SkiMR: Dwell-free Eye Typing in Mixed Reality

Category: Research

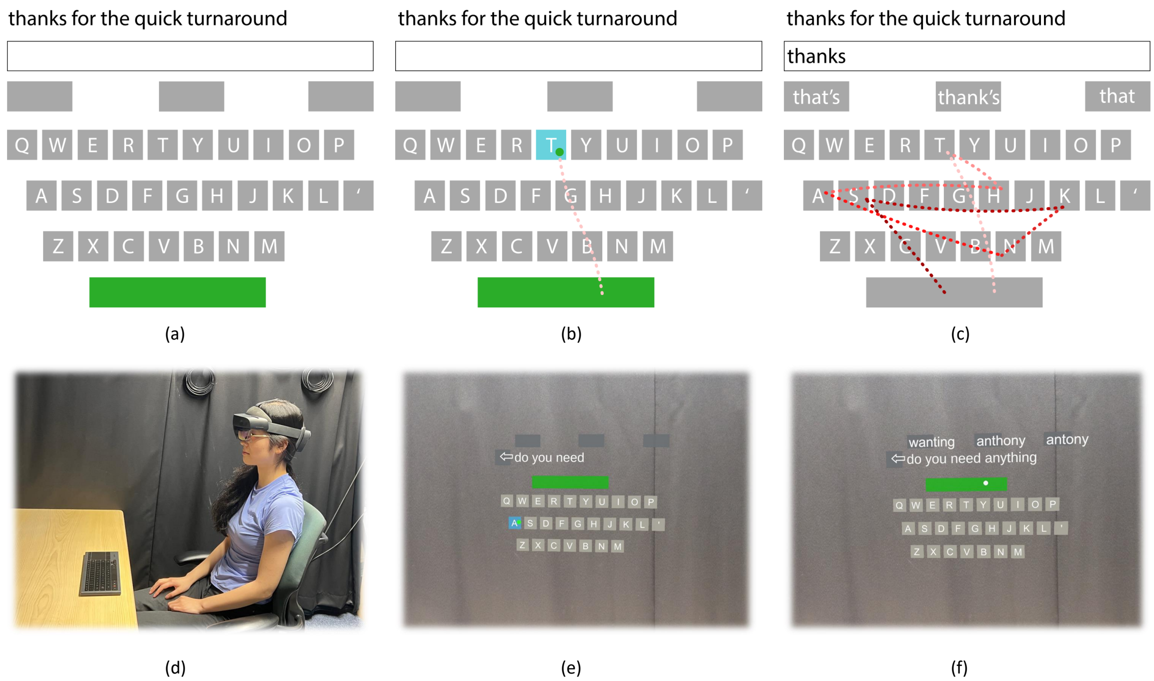


Figure 1: (a)-(c): illustrations of the dwell-free system used in Study 1. The user initiates the typing of the word 'thanks' by gazing at the *SPACE* key as shown in (a). The *SPACE* will be highlighted once triggered. Then, by sequentially fixating on each letter key, the user constructs the word. After finalizing the sequence by skimming the *SPACE* key again the recognized word 'thanks' is typed in (c). (d): The study setting of a participant wearing a mixed reality headset. (e) and (f) show the actual typing scene in Study 2 with the refined keyboard design.

Abstract

We present SkiMR: a dwell-free eye typing system that enables fast and accurate hands-free text entry on mixed reality headsets. SkiMR uses a probabilistic statistical decoder to infer users' intended text based on users' eye movements on a virtual keyboard. It does not rely on dwell timeouts, which enables it to be faster than traditional eye typing. We study this dwell-free eye typing system, deployed on a HoloLens 2, in two studies. In the first study ($n = 12$) we show that dwell-free eye typing results in a significantly faster text entry rate compared to traditional dwell-based eye typing with word prediction support and a hybrid dwell-free method that uses dwell timeouts to delimit word entry. Based on the insights from the first study we evaluated the feasibility of a refined system in a more realistic composition task and using an interaction mechanism that provided real-time predictions during dwell-free eye typing. The second study ($n = 16$) demonstrates that this final system allowed users to compose original text at 12 words per minute with a corrected character error rate of 1.1%. Overall, this work demonstrates the high potential for fast and accurate hands-free text entry using dwell-free eye typing for mixed reality headsets.

Keywords: Text input, Mixed / augmented reality, Keyboards

1 Introduction

Text entry is an important everyday activity and as a consequence mixed reality (MR) headsets have explored a range of possible methods, including touch-typing and gesture typing in thin air by monitoring the user's fingers through hand tracking (e.g. [4, 5, 8, 13, 28]).

In this paper we explore a *complementary* text entry modality by allowing users to type on a MR headset by eye gaze. This allows users to enter text in situations where hand interaction is not practical or seamless, such as when users are encumbered, for example by holding bags, or using their hands to manipulate something, such as machinery in a manufacturing plant. It also allows text entry in a more discrete way in public spaces compared to typing in thin air.

The traditional way to realize typing by eye gaze is called eye typing [27]. The system tracks the user's eye gaze through an eye tracker and the user's gaze point is used to signal the intended letter key. Due to the Midas touch problem [12], the system cannot disambiguate whether a user wants to *look* at a key or *type* a key, the system uses a dwell-timeout, which is typically 400–1000 ms [27]. However, traditional eye typing has three limitations [16]: (1) text entry performance is limited by the dwell timeouts; (2) the eyes are sensory organs rather than control organs foremost and hence it is straining for users to force a fixation for a set duration during typing; and (3) when users write they prefer to think in terms of words, phrases and sentences, however, dwell-based eye typing forces users to think about the writing task as a series of low-level character input actions.

As an alternative, prior work has proposed dwell-free eye typing for nonspeaking individuals with motor disabilities [16], which has been realized as a product as part of the Tobii-Dynavox system [15]. This is a system that runs on a computer coupled with an eye tracker.

The system is typically mounted on a wheelchair or a desk in front of the user. Dwell-free eye typing eliminates dwell-timeouts by decoding the user's eye gaze sequence over the keyboard into text.

In this paper we explore the potential of dwell-free eye typing as a hands-free text entry modality for mixed reality headsets. We present a system called SkiMR that realizes dwell-free eye-typing on the HoloLens 2 mixed reality headset and performs decoding on-device, thus allowing dwell-free eye typing as a truly mobile input modality for mixed reality systems. The system enables dwell-free eye typing by allowing users to enter text by sequentially gazing at the intended letter keys, including space, without an explicit dwell timeout. In this way the system allows the user to "skim" the keys. This novel mobile mixed reality text entry system allows users to write at a reasonable speed while simultaneously leaving both of the users' hands free. This makes it possible for users to, for example, compose brief messages, command sequences, or queries while using both of their hands to do physical work. For instance, users may be encumbered by holding bags in their hands, or by working in an environment that demands both hands for other interactions, such as if a user is working with machinery in a manufacturing setting.

In the first study we use this system to compare dwell-free eye typing with a baseline method that uses traditional dwell-based eye typing with word prediction (DWELL+WORDPREDICTION) and a variant of dwell-free eye typing where each word is delimited by dwell timeouts (DWELLDELIMITEDWORD). This study demonstrates that dwell-free eye typing is significantly faster than the other two hands-free text entry methods in a transcription task. However, it also induces a higher uncorrected character error rate.

We then improved the system based on our observations from the first study. This allowed us to study a refined version of dwell-free eye typing on a mixed-reality headset in a composition task, which is a more realistic task as in reality users *compose* their own text rather than copy stimulus words. We explored the interaction mechanism of facilitating word predictions and found that providing a next-word suggestion did not affect performance and that users were able to enter text at 12 WPM with a corrected character error rate of 1.1%.

In summary, this paper makes the following contributions:

- We present SkiMR: a system that enables hands-free text entry using dwell-free eye typing on a mixed reality headset. We demonstrate that this system allows an entry rate of 12 WPM with a corrected character error rate of 1.1% in a composition task.
- We find that dwell-free eye-typing is significantly faster than traditional dwell-based eye typing with prediction and a variant of dwell-free eye typing in which each word is delimited by a dwell timeout.

2 RELATED WORK

Dwell-based eye typing has been a subject of intense research for many years [27], primarily serving as a communication tool for individuals with physical disabilities. The conventional method necessitates users to select each key by maintaining their gaze on it for a specified duration. While various research efforts have sought to optimize performance through means such as reducing dwell time [26, 29], minimizing visual search time [31], and employing text completions and predictions [25], the inherent constraints of character-by-character entry coupled with dwell time thresholds restrict the maximum achievable entry rate. Numerous methods have been explored to eliminate dwell-time dependency in eye-typing.

2.1 Multimodal Eye Typing

One approach that can eliminate dwell-time is the incorporation of multimodal input channels as additional means to explicitly delimit commands or selections, combined with a relatively simple word decoder. For instance, one novel method combined gaze with humming [11], where humming signified the beginning and end of a word. Similarly, HGaze Typing [7] amalgamated head gestures with gaze to handle text entry, selection, deletion, and revision tasks, showcasing enhanced efficiency and user satisfaction over dwell-time-based keyboards. In another inventive approach, a multimodal setup paired gaze with foot input, yielding significantly improved typing speeds while maintaining comparable error rates [34]. Finally, swipe touch input either on a smartphone [18] or on smart glasses [1] was utilized in conjunction with gaze to expedite typing. These studies exemplify the potential of multi-modal methods in text entry systems to successfully tackle dwell-time dependency, a concept directly aligned with the primary objective of our research on dwell-free eye typing.

2.2 Smooth Pursuit

Smooth pursuit eye movements have been effectively utilized to achieve dwell-free eye typing, enabling a more fluid and continuous method of text entry. Dasher [41, 42] embodies this principle by leveraging the user's gaze direction across a vertically arranged alphabet for letter selection, achieving a continuous writing experience. StarGazer [10] extends this concept by employing a zooming interface, wherein the gaze direction dictates the pan and zoom, allowing for the selection of desired letters. pEYEwrite [37] takes a different approach by employing a hierarchical pie menu. It provided an alternative way for the selection of letter keys via smooth pursuit movements across the borders of pie slices, promoting a continuous text entry experience. EyeK [35] used a gaze gesture of inside-outside-inside to determine the selection of each key.

2.3 Dwell-free Eye Typing

Although the methods mentioned above eliminated the dwell-time in eye typing through various means, we specifically use the term *dwell-free eye typing* as originally posited by Kristensson et al. [16]. In this prior work, dwell-free systems enable users to input text by sequentially fixating on the intended letters, without the need to exceed a dwell threshold by maintaining the gaze point on a key for a particular period of time. It also envisioned that the complete elimination of dwell-timeout mechanisms could allow for the input of text at various granularities, including characters, words, and sentences, through the use of a statistical decoder supported by a language model similar to that used in conventional gesture keyboards [14, 17, 45]. Theoretically, the speed of such dwell-free systems could surpass that of existing dwell-based eye-typing techniques with the maximum attainable performance reaching 46 WPM in an ideal situation [16]. The system was later realized as a commercial product as an alternative option to a dwell-based method for disabled users who solely rely on eye typing for daily communication. Another study [15] presented the deployment of the commercial system and used computational simulations to understand the barriers and potential improvements to dwell-free eye-typing adoption. The research, conducted in users' homes, uncovered design issues and performance implications for dwell-free eye-typing systems, culminating in recommendations for enabling its wider adoption in practice.

Other research work has primarily concentrated on word-at-a-time inputs for dwell-free eye typing. Such systems generally operate in two steps: first, the delimitation mechanism identifies the start and end of a word; second, a word gesture recognizer interprets the gaze trace into potential words. However, these steps can vary across different systems. Filteryedping [32], for instance, relies solely on sequential gaze-at letters. It generates a candidate list of words composed of subsequences of the input letter sequence. Words containing all the letters from the input sequence are shortlisted. These candidate words are then sorted by length and dictionary frequency. However, this algorithm struggles with neighbor-letter errors and missing-letter errors. Both GazeTry [22] and LCSMapping [21] focused on reducing recognition errors. MoWing [22] uses a cost function to determine the conversion cost between the input letter sequence and each word, ranking candidates according to this cost. LCSMapping [21], on the other hand, also considers letter location and gaze duration using the fixation-saccades model in addition to gazed letters. EyeSwipe [19] introduces a selection mechanism called 'reverse crossing' to indicate the first and last characters of a word. GlanceWriter [2] uses a word delimitation mechanism that is similar to ours with an upper keyboard bound near the 'space' key to deliniate the start/end of a word, thereby serving a similar function to our 'space bar'. However, the entry rates achieved by GlanceWriter [2] in two user studies were relatively low (10.89 WPM and 9.54 WPM) compared to our results.

Overall, the dwell-free eye-typing systems previously discussed have predominantly been implemented and evaluated within a desktop environment using commercial eye trackers with high operating frequencies. Our work represents the first deployment and examination of such a system on a MR headset. Despite the significant constraints of the eye-tracking capabilities on HoloLens 2, our system demonstrates strong performance for both speed and accuracy.

2.4 Hands-free Text Entry in HMDs

While eye typing has been extensively studied previously on desktop environments, only one study examined eye typing in VR head-mounted displays (HMDs) [33]. The findings suggest that dwell-based gaze typing in VR was viable. Users performed best (10.15 WPM) when the entire keyboard was within view, while the larger-than-view keyboard (9.15 WPM) induced physical strain due to increased head movements. The study also found users performed significantly better while stationary than when in motion. Furthermore, gaze with confirmation of selection by clicking on a controller was found to be a better selection method than dwell-only interaction.

Other previous works mainly focused on head gaze input on HMDs, that is using a projected cursor controlled by head movements rather than eye movements. Yu et al. [44] explored three text entry methods using head gaze motions on a mobile VR headset. Two character-based methods utilized button tapping or dwelling for delimitation. The third method, based on template matching, allowed for word-level text entry with a controller button pressed to delimit word start and end. The initial tests revealed typing speeds of 10.59 WPM, 15.58 WPM, and 19.04 WPM for each method respectively. Users found all methods easy to learn and rated fatigue levels as acceptable. After optimizing the recognition algorithm for the GestureType method, users achieved a typing speed of 24.73 WPM after an hour of training. Later, more delimitation mechanisms were investigated such as eye blinks and neck movements [24].

RingText [43] utilized a circular keyboard layout comprised of two concentric circles. The outer circle housed the letters, and the resultant text was displayed within the inner circle. Gazing at a letter region with the cursor would trigger a selection, while a retreat into the inner circle would reset the selection. A refined version resulted in an entry rate of 11.30 WPM after 60 minutes of training.

An imaginary (invisible) keyboard [23] on AR HMDs was evaluated using three selection mechanisms combined with head gaze motions: eye blinks, dwell, and swipe gestures. The blinks mechanism was found to be the most effective, resulting in an average text entry speed of 11.95 WPM.

Eye gaze, employed as a hands-free approach, has demonstrated potential higher speed over head gaze movements in virtual environments [3]. This could result in more responsive and natural tracking for small-scale movements. The precision of eye movements allows for accurate selection or tracking within a confined field of view, making it more compatible with MR headsets that aim for minimal field-of-view occlusion. Large, sustained head movements may lead to fatigue or neck strain, whereas the musculature controlling eye movement may not tire as easily [3].

3 System Design

Our system runs on a Microsoft HoloLens 2 headset. We use the built-in eye tracking using the Extended Eye Tracking API on the device to obtain eye gaze data at a frequency of 90 Hz. The system itself is developed in Unity. The application is deployed directly to the Hololens 2 headset and runs on the device. The statistical decoding also runs on the device itself.

We implemented two variants of eye-typing that reduce or omit dwell timeouts: (1) DwellFree: dwell-free eye typing, which enables users to type a sentence without dwell timeouts; and (2) DwellDelimitedWord: a word-delimited solution that only requires dwell-based selection on the first and last character in the word. These two variants are now described in more detail.

3.1 DwellFree

The dwell-free eye typing process commences when a user looks at the space bar, shown in Figure 1(a), which initiates the gaze trace. The user then gazes sequentially at the letters of the intended word and completes the trace by gazing at the space bar once again as illustrated in Figure 1(b)(c). The system records the sequence of gaze points throughout this process. These points are first filtered and classified using the I-VT (identification-velocity threshold) algorithm [30], which helps in identifying fixations and saccades.

The system preprocesses the trace by only considering the gaze trace from the first fixation to the last fixation. The gaze points from the space bar to the first letter and from the last letter back to the space bar are discarded.

Thereafter a gesture decoder processes the sequence of gaze points and returns a set of potential words. We adapted a word gesture decoder [6] for this purpose. The decoder uses a probabilistic model that interprets the gaze trace by considering the statistical likelihood of letter sequences and the spatial proximity of the keys on the keyboard, with the ability to handle deletion, insertion and substitution errors.

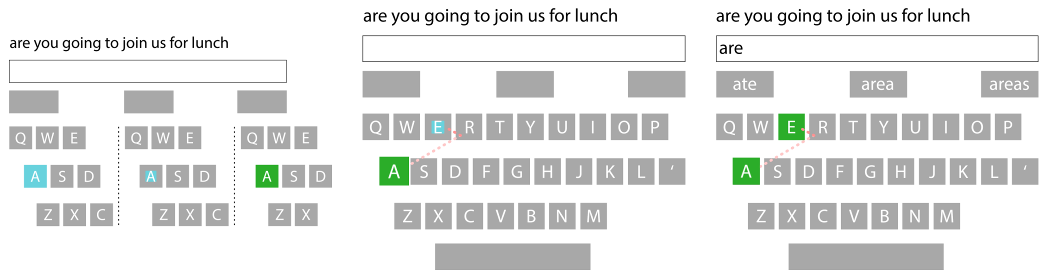
This dwell-free eye typing system thus enables eye typing without requiring the user to fixate on a key for a set dwell time. While the system decodes at a word level, it creates a quasi-sentence-level typing experience, as it treats the end of one word and the start of the next as a single gaze action on the space bar.

Data Collection

In order to fine-tune the parameters for both fixation detection and the decoder, we collected typing data in a pilot study involving four participants, including one of the authors. We set the I-VT classification threshold, used for determining the start and end fixated keys, at 30 degrees per second based on this data. Additionally, we used this data to optimize the free parameters for the decoder, such as the variances and language model scale factors.

3.2 DwellDelimitedWord

In addition to a pure dwell-free eye typing technique, we also developed a semi-dwell-free eye typing system as a basis for comparison. The behavior of this technique is illustrated in Figure 2. This system strikes a balance between traditional dwell-based systems and a completely dwell-free method. While it still operates based on the principles of gaze traces and sequential letter fixation for the intermediary letters as shown in Figure 2(b), this semi-dwell-free system incorporates brief dwell times on the start (Figure2(a)) and end keys (Figure2(c)) for each word. We used the same word gesture decoder [6] as in the dwell-free eye typing technique for this purpose. By retaining some elements of dwell-time interaction, the semi-dwell-free system provides users with a more familiar and controlled typing experience. We conjecture that it may reduce instances of unintentional key selections.



(a) Dwell at 'A' key to start typing (b) Gesture across middle keys without dwell (c) Dwell at the end key 'E' to finish

Figure 2: Illustration of the DWELLDELIMITEDWORD condition

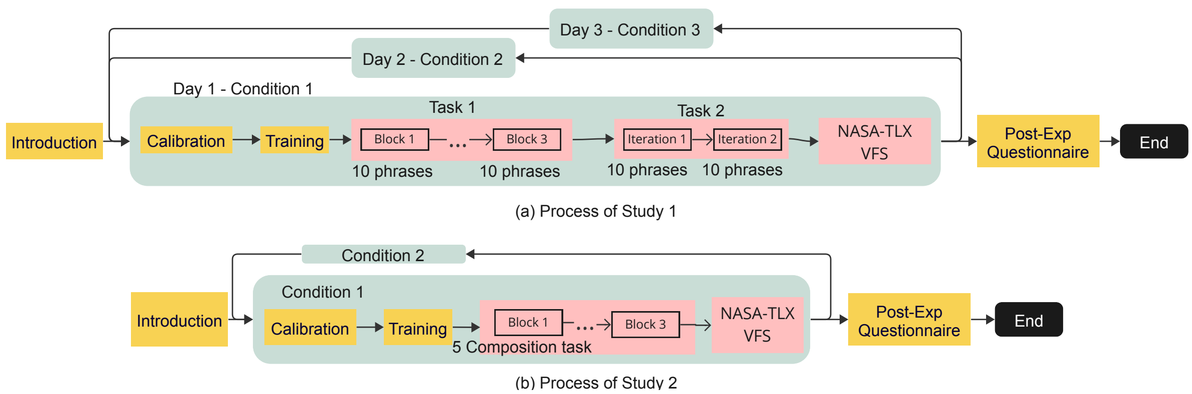


Figure 3: Procedure of Study 1 and 2

4 STUDY 1

The purpose of Study 1 was to assess the efficacy of DWELL-FREE and DWELLDELIMITEDWORD eye-typing in comparison with conventional dwell-based eye typing method. We refer to the conventional baseline method as DWELL+WORDPREDICATION. As such, this was a 3×1 design with one independent variable TYPINGMETHOD with three levels: DWELL+WORDPREDICTION, DWELLDELIMITEDWORD and DWELLFREE.

4.1 DWELL+WORDPREDICTION

In this study, the design of the baseline condition is determined by reference to a commercially available standard dwell-based keyboard [15], augmented with word suggestion functionality. In this baseline condition, the users must focus their gaze on a key for a fixed period (500 ms). Upon reaching this dwell time threshold, the system selects the character that the user's gaze is fixated on. To reduce the need for individual character selection, a list of three potential word completions that align with the characters already entered will be displayed at the top of the keyboard. The suggested words can be selected by the same dwell mechanism.

As the HoloLens 2 does not have any default system eye-gaze typing method, we chose the non-AR commercial eye typing standard as the baseline method for the following reasons. First, it is fair to use the same statistical decoder to intelligently provide word suggestions in the baseline method as this is what actual eye typing users rely on [15]. Thus, we compare our dwell-free system to a more capable baseline than the baseline mentioned in a prior paper [7], which only used dwell timeout without intelligently providing word suggestions. Second, an error-free dwell time baseline could be set as low as 400 ms for desktop applications [29]. We chose the average of 400 ms and 600 ms (which is the baseline dwell time in prior work [7]), resulting in 500 ms as a meaningful investigation point for HoloLens 2. As a commercially available standard, this dwell-based keyboard with word suggestions represents the current norm in eye typing technologies. As such, it serves as a baseline for comparison when assessing the two alternative typing methods under examination in our study.

4.2 Tasks

We used a transcription task in the first study. In this task, participants are presented with stimuli phrases at the top of the keyboard, and they are asked to transcribe these phrases as quickly and accurately as possible using the given typing method.

We used phrases from the Enron mobile message dataset's memorable phrases subset [39], filtered to include sentences of 40 or fewer characters, with at least four words consisting solely of A-Z letters and apostrophes. This ensured the phrases fit within a single line of the visible display and have comparable complexity. All punctuation was removed. We also eliminated phrases containing words that fell outside of the decoder's 64,000-word vocabulary. Unique sets of stimulus phrases were used for the three conditions but remained consistent across participants for fairness.

4.2.1 Task 1: Unique Phrases

In each condition, participants were asked to transcribe a set of 30 phrases in Task 1. These phrases were divided into three equal blocks, with a short break allowed between each block to minimize fatigue. To maintain a uniform level of difficulty across the phrase sets, the complexity of each set was balanced based on the number of characters and words per phrase.

4.2.2 Task 2: Repeated Phrases

Task 2 was designed to evaluate the proficiency of participants in the three conditions once they had achieved a degree of familiarity with each method. The task involved repetitive transcription of a single phrase. In our pilot study, participants reported that they were not able to navigate to the exact location of keys instinctively as they did with their fingers. Given that eye typing may present a steeper learning curve for able-bodied users, who are accustomed to using their fingers rather than their eyes to navigate a standard keyboard, this task aimed to assess their performance once they became more adept at using their eyes and learning effects had saturated.

4.3 Procedure

Each participant was asked to complete three conditions including the baseline condition, each on a separate day as summarized in Figure 3(a). This strategy was adopted to counter potential fatigue effects. We also took extreme care to mitigate visual fatigue during the study by giving multiple small breaks during each session. Each experimental session was designed to span approximately one hour. The order of the three conditions was fully counterbalanced across the participants.

Before the study, participants were given a demographic questionnaire including questions about their previous experience with AR/VR headsets and familiarity with any eye-typing systems. At the beginning of the first session, participants were initially familiarised with the HoloLens 2 and the three typing methods to be used in the study during a brief five-minute training period.

In each of the three sessions, participants began by typing five phrases to familiarize themselves with the current typing modality. Task 1 was then segmented into three blocks, with participants typing ten phrases per block followed by a one-minute break. Upon completing the task, participants were asked to fill out a Visual Fatigue Scale questionnaire. A break of five minutes was given to rest their eyes and to prevent fatigue-induced effects from Task 1 propagating into Task 2. Then, participants were asked to complete Task 2 using the same typing method, leveraging the proficiency gained from Task 1 in Task 2. The study delved deeper into the expert-level performance achievable with the current method in Task 2 where the participants repeatedly typed the same phrases ten times, with each iteration followed by a one-minute break. The raw NASA-TLX and Visual Fatigue Scale questionnaires were filled out again upon completing Task 2.

At the end of the third session, after completing all three conditions, participants were queried regarding their overall preference among the three conditions, as well as potential applications in their daily activities. Their general comments and feedback were also collected.

4.4 Apparatus

The virtual keyboard was placed at a distance of 1.8 m from the user. This placement was fixed across the three conditions. The individual letter keys were 0.08×0.08 m, corresponding to an angular size of 2°. The overall keyboard size was 0.38 m in height and 0.935 m in width, corresponding to angular sizes of 12° and 29° respectively.

A physical keyboard was paired with the headset via Bluetooth, which was employed solely for controlling the study's progression. Participants were instructed to press the physical *SPACE* key to display the keyboard and initiate a new phrase entry. Pressing the physical *RETURN* key submits the current phrase entry and deactivates the keyboard until the participant again pressed the *SPACE* key.

4.5 Participants

For Study 1, we recruited a total of 12 participants (4 females and 8 males), falling within the age range of 20 to 30. Among these participants, five had previous experience with VR headsets, two had familiarity with the HoloLens, but none had prior experience with eye-typing applications. Participants spent approximately three hours in total entering phrases across the three conditions, with the condition sessions distributed over three separate days. In acknowledgement of their contribution, participants were compensated with a voucher for their time.

4.6 Results

4.6.1 Task 1: Unique Phrases

The entry rate is quantified in terms of words-per-minute (WPM), where we consider a word as a sequence of five characters, spaces included. The speeds for the three conditions over three blocks in Task 1 are illustrated in Figure 4(a). A two-way repeated measures ANOVA was applied to analyze the entry rate data. Mauchly's test indicated that the assumption of sphericity had not been violated for both Condition and Block. We found a significant main effect of Condition, ($F(2,22) = 28.37$, $p < 0.05$, $\eta_p^2 = 0.433$) and Block ($F(2,22) = 28.37$, $p < 0.05$, $\eta_p^2 = 0.13$) on the text entry rate. Bonferroni-corrected pairwise comparisons showed that the DWELLFREE method was significantly faster than the two other conditions, with average typing speeds of 13.32 WPM in the first block and 15.42 WPM in the final block. Although DWELLDELIMITEDWORD had higher means across three blocks than DWELL+WORDPREDICTION, there was no significant difference. Additionally, the entry rate changed across blocks with Block 3 significantly faster than Block 1 and Block 2, indicating performance improvement across blocks. No significant interaction was found, suggesting that the effect of Condition on the entry rate did not change over Blocks.

However, we also observed a relatively high uncorrected error rate for the two gesture-based methods as we didn't provide any error correction methods or the deletion button for editing the results. In the final block, the average error rate for the three conditions is 2.31% (DWELL+WORDPREDICTION), 5.34% (DWELLDELIMITEDWORD) and 7.45% (DWELLFREE) respectively measured in character-error-rate.

4.6.2 Task 2: Repeated Phrases

The entry rates obtained in Task 2 are summarized in Figure 4(c). We also found a significant main effect of Condition ($F(2,22) = 31.03$, $p < 0.05$, $\eta_p^2 = 0.70$) and Block ($F(1,11) = 5.24$, $p < 0.05$, $\eta_p^2 = 0.01$). In Task 2, the DWELLFREE method remained to be significantly faster than both the DWELL+WORDPREDICTION and DWELLDELIMITEDWORD methods, achieving a typing speed of 16.73 WPM in the second block. Moreover, we observed that the DWELLDELIMITEDWORD method also outperformed the DWELL+WORDPREDICTION method. This suggests that, over time, participants were able to improve their typing speed using both gaze trace methods. However, there appeared to be a speed limit for the DWELL+WORDPREDICTION approach, indicating that proficiency plateaued with continued practice. There was no significant interaction effect.

4.6.3 Perceived Workload

In our analysis of the perceived workload across the three conditions, we applied the non-parametric Friedman test due to the ordinal nature of NASA-TLX data. The Friedman test revealed a significant main effect of Condition ($\chi^2(2) = 7.348, p < .05$).

To explore this effect further, post-hoc pairwise comparisons were conducted using Conover's test. Surprisingly, these tests indicated that the perceived workload in the DWELLDELIMITEDWORD condition was significantly higher than in the DWELL+WORDPREDICTION condition ($p = 0.013$). The participants reported higher ratings regarding Mental Demand, Temporal Demand, Perceived Performance, and Effort. This could be caused by the constant switching between dwelling and gesturing using

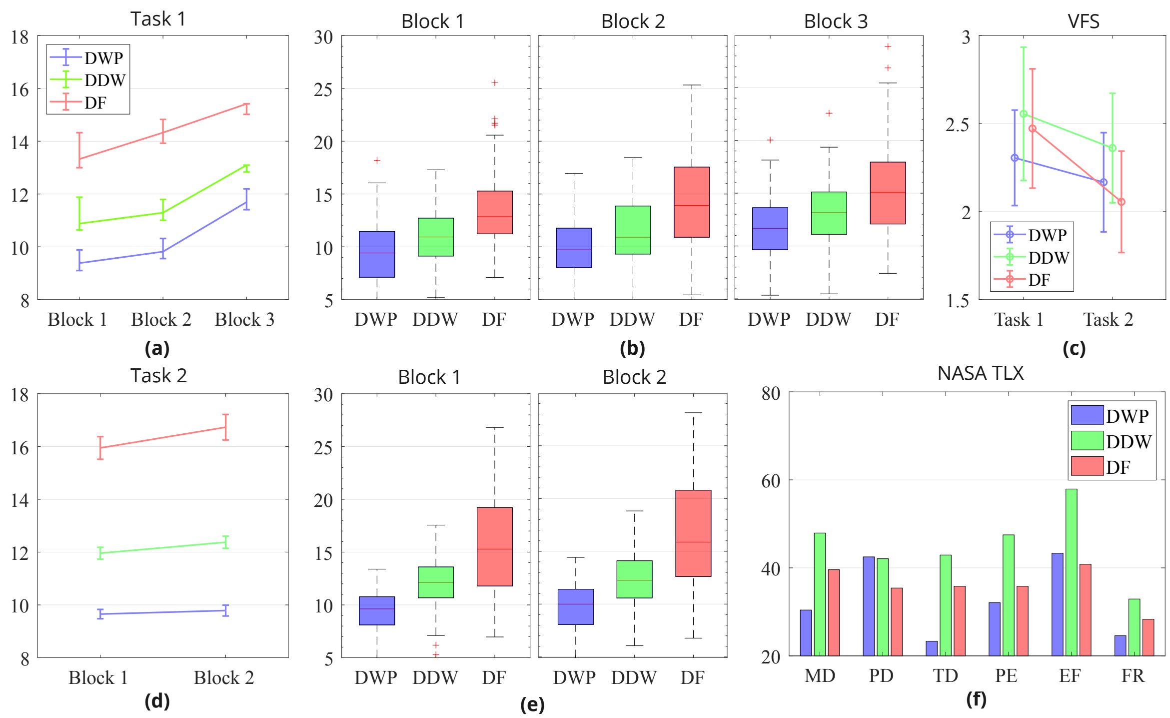


Figure 4: Results from Study 1. DWP, DDW and DF stand for Dwell+WordPrediction, DwellDelimitedWord and DwellFree respectively. Each color stands for the same condition throughout the figures. All entry rates are calculated in word-per-minute (WPM). Boxplots in (b)(e) show the first quartile (Q1) and the third quartile (Q3) with a median represented by '-'. Whiskers show the minimum and maximum values. (a):the entry rate in task 1 with line graph. (b): entry rate of task 1 shown in box plots across three bocks. (c): the results of Visual Fatigue Scale filled after task 1 and task 2 for each condition. (d): the entry rate in Task 2. (e) box plots of entry rate in task 2. (f): NasaTLX results.

eye gaze. Participants also reported higher visual fatigue by using DWELLDELIMITEDWORD method but with no significant difference. The other pairwise comparisons were not significant.

4.6.4 Summary

In summary, Study 1 showed that the baseline DWELL+WORDPREDICTION method resulted in roughly the same text entry rate as DWELLDELIMITEDWORD. DWELLFREE eye-typing resulted in significantly higher entry rates, both when typing unique and repeated phrases. However, DWELLFREE eye typing and DWELLDELIMITEDWORD also resulted in significantly higher error rates. Finally, DWELLDELIMITEDWORD induced a significantly higher perceived workload compared to the baseline DWELL+WORDPREDICTION and DWELLFREE eye typing conditions.

5 STUDY 2

Study 1 demonstrated the potential for dwell-free eye typing to deliver significantly higher entry rates while maintaining the same perceived workload as an established dwell-based commercial baseline. Building on feedback from our initial study, we iteratively refined the design of the dwell-free eye typing system with the deletion functionality. In Study 2, we investigated the feasibility of this enhanced system in a real-world scenario through a composition task [40]. Further, we investigated how to provide users with real-time feedback and facilitate adaptive responses during the eye-typing process.

5.1 Task

We used a more externally valid task to evaluate the feasibility of DWELLFREE method in a more realistic scenario. The composition tasks for this study were informed by the approach described by Vertanen and Kristensson [40]. Each task involved guided composition conditions, where participants were given certain scenarios and instructed to type a message accordingly. We specifically employed the 'Situation' stimuli to mimic the most common scenarios where users consciously formulate their responses based on the context in their minds, which is a context that the system does not have access to.

To ensure a consistent reference for the results, we adopted the dictation method described by Gaines et al. [9]. Participants were directed to articulate their intended text to the experimenter before they commenced typing their responses. This procedure aimed to keep the reference sentence intact and not influenced by the next-word predictions provided by the language model. As such, participants were asked to type exactly the same sentence as they dictated. The stimulus situation text was not displayed after the participant pressed the *SPACE* key to begin their entry.

5.2 System Refinement

Resizing the Keyboard We aimed to strike a balance between two primary considerations. Firstly, the keyboard's size needed to minimize occlusion of the user's Field of View (FoV) to preserve the immersive AR experience on the HoloLens 2 [5]. This reduced size not only prevents users from feeling overwhelmed during eye-typing but also adheres to Fitts' law, ensuring short paths between keys for faster cursor movements [36]. Secondly, the keyboard had to be sufficiently large to enable users to fixate accurately on individual keys and produce a gaze trace that the decoder could effectively interpret.

After iterative testing, we adjusted the key size from an initial 0.08×0.08 m to a final size of 0.06×0.06 m, corresponding to the visual angles of 2°32′ to 1°54′. We also adjusted the keyboard size to a height and width equal to 0.27 m and 0.7725 m respectively, or (8°34′ and 24°13′). The height-to-width ratio of the keyboard was therefore reduced to 0.35. This ratio is larger than that of a physical keyboard serving to enhance recognition accuracy, particularly in differentiating keys between horizontal rows.

Relocating the *SPACE* Key Participants in Study 1 provided feedback indicating that they felt compelled to check the resulting word and word alternatives after making the gesture. However, the typing requirement that necessitated fixation on the space key located at the bottom of the keyboard created an issue. This led to a pattern of constant back-and-forth eye movements across the keyboard layout as they were forced to look down again to restart typing by looking at the space key. This movement interrupted the desired linear flow of their gaze during dwell-free typing. Thus, we relocated the space key to the top of the keyboard. This arrangement integrated the process of checking typed results, alternative words, and stimulus phrases into the typing flow, as ending the word trajectories at the space key was directionally consistent with viewing the results and other functional content above. This strategic placement of the space key also served as a boundary, differentiating between the time-out threshold-based interaction (above) and the dwell-free, trajectory-based interaction (below), see Figure 1(f).

Adding the *DELETION* function For a more realistic user experience, we provided a degree of editing functionality by adding *Backspace* key. Users can now correct errors by deleting and retyping words.

5.3 Auto-completion

As indicated in prior research [15, 38], users tend to minimize their typing efforts when using eye typing. With the advancements in Large Language Models (LLMs), it is capable of contextually predicting and auto-completing words, phrases, or sentences as users type [38]. This approach can potentially augment dwell-free eye typing experiences, potentially boosting text entry speeds and enhancing user satisfaction.

While predictions can be beneficial, they can also be intrusive, diverting the user's attention and leading to cognitive overload [21]. The intricate cognitive operations required during eye typing, encompassing tasks like visual search, decision-making, and processing visual feedback, underscore the importance of not overwhelming users with excessive visual cues. Moreover, none of the previous research into hands-free text entry incorporated intelligent word prediction features, possibly emphasizing the judicious use of the limited FoV in MR headsets to prevent cognitive overload.

Therefore, the challenge was to develop a feedback mechanism that seamlessly integrates real-time prediction feedback with user interactions, maximizing the utility of the tool. Given these considerations, our primary exploration focused on exploring the supporting mechanism with the example of providing next-word prediction using LLMs.

In a preliminary study, we examined two potential feedback mechanisms for providing suggestions. The first mechanism involved a visual feedback approach utilizing grey next-word predictions. This is an approach used in Outlook and Word on Microsoft Windows systems. The second mechanism involved auditory feedback, where a Text-to-Speech generator read the predicted word aloud. The audio feedback method used another channel from the primary input channel in this mode of typing and demonstrated better performance and higher user satisfaction among two participants in a pilot study.

Our final system with predictions utilized a fine-tuned model from OpenAI (text-davinci-003) specializing in text correction to generate the next-word prediction in our system. The function was directly implemented on the HoloLens 2 with a network connection to send prompts and receive results. The prompts, formatted as "Give the most likely one next word only: original phrase; the next word {*with the phrases already typed*}", were sent each time a new word was typed or a word alternative was selected. Once the predicted word was received, it would be read out by the text-to-speech system, and users could select the predicted word by dwelling at the end of the current typed phrase. There was a delay of 1 to 2 seconds in voice prediction output, attributable to LLM processing and network latency.

5.4 Study Design

Study 2 used a within-subjects design consisting of two conditions. In the first condition NOPREDICTION, participants interacted with a system that included a deletion key but did not offer next-word predictions. The second condition PREDICTION mirrored the first but introduced an additional feature. Here, an audio feedback mechanism was employed to provide real-time next-word predictions to the users, aiming to examine the effect of this feature on user performance and experience.

5.5 Procedure

The study, as summarized in Figure3(b) began with introductory training on using the HoloLens 2 and the relevant typing methods. This was followed by two experimental conditions, conducted in a fully counterbalanced order to avoid order effects. At the start of each condition, careful eye calibration was undertaken, and participants were instructed to maintain a consistent headset position. The Visual Fatigue Scale (VFS) was deployed at both the beginning and end of each condition to gauge any induced eye fatigue.

Each condition commenced with a familiarization phase involving five situation-based questions. This was followed by the main composition task consisting of three blocks of five questions, with brief breaks recommended after each block.

Post-completion of each condition, participants filled out the Raw NASA-TLX questionnaire and the VFS, along with subjective ratings for comfort, learnability, perceived accuracy and speed. A 10-minute break was provided between the two conditions. In the end, we asked participants which of the two typing conditions they preferred.

5.6 Participants

An additional cohort of 16 participants (4 female, 12 male, aged between 21 to 39) was recruited for Study 2. Among these participants, 11 had previous experience with VR headsets, 4 had used HoloLens before, and only 2 had interacted with eye-gaze-related applications. None of these participants were involved in Study 1. Each participant was engaged in a single session that lasted approximately 1.5 hours and received a voucher as compensation for their time.

5.7 Results

5.7.1 Entry Rate

We performed a 2×3 factorial ANOVA to examine the entry rate across the two conditions of prediction (NOPREDICTION, PREDICTION) and three Blocks of trials. There was a significant main effect of Block ($F(2,30) = 45.16, p < .001$, $\eta_p^2 = 0.425$), which suggests that the entry rate changed over the blocks of trials. Post-hoc analysis with Bonferroni correction indicated a significant increase in entry rate from one Block to the next, suggesting a learning effect—as expected, participants' skills and efficiency at the task improved with continued practice. Every pair of the three blocks demonstrated a significant difference. By the final typing block, the entry rate had increased to 12.05 WPM and 11.70 WPM for the PREDICTION and NOPREDICTION conditions respectively. These results are likely to be surpassed with continued practice, as the participants only engaged in approximately 30 minutes of typing for each condition. The task involved a comprehensive process of reading the situation, formulating a reasonable response, dictating to the experimenter, and then typing using the designated method.

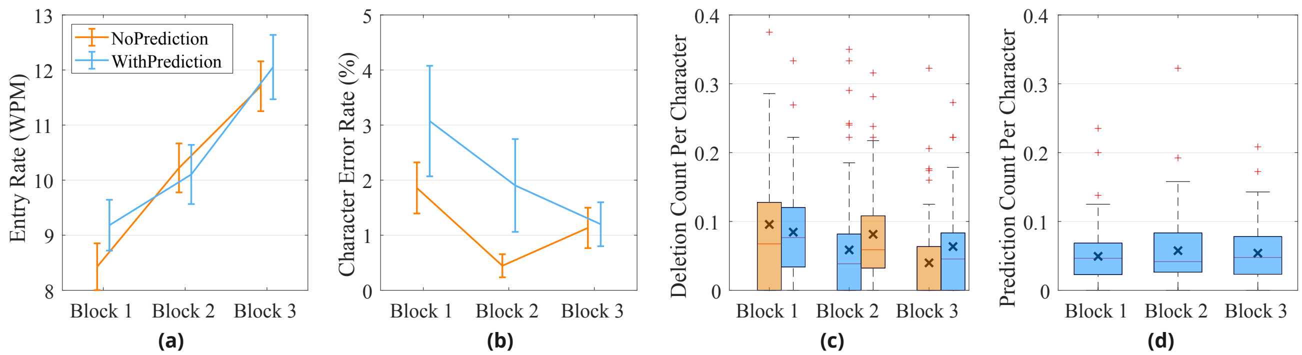


Figure 5: Results from Study 2: (a)(b) Mean text entry rate, error rate across three blocks. Error bars indicates the standard error of the mean. (c)(d) Box plots of the average deletion and prediction (only in PREDICTION condition) selected per character. The boxplots show the median (the horizontal line), the first and third quartile (the box) and the minimum and maximum (the whiskers). The mean is indicated by the '×' mark. Each color represents the same condition in all subfigures

No significant effects were identified between the two conditions or in the interaction between Block and Condition. This could potentially be due to an increased number of deletions observed in the presence of predictions, possibly triggered by the short dwell time of the prediction selection button, as indicated in Figure5(c). Another contributing factor might be the delay in receiving the prediction from the LLM. Participants might wait for the audio feedback of the next-word prediction to confirm its relevance. However, given that the audio feedback does not interfere with the typing flow, some participants disregarded delayed predictions.

