Exploring Gaze for Assisting Freehand Selection-based Text Entry in AR

MATHIAS N. LYSTBÆK, Aarhus University, Denmark
KEN PFEUFFER, Aarhus University, Denmark and Bundeswehr University Munich, Germany
JENS EMIL GRØNBÆK, Aarhus University, Denmark
HANS GELLERSEN, Lancaster University, UK and Aarhus University, Denmark

With eye-tracking increasingly available in Augmented Reality, we explore how gaze can be used to assist freehand gestural text entry. Here the eyes are often coordinated with manual input across the spatial positions of the keys. Inspired by this, we investigate gaze-assisted selection-based text entry through the concept of spatial alignment of both modalities. Users can enter text by aligning both gaze and manual pointer at each key, as a novel alternative to existing dwell-time or explicit manual triggers. We present a text entry user study comparing two of such alignment techniques to a gaze-only and a manual-only baseline. The results show that one alignment technique reduces physical finger movement by more than half compared to standard in-air finger typing, and is faster and exhibits less perceived eye fatigue than an eyes-only dwell-time technique. We discuss trade-offs between uni and multimodal text entry techniques, pointing to novel ways to integrate eye movements to facilitate virtual text entry.

CCS Concepts: • **Human-centered computing** → **Mixed / augmented reality**; **Pointing**; **Interaction design theory, concepts and paradigms**.

Additional Key Words and Phrases: augmented reality, eye-tracking, gaze interaction, multimodal UI, text entry, virtual keyboard

ACM Reference Format:
Mathias N. Lystbæk, Ken Pfeuffer, Jens Emil Grønbæk, and Hans Gellersen. 2022. Exploring Gaze for Assisting Freehand Selection-based Text Entry in AR. *Proc. ACM Hum.-Comput. Interact.* 6, ETRA, Article 141 (May 2022), 16 pages. https://doi.org/10.1145/3530882

1 INTRODUCTION

Modern Head Mounted Displays (HMDs) for Augmented Reality (AR) offer users intuitive interaction via freehand mid-air gestures. A key challenge for such interfaces is text entry, e.g., in tasks such as filling in forms or annotating the real-world environment. Without a physical keyboard or a mobile device, users can perform *'AirTaps'*, where they engage their fingers in the air for directly pressing the keys of a virtual keyboard. However, a major challenge is physical fatigue, as gesturing in mid-air and lack of haptic feedback can be demanding for continuous tasks with high selection frequency like text entry [10, 21, 39].

During typing, users naturally lock their eyes on the key before the manual acquisition. This is a common phenomenon in eye-hand coordination, to obtain information for on-line planning and correction of hand movements [1, 43]. Particularly in mid-air text entry, the user's gaze often remains on target to confirm that the finger reached the correct depth of the keyboard to trigger selection. Thus, in many cases, gaze is already part of the task, representing an opportunity to cascade gaze with manual input to enhance typing.

Authors' addresses: Mathias N. Lystbæk, mathiasl@cs.au.dk, Aarhus University, Denmark; Ken Pfeuffer, ken@cs.au.dk, Aarhus University, Denmark and Bundeswehr University Munich, Germany; Jens Emil Grønbæk, jensemil@cs.au.dk, Aarhus University, Denmark; Hans Gellