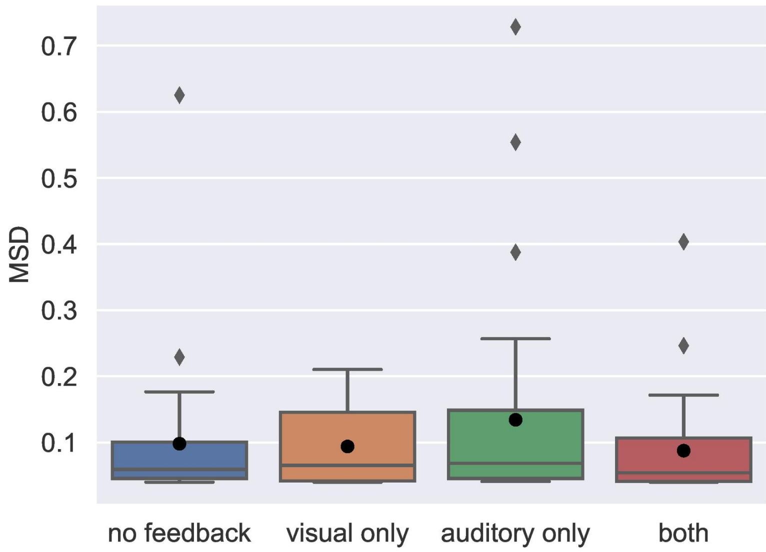

Figure 7.: Minimun string distance error rate for each feedback, black dots visualize the average values, diamonds are outlier observations.

The subjective reasons for feedback preference were collected and are summarized in Table 2. The sole-visual feedback was preferred by most participants because it not only helped the user to confirm the input but also provided information about which character is selected. Some participants reported that the tone used for the auditory

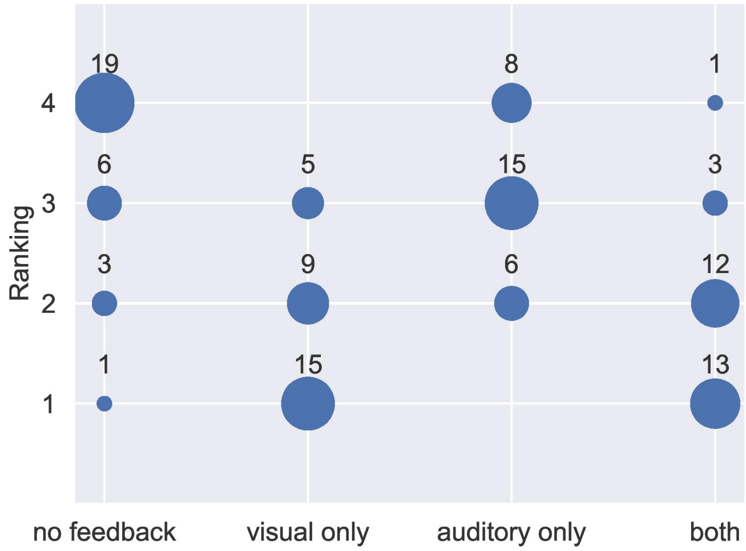

Figure 8.: Ranking result counts for the four feedback mechanisms, 1 represents the most preferred, 4 represents the lowest ranking. Above the blue circle is the count value.

feedback was distracting and disturbing.

The answer regarding the moving speed, search time, and item size were clustered and are presented in Table 3. About one-third of participants considered the movement speed of the interface as too fast which leads to unintentional input. Some participants felt that the moving speed is fast only at the beginning of the use. And some stated that some characters which were not easy to find were moving fast. Besides, there were a few participants who considered the moving speed for some characters as too slow.

At the end of the questionnaire, participants gave suggestions for potential improvement. Multiple participants suggested that the visual feedback should be more salient, and e. g., color should be added to the selected character or the font should be bold, indicating that other visual feedback might be difficult to notice. In addition, participants suggested that the feedback tone should be more user-friendly. Another suggestion was to use different types of auditory feedback for special keys, such as the delete key. Some participants indicated that the delete key could be re-positioned to a better location. Some participants were asking for the implementation of a QWERTY layout design. A reduction of movement speed and the addition of a word prediction functionality were also mentioned as potential advancements.

6. Discussion

The goal of this study was to develop a novel eye typing interface with a relatively low requirement for spatial accuracy. An experimental laboratory study was conducted to evaluate users' performance and preferences. In addition, multiple feedback stimuli were compared to facilitate immediate input confirmation during the use of the

Table 2.: Summary the reasons behind feedback preference.

Feedback	Reasons for ranking	Count
No Feedback	Confirmation of input detection is missing	13
	Need to confirm the input	6
Visual only	Provide direct feedback	6
	Better concentration on typing task	5
	Given confirmation of the input	2
	Checking the input is omitted	2
	Less mistakes	2
Auditory only	Sound is disturbing	6
	Eye response to sound causes errors	2
Both	More robust system interaction	10
	Double support in interaction	2

interface.

Our assessment of effectiveness and efficiency variables, as well as user feedback for the hybrid eye typing interface, shows that the system represents an enjoyable and well functioning human-machine interaction approach. The results show that the participants were able to learn how to efficiently use the system within an hour. While the text entry speed count of 4.7 WPM found in this study is not high in comparison to other gaze-based spellers which have a high requirement for personal calibration, the result has to be seen in the light of the nature of participants in this study, all of whom were novice users of this novel layout. Most of them had no previous experience with gaze interaction. Besides, there are still critical advantages over existing spellers. The calibration process for the proposed hybrid design is reduced to three seconds. Such one-point calibration eye typing methods need a shorter time required for the calibration process of the eye tracker, similar to the process of unlocking a cell phone. Since the angle of the gaze-path is mostly unaffected by a gaze point offset that is sometimes found with one-point calibration, this gaze speller is well suited for immediate interactions. The relatively simple arrangement of usable objects further supports this. The hybrid eye typing interface alleviates the challenges of a lack of calibration in walk-up-and-use eye-tracking solutions. There is also no learning of specific eye-movements necessary (like in gesture-based systems), as users do not need to remember or learn gestures to use the hybrid system. It can be expected that any introduction of gaze-based interfaces in the public will initially depend on a high level of detection robustness, the simpleness of the interface, and the understandability of the interaction design.

There is no paper with a systematic comparison of the typing performance of eye typing systems without personal calibration and existing papers are inconsistent in terms of participants, experimental conditions, and evaluation metrics (Abdrabou et al., 2019; Bafna et al., 2021; Lutz et al., 2015; Porta et al., 2021; Zeng & Roetting, 2018). Nonetheless, in the following we provide a rough comparison and discussion

Table 3.: Perception of the text entry system.

Topic	Description	Count
Speed	Too fast	11
	Feel fast at the beginning of the use	6
	Too fast for unfamiliar letters	6
	Appropriate	3
	Some keys (e.g., space) are slow	3
Search time	Time is enough to search the desired tiles	14
	Not enough time for searching	13
	After using it for a while, the time is enough	2
Item size	Appropriate	23
	Some keys (e.g., delete) are small	3
	Too small	3

here based on the data reported in each paper.

The SMOOVS system using one-point calibration reports that the typing speed varies in the range of 2.9 to 3.34 WPM (Lutz et al., 2015), the average typing speed was improved to 4.5 WPM by providing one-word prediction (Zeng & Roetting, 2018). There are also other calibration-free systems. Abdrabou et al. (2019) first designed the calibration-free eye-typing interface and this interface achieved an average typing speed of 3.41 WPM. This system adds word prediction function, which gives six predicted words. The reported typing speed of EyeTell, which uses a front-facing camera, and provides 3 predicted letters, was 1.27 WPM (Bafna et al., 2021). More recently, SPEye achieved a typing speed of 0.85 WPM without word prediction and 1.15 WPM with word prediction (Porta et al., 2021).

As a main advantage over existing systems, the exclusive actionability of select object groups through direct feedback increases user feedback and user understanding during the use of the system. This new eye typing interface features a hybrid design, i. e., only the items in the gaze focus are moving. The interaction is divided into two stages. The cluster selection is corresponding to the static stage, where the cluster looked at by the user is highlighted, but all characters on this interface remain static which enables the user to search the desired character more easily. Only the characters in one cluster begin to move when the user is fixating on this cluster for a certain time. The final selection of characters is done in the dynamic stage. Although there is a short one-point calibration, our system obtains a smoother and faster typing performance.

Porta et al. (2021) compared the typing errors for the typing systems mentioned above, the SPEye system achieved a relatively low error rate (close to zero). The error rates of the system proposed in this paper were relatively high, and participants reported that the system was relatively sensitive. Some participants commented that the moving speed was fast and they didn't have enough time to search the desired letter, with one participant stating "It's a bit fast when selecting letter groups, which is why it often opens some groups that you don't want to open." To achieve a more robust detection, a longer dwell time could be introduced, especially for novices who are not familiar with the layout. The results suggest that the movement speed of the interface in this study was potentially set too high for novices, although users could get used to faster speeds as usage time grows. Besides, the typing performance might be increased with consideration of learning effects and individual differences. For example, users could be allowed to adjust the parameter themselves (Špakov & Miniotas, 2004) or the system could automatically adjust to meet the need based on past typing records (Mott et al., 2017; Pi et al., 2020).

The results also indicated that the type of feedback used in the interface significantly relates to the typing speed. The highest typing speed was found in the feedback condition with the combined visual and auditory feedback. No significant difference was found either on KSPC or MSD error regarding feedback conditions. Hence, it can be concluded that for typing efficiency in the hybrid eye typing interface, it is optimal to provide combined auditory and visual feedback, as the experimental data suggests that it works better than no feedback.

Although the evaluation of user performance indicated that the combined visual and auditory feedback resulted in higher typing speed and lower error rates, the evaluation of user preference showed that 52% of participants ranked the sole-visual feedback to be the most preferred feedback, and slightly fewer participants (45%) preferred the combined visual and auditory feedback most. No participants felt that visual feedback caused extra visual noise that made it hard to concentrate on typing. There was no significant increase in demand for visual resources. Participants appreciated that visual feedback informed them that input was registered and which action was activated. The reason that fewer participants like the auditory feedback were similar to what Špakov et al. (2016) reported about auditory feedback. Participants found the auditory feedback to be useful but distracting in both studies.

While the character set of the interface used in this study was limited to letters, future implementations of the interface functionality could include a character-type selection screen, increasing the available character space. Additionally, future work could investigate to use pursuit calibration to improve the calibration accuracy (Drewes, Pfeuffer, & Alt, 2019; Pfeuffer, Vidal, Turner, Bulling, & Gellersen, 2013). The detection algorithm could also be improved, such as using a linear regression-based algorithm (Drewes, Khamis, & Alt, 2019), Profile Matching and 2D Correlation (Velloso, Coutinho, Kurauchi, & Morimoto, 2018), to allow the interface accommodating more targets and make the detection more robust. Besides, provision of word prediction for this hybrid eye typing interface could also be considered to further speed up the typing efficiency. Another possible area of future research is the exploration of combined input modalities for more efficient and natural user experiences, such as head movement (Sidenmark, Mardanbegi, Gomez, Clarke, & Gellersen, 2020) and touch (Kumar, Hedeshy, MacKenzie, & Staab, 2020). Further, this eye typing interface can be utilized in applications of augmented and virtual reality. Additionally, since the eye-tracker based gaze registration is subject to the limitation in the detectable tracking range, a webcam-based gaze estimation method could be used to increase the available area for gaze registration. Compared with the eye tracker used in this study (Tobii EyeX), a webcam-based gaze estimation extends the maximum tolerance distance range from 75 cm to 180 cm (Zhang, Sugano, & Bulling, 2019). In addition, it would considerably reduce the price of equipment and also broaden use-case scenarios.

7. Conclusion and Future Work

The eye typing interface proposed in this study represents a prime candidate for implementation in public spaces. This hand-free text entry system can help to limit touches and helps to promote hand hygiene. The system is also considered to be more convenient, e. g., for people with disabilities. It could be complementary, where eye input can be used as an alternative interaction method, coexisting with physical buttons and touch screens.

This new eye typing interface enables the user to enter text by eye after a three seconds calibration process, which shortens the time for calibration, and it provided robust typing performance as well as relatively high user acceptance. The interface successfully avoids the Midas touch problem, through the implementation of a static stage, that allows users to search the desired character. In later dynamic stage, only objects within the attention area are moving. Users follow the movement of one desired object for selection. While the combined visual and auditory feedback resulted in the highest typing speed, subjective data revealed that users prefer sole-visual to combined visual and auditory feedback, because the sound was distracting sometimes.

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.

Acknowledgments The author Zhe Zeng would like to thank the China Scholarship Council (CSC) for financially supporting her PhD study at TU Berlin, Germany.

References

Abdrabou, Y., Mostafa, M., Khamis, M., & Elmougy, A. (2019). Calibration-free text entry using smooth pursuit eye movements. New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3314111.3319838

Almoctar, H., Irani, P., Peysakhovich, V., & Hurter, C. (2018). Path word: A multimodal password entry method for ad-hoc authentication based on digits' shape and smooth pursuit eye movements. In *Proceedings of the 20th acm international conference on multimodal interaction* (p. 268–277). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3242969.3243008

Bafna, T., Bækgaard, P., & Paulin Hansen, J. P. (2021). Eyetell: Tablet-based calibration-free eye-typing using smooth-pursuit movements. In *Acm symposium on eye tracking research and applications.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3448018.3458015

Blattgerste, J., Renner, P., & Pfeiffer, T. (2018). Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the workshop on communication by gaze interaction.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3206343.3206349

Cymek, D. H., Venjakob, A. C., Ruff, S., Lutz, O. H.-M., Hofmann, S., & Roetting, M. (2014). Entering pin codes by smooth pursuit eye movements. *Journal of Eye Movement Research*, *7*(4), 1. Retrieved from https://doi.org/10.16910/jemr.7.4.1

Drewes, H., Khamis, M., & Alt, F. (2019). Dialplates: Enabling pursuits-based user interfaces with large target numbers. In *Proceedings of the 18th international conference on mobile and ubiquitous multimedia.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3365610.3365626

Drewes, H., Pfeuffer, K., & Alt, F. (2019). Time- and space-efficient eye tracker calibration. In *Proceedings of the 11th acm symposium on eye tracking research & applications.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3314111.3319818

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.

Eid, M. A., Giakoumidis, N., & El Saddik, A. (2016). A novel eye-gaze-controlled wheelchair system for navigating unknown environments: Case study with a person with als. *IEEE Access*, *4*, 558-573.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, pp. 226–231).

Esteves, A., Velloso, E., Bulling, A., & Gellersen, H. (2015). Orbits: Gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th annual acm symposium on user interface software & technology* (pp. 457–466).

ETA. (2020, October). *Computer qwerty keyboard key frequency.* http://letterfrequency.org/.

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. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 1118–1130). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025453.3025599

Hansen, D. W., Skovsgaard, H. H. T., Hansen, J. P., & Møllenbach, E. (2008). Noise tolerant selection by gaze-controlled pan and zoom in 3d. In (p. 205–212). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1344471.1344521

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. In *Interact* (Vol. 3, pp. 121–128).

Huckauf, A., & Urbina, M. H. (2008). Gazing with peyes: Towards a universal input for various applications. In *Proceedings of the 2008 symposium on eye tracking research & applications* (p. 51–54). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1344471.1344483

Hwang, C.-S., Weng, H.-H., Wang, L.-F., Tsai, C.-H., & Chang, H.-T. (2014). An eye-tracking assistive device improves the quality of life for als patients and reduces the caregivers' burden. *Journal of motor behavior*, *46*(4), 233–238. Retrieved from https://doi.org/10.1080/00222895.2014.891970

Jacob, R. J. (1990). What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 11–18).

Jacob, R. J. (1995). Eye tracking in advanced interface design. *Virtual environments and advanced interface design*, 258–288.

Kangas, J., Akkil, D., Rantala, J., Isokoski, P., Majaranta, P., & Raisamo, R. (2014). Gaze gestures and haptic feedback in mobile devices. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 435–438). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2556288.2557040

Kangas, J., Špakov, O., Isokoski, P., Akkil, D., Rantala, J., & Raisamo, R. (2016). Feedback for smooth pursuit gaze tracking based control. In *Proceedings of the 7th augmented human international conference 2016.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2875194.2875209

Katsini, C., Abdrabou, Y., Raptis, G. E., Khamis, M., & Alt, F. (2020). The role of eye gaze in security and privacy applications: Survey and future hci research directions. In *Proceedings of the 2020 chi conference on human factors in computing systems* (p. 1–21). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3313831.3376840

Ke, S. R., Lam, J., Pai, D. K., & Spering, M. (2013). Directional asymmetries in human smooth pursuit eye movements. *Investigative ophthalmology & visual science*, *54*(6), 4409–4421. Retrieved from https://doi.org/10.1167/iovs.12-11369

Köpsel, A., Majaranta, P., Isokoski, P., & Huckauf, A. (2016). Effects of auditory, haptic and visual feedback on performing gestures by gaze or by hand. *Behaviour & Information Technology*, *35*(12), 1044–1062.

Krukowski, A. E., & Stone, L. S. (2005). Expansion of direction space around the cardinal axes revealed by smooth pursuit eye movements. *Neuron*, *45*(2), 315–323.

Kumar, C., Hedeshy, R., MacKenzie, I. S., & Staab, S. (2020). Tagswipe: Touch assisted gaze swipe for text entry. In *Proceedings of the 2020 chi conference on human factors in computing systems* (p. 1–12). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3313831.3376317

Kurauchi, A., Feng, W., Joshi, A., Morimoto, C., & Betke, M. (2016). Eyeswipe: Dwell-free text entry using gaze paths. In *Proceedings of the 2016 chi conference on human factors in computing systems* (p. 1952–1956). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2858036.2858335

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710).

Ludvigh, E., & Miller, J. W. (1958). Study of visual acuity during the ocular pursuit of moving test objects. i. introduction. *JOSA*, *48*(11), 799–802.

Lutz, O. H.-M., Venjakob, A. C., & Ruff, S. (2015). Smoovs: Towards calibration-free text entry by gaze using smooth pursuit movements. *Journal of Eye Movement Research*, *8*(1), 2.

MacKenzie, I. S., & Soukoreff, R. W. (2003). Phrase sets for evaluating text entry techniques. In *Chi '03 extended abstracts on human factors in computing systems* (p. 754–755). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/765891.765971

MacKenzie, I. S., & Tanaka-Ishii, K. (2010). *Text entry systems: Mobility, accessibility, universality*. Elsevier.

Majaranta, P., Ahola, U.-K., & Špakov, O. (2009). Fast gaze typing with an adjustable dwell time. In *Proceedings of the sigchi conference on human factors in computing systems* (p. 357–360). New York, NY, USA: Association for Computing Machinery. Retrieved from 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. In *Chi '03 extended abstracts on human factors in computing systems* (p. 766–767). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/765891.765979

Majaranta, P., MacKenzie, I. S., Aula, A., & Räihä, K.-J. (2006). Effects of feedback and dwell time on eye typing speed and accuracy. *Universal Access in the Information Society*, *5*(2), 199–208.

Majaranta, P., & Räihä, K.-J. (2002). Twenty years of eye typing: Systems and design issues. In *Proceedings of the 2002 symposium on eye tracking research & applications* (pp. 15–22). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/507072.507076

Menges, R., Kumar, C., Müller, D., & Sengupta, K. (2017). Gazetheweb: A gaze-controlled web browser. In *Proceedings of the 14th web for all conference on the future of accessible work.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3058555.3058582

Morimoto, C. H., & Amir, A. (2010). Context switching for fast key selection in text entry applications. In (p. 271–274). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1743666.1743730

Mott, M. E., Williams, S., Wobbrock, J. O., & Morris, M. R. (2017). Improving dwell-based gaze typing with dynamic, cascading dwell times. In *Proceedings of the 2017 chi conference on human factors in computing systems* (p. 2558–2570). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3025453.3025517

Nielsen, J. (1994). *Usability engineering*. Morgan Kaufmann.

Pedrosa, D., Pimentel, M. D. G., Wright, A., & Truong, K. N. (2015, March). Filteryedping: Design challenges and user performance of dwell-free eye typing. *ACM Trans. Access. Comput.*, *6*(1). Retrieved from https://doi.org/10.1145/2724728

Penkar, A. M., Lutteroth, C., & Weber, G. (2012). Designing for the eye: Design parameters for dwell in gaze interaction. In *Proceedings of the 24th australian computer-human interaction conference* (pp. 479–488). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2414536.2414609

Perlin, K. (1998). Quikwriting: continuous stylus-based text entry. In *Proceedings of the 11th annual acm symposium on user interface software and technology* (pp. 215–216).

Pfeuffer, K., Vidal, M., Turner, J., Bulling, A., & Gellersen, H. (2013). Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proceedings of the 26th annual acm symposium on user interface software and technology* (p. 261–270). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2501988.2501998

Pi, J., Koljonen, P. A., Hu, Y., & Shi, B. E. (2020). Dynamic bayesian adjustment of dwell time for faster eye typing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 1-1.

Porta, M., Dondi, P., Pianetta, A., & Cantoni, V. (2021). Speye: A calibration-free gaze-driven text entry technique based on smooth pursuit. *IEEE Transactions on Human-Machine Systems*, 1-12.

Rajanna, V., & Hansen, J. P. (2018). Gaze typing in virtual reality: Impact of keyboard design, selection method, and motion. In *Proceedings of the 2018 acm symposium on eye tracking research & applications.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3204493.3204541

Rantala, J., Majaranta, P., Kangas, J., Isokoski, P., Akkil, D., Špakov, O., & Raisamo, R. (2020). Gaze interaction with vibrotactile feedback: Review and design guidelines. *Human{ Computer Interaction*, *35*(1), 1–39.

Robinson, D. (1968). The oculomotor control system: A review. *Proceedings of the IEEE*, *56*(6), 1032-1049.

Sidenmark, L., Clarke, C., Zhang, X., Phu, J., & Gellersen, H. (2020). Outline pursuits: Gaze-assisted selection of occluded objects in virtual reality. In *Proceedings of the 2020 chi conference on human factors in computing systems* (p. 1–13). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3313831.3376438

Sidenmark, L., Mardanbegi, D., Gomez, A. R., Clarke, C., & Gellersen, H. (2020). Bimodalgaze: Seamlessly refined pointing with gaze and filtered gestural head movement. In *Acm symposium on eye tracking research and applications.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3379155.3391312

Velloso, E., Coutinho, F. L., Kurauchi, A., & Morimoto, C. H. (2018). Circular orbits detection for gaze interaction using 2d correlation and profile matching algorithms. In *Proceedings of the 2018 acm symposium on eye tracking research & applications.* New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3204493.3204524

Vidal, M., Bulling, A., & Gellersen, H. (2013). Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 acm international joint conference on pervasive and ubiquitous computing* (p. 439–448). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2493432.2493477

Špakov, O., Isokoski, P., Kangas, J., Akkil, D., & Majaranta, P. (2016). Pursuitadjuster: An exploration into the design space of smooth pursuit based widgets. In *Proceedings of the ninth biennial acm symposium on eye tracking research & applications* (p. 287–290). New York, NY, USA: Association for Computing Machinery. Retrieved from https://doi.org/10.1145/2857491.2857526

Špakov, O., & Miniotas, D. (2004). On-line adjustment of dwell time for target selection by gaze. In *Proceedings of the third nordic conference on human-computer interaction* (p. 203–206). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/1028014.1028045`

Ward, D. J., & MacKay, D. J. (2002). Fast hands-free writing by gaze direction. *Nature*, *418*(6900), 838–838.

Wobbrock, J. O., Rubinstein, J., Sawyer, M. W., & Duchowski, A. T. (2008). Longitudinal evaluation of discrete consecutive gaze gestures for text entry. In *Proceedings of the 2008 symposium on eye tracking research & applications* (p. 11–18). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/1344471.1344475`

Zeng, Z., & Roetting, M. (2018). A text entry interface using smooth pursuit movements and language model. In *Proceedings of the 2018 acm symposium on eye tracking research & applications* (pp. 69:1–69:2). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/3204493.3207413`

Zeng, Z., Siebert, F. W., Venjakob, A. C., & Roetting, M. (2020). Calibration-free gaze interfaces based on linear smooth pursuit. *Journal of Eye Movement Research*, *13*(1), 3.

Zhang, X., Sugano, Y., & Bulling, A. (2019). Evaluation of appearance-based methods and implications for gaze-based applications. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–13).

About the Authors

Zhe Zeng received her M.S. and Ph.D. degree from the Technical University of Berlin, Germany. Her research interests include eye tracking technology, natural gaze interaction and human factors.

Elisabeth Sumithra Neuer received her M.S. degrees in Human factors from Technical University of Berlin. Her interests include eye tracking and gaze interaction.

Matthias Roetting is professor for Human-Machine Systems at the Technical University of Berlin, Germany. He received his Ph.D. degree from Aachen University of Technology, in 2001. His research interests include eye tracking technology, human–machine interaction, as well as driver assistance systems.

Felix Wilhelm Siebert received his Ph.D. degree in psychology from the Leuphana University of Lüeneburg, Germany. He holds an assistant professor position for transport psychology at the Department of Technology, Management and Economics at the Technical University of Denmark.