5.7.2 Character Error Rate

There was no significant effect observed for character error rates. The character error rates decreased to an average of 1.19% and 1.13% in the final block, rates that are notably low and acceptable for an eye typing system, or any intelligent text entry method in general [20]. Nevertheless, this could imply that errors have been indirectly accounted for in the entry rate, as more deletions were observed. This suggests that the entry rate measurement already factors in the time spent correcting mistakes, thus the low character error rates do not indicate a lack of mistakes but efficient error correction within the typing process.

5.7.3 Deletion Count per Character

A 2 (conditions) $\times$ 3 (blocks) factorial ANOVA of deletion counts per character revealed a statistically significant effect of Block. Post-hoc analysis showed that the deletion count per character in Block 3 was significantly smaller than that of Block 1. Block 2 had no significant difference with either of the remaining Blocks. This is in line with the decrease of error rate and increase of entry rate observed.

5.7.4 Prediction Count per Character

We used a one-way ANOVA to analyze the prediction counts per character in the PREDICTION condition. There's no significant difference in the selection prediction over time.

5.7.5 Subjective Ratings

There were no significant differences in the perceived workload from NASA-TLX ratings or visual fatigue from VFS. Also, the subjective ratings on comfort, learnability, perceived accuracy, and speed show no significant differences.

6 DISCUSSION

In this paper we have presented SkiMR: the first system to allow hands-free text entry on mixed reality headsets through dwell-free eye typing. Study 1 demonstrated the viability of using three gaze-based text entry methods. We found that dwell-free eye typing resulted in a significantly faster text entry rate. Participants achieved a 15.7% increase in text entry rate after limited practice. Further practice with repeated phrased led to an additional 8.5% improvement. Study 1 used a decoder that closely mirrors real-world usage, equipped with a considerable vocabulary of 64,000 words. This wide lexicon supports diverse and naturalistic communication, allowing for a more authentic evaluation of eye-typing methods in MR headsets.

We then refined the system based on observations and insights from Study 1. Using this system, Study 2 demonstrated that this system allowed users to type at 12 WPM with a corrected error rate of 1.1%. While participants' exposure to dwell-free eye typing was short and thus their learning was unlikely to fully saturate, our results still surpass those reported for prior hands-free systems [23, 33, 43] implemented in MR headsets.

Speed Error Trade-off Comparing Study 1 and Study 2, Study 2 resulted in a 21.9% decrease in speed (from 15.4 WPM to 12.1 WPM), but also a significant 84% reduction in character error rate (CER, from 7.5% to 1.2%). This disproportionate trade-off indicates that the system's accuracy was vastly improved at a relatively low cost to speed. Note that in Study 1 we compared uncorrected error rates; that is, participants were not given the opportunity to delete or edit their entries, which would naturally lead to a higher CER. On the other hand, Study 2 involved a more externally valid composition task, as opposed to a transcription task used in Study 1. This change likely resulted in longer completion times due to the added cognitive demands of generating original text [40]. This analysis suggests that the improvement made between the two studies is indeed an advancement. The technique was refined to not just reduce errors but to also support a more complex and realistic task, which does indeed reduce the typing rate, but not necessarily in a way that negates the benefits observed in Study 1.

As a further calibration point, a recent study of dwell-free eye typing on a desktop environment with a higher screen refresh rate and a higher eye tracking sample rate reported the highest measured entry rate as 10.1 WPM [2]. In summary, this paper has demonstrated the potential for relatively fast hands-free text entry on mixed reality headsets through dwell-free eye typing.

Facilitating Auto-completions Eye typing inherently leads to fatigue due to the sensory nature of eyes, highlighting the value of auto-generated text completions to mitigate this challenge. However, introducing new visual elements can risk overloading the user cognitively, especially when introducing additional visual content through the same channel. Providing auto-correct, auto-suggest, and so on, also introduces uncertainties and can potentially also be detrimental to performance. Thus, how to deliver prediction feedback plays a vital role.

Our study found no notable speed advantage in integrating next-word prediction with dwell-free eye typing. The costs and benefits of using predictions appear to be in balance. Notably, the need to pause the gaze and dwell for predictions disrupted the continuous typing flow, adding to the overall time taken. Participants reported that typing out shorter words was faster than selecting their predicted equivalents. Another factor was the slight delay in procuring predictions, stemming from the inherent processing time of the current LLMs. Given these findings, there's room to consider more extended predictions. While our study primarily aimed to showcase the interaction mechanism supporting predictions, future research could explore the feasibility of auto-completing not just individual words but entire phrases or sentences.

In general, our approach of using auditory feedback did not introduce more cognitive workload or discomfort in the overall user experience. We see the potential for incorporating the capabilities of large language models into dwell-free eye typing for mixed reality. We see fruitful future work in exploring how to optimize such features while ensuring the overall system is kept effective and efficient.

Use Cases Participants' comments provided insightful use cases that underlined the potential advantages of our dwell-free eye typing application in the headset. A compelling scenario emerged from a participant engaged in chemistry experiments. They described the challenges faced when handling chemicals - the imperative to wear gloves and goggles for protection but also the simultaneous need to record intricate details and data. Here, a hands-free eye typing system integrated into their protective goggles could revolutionize their documentation process. Instead of pausing their work, removing protective equipment, and risking contamination, they could effortlessly make digital records in real time. This innovative application extends beyond the lab. Other participants highlighted its utility in daily scenarios where hands are preoccupied, such as enjoying a meal, relaxing while lying down, or traveling on an airplane. Such feedback underscores the system's potential to transform tasks requiring data input in both specialized and everyday contexts.

Finally, we also see promising future work in investigating means of alleviating eye fatigue and in studying the performance of dwell-free eye typing when users are non-stationary, such as when they are walking around a manufacturing plant.

7 Conclusions

In this paper we have presented SkiMR, a dwell-free eye typing system for mixed reality headsets that enables hands-free text entry with a relatively fast text entry rate and a low corrected character error rate. Our refined system enabled users to compose original text at an average text entry rate of 12 wpm with a corrected character error rate of 1.1%.

Our initial study further showed that dwell-free eye typing results in a significantly faster text entry rate than a traditional dwell-based eye typing system further supported with word prediction. In addition, dwell-free eye typing was also significantly faster than a hybrid dwell-free solution that relied on dwell timeouts to delimit individual words.

Overall, this paper demonstrates that using an eye gaze modality in combination with statistical decoder enables fast and accurate hands-free text entry for mixed reality headsets. We hope this work will stimulate further research in this area.

References

[1] S. Ahn and G. Lee. Gaze-assisted typing for smart glasses. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 857–869, 2019.

[2] W. Cui, R. Liu, Z. Li, Y. Wang, A. Wang, X. Zhao, S. Rashidian, F. Baig, I. Ramakrishnan, F. Wang, et al. Glancewriter: Writing text by glancing over letters with gaze. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2023.

[3] T. J. Dube and A. S. Arif. Text entry in virtual reality: A comprehensive review of the literature. In *Human-Computer Interaction. Recognition and Interaction Technologies: Thematic Area, HCI 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II 21*, pp. 419–437. Springer, 2019.

[4] J. Dudley, H. Benko, D. Wigdor, and P. O. Kristensson. Performance envelopes of virtual keyboard text input strategies in virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 289–300. IEEE, 2019.

[5] J. J. Dudley, K. Vertanen, and P. O. Kristensson. Fast and precise touch-based text entry for head-mounted augmented reality with variable occlusion. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 25(6):1–40, 2018.

[6] J. J. Dudley, J. Zheng, A. Gupta, H. Benko, M. Longest, R. Wang, and P. O. Kristensson. Evaluating the performance of hand-based probabilistic text input methods on a mid-air virtual qwerty keyboard. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, forthcoming.

[7] W. Feng, J. Zou, A. Kurauchi, C. H. Morimoto, and M. Betke. Hgaze typing: Head-gesture assisted gaze typing. In *ACM Symposium on eye tracking research and applications*, pp. 1–11, 2021.

[8] C. R. Foy, J. J. Dudley, A. Gupta, H. Benko, and P. O. Kristensson. Understanding, detecting and mitigating the effects of coactivations in ten-finger mid-air typing in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–11, 2021.

[9] D. Gaines, P. O. Kristensson, and K. Vertanen. Enhancing the composition task in text entry studies: Eliciting difficult text and improving error rate calculation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–8, 2021.

[10] D. W. Hansen, H. H. Skovsgaard, J. P. Hansen, and E. Møllenbach. Noise tolerant selection by gaze-controlled pan and zoom in 3d. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pp. 205–212, 2008.

[11] R. Hedeshy, C. Kumar, R. Menges, and S. Staab. Hummer: Text entry by gaze and hum. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–11, 2021.

[12] R. J. Jacob and K. S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*, pp. 573–605. Elsevier, 2003.

[13] F. Kern, F. Niebling, and M. E. Latoschik. Text input for non-stationary xr workspaces: Investigating tap and word-gesture keyboards in virtual and augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2658–2669, 2023.

[14] P. O. Kristensson. *Discrete and continuous shape writing for text entry and control*. PhD thesis, Institutionen för datavetenskap, 2007.

[15] P. O. Kristensson, M. Mjelde, and K. Vertanen. Understanding adoption barriers to dwell-free eye-typing: Design implications from a qualitative deployment study and computational simulations. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, pp. 607–620, 2023.

[16] P. O. Kristensson and K. Vertanen. The potential of dwell-free eye-typing for fast assistive gaze communication. In *Proceedings of the symposium on eye tracking research and applications*, pp. 241–244, 2012.

[17] P.-O. Kristensson and S. Zhai. Shark2: a large vocabulary shorthand writing system for pen-based computers. In *Proceedings of the 17th*

annual ACM symposium on User interface software and technology, pp. 43–52, 2004.

[18] C. Kumar, R. Hedeshy, I. S. MacKenzie, and S. Staab. Tagswipe: Touch assisted gaze swipe for text entry. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020.

[19] A. Kurauchi, W. Feng, A. Joshi, C. H. Morimoto, and M. Betke. Swipe&switch: Text entry using gaze paths and context switching. In *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 84–86, 2020.

[20] M. LaLomia. User acceptance of handwritten recognition accuracy. In *Conference companion on Human factors in computing systems*, pp. 107–108, 1994.

[21] Y. Liu, B.-S. Lee, and M. J. McKeown. Robust eye-based dwell-free typing. *International Journal of Human–Computer Interaction*, 32(9):682–694, 2016.

[22] Y. Liu, C. Zhang, C. Lee, B.-S. Lee, and A. Q. Chen. Gazetry: Swipe text typing using gaze. In *Proceedings of the annual meeting of the australian special interest group for computer human interaction*, pp. 192–196, 2015.

[23] X. Lu, D. Yu, H.-N. Liang, and J. Goncalves. itext: Hands-free text entry on an imaginary keyboard for augmented reality systems. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pp. 815–825, 2021.

[24] X. Lu, D. Yu, H.-N. Liang, W. Xu, Y. Chen, X. Li, and K. Hasan. Exploration of hands-free text entry techniques for virtual reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 344–349. IEEE, 2020.

[25] I. S. MacKenzie and X. Zhang. Eye typing using word and letter prediction and a fixation algorithm. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pp. 55–58, 2008.

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

[27] P. Majaranta and K.-J. Räihä. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pp. 15–22, 2002.

[28] A. Markussen, M. R. Jakobsen, and K. Hornbæk. Vulture: a mid-air word-gesture keyboard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1073–1082, 2014.

[29] M. E. Mott, S. Williams, J. O. Wobbrock, and M. R. Morris. Improving dwell-based gaze typing with dynamic, cascading dwell times. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 2558–2570, 2017.

[30] A. Olsen. The tobii i-vt fixation filter. *Tobii Technology*, 21:4–19, 2012.

[31] P. Panwar, S. Sarcar, and D. Samanta. Eyeboard: A fast and accurate eye gaze-based text entry system. In *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, pp. 1–8. IEEE, 2012.

[32] D. Pedrosa, M. D. G. Pimentel, A. Wright, and K. N. Truong. Filteryedping: Design challenges and user performance of dwell-free eye typing. *ACM Transactions on Accessible Computing (TACCESS)*, 6(1):1–37, 2015.

[33] V. Rajanna and J. P. Hansen. 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*, pp. 1–10, 2018.

[34] V. Rajanna, M. Russel, J. Zhao, and T. Hammond. Presstapflick: Exploring a gaze and foot-based multimodal approach to gaze typing. *International Journal of Human-Computer Studies*, 161:102787, 2022.

[35] S. Sarcar, P. Panwar, and T. Chakraborty. Eyek: an efficient dwell-free eye gaze-based text entry system. In *Proceedings of the 11th asia pacific conference on computer human interaction*, pp. 215–220, 2013.

[36] I. Schuetz, T. S. Murdison, K. J. MacKenzie, and M. Zannoli. An explanation of fitts' law-like performance in gaze-based selection tasks using a psychophysics approach. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019.

[37] M. H. Urbina and A. Huckauf. Alternatives to single character entry and dwell time selection on eye typing. In *Proceedings of the 2010 symposium on eye-tracking research & applications*, pp. 315–322, 2010.

[38] S. Valencia, R. Cave, K. Kallarackal, K. Seaver, M. Terry, and S. K. Kane. "the less i type, the better": How ai language models can enhance or impede communication for aac users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2023.

[39] K. Vertanen and P. O. Kristensson. A versatile dataset for text entry evaluations based on genuine mobile emails. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pp. 295–298, 2011.

[40] K. Vertanen and P. O. Kristensson. Complementing text entry evaluations with a composition task. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 21(2):1–33, 2014.

[41] D. J. Ward, A. F. Blackwell, and D. J. MacKay. Dasher—a data entry interface using continuous gestures and language models. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*, pp. 129–137, 2000.

[42] D. J. Ward and D. J. MacKay. Fast hands-free writing by gaze direction. *Nature*, 418(6900):838–838, 2002.

[43] W. Xu, H.-N. Liang, Y. Zhao, T. Zhang, D. Yu, and D. Monteiro. Ringtext: Dwell-free and hands-free text entry for mobile head-mounted displays using head motions. *IEEE transactions on visualization and computer graphics*, 25(5):1991–2001, 2019.

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

[45] S. Zhai and P.-O. Kristensson. Shorthand writing on stylus keyboard. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 97–104, 2003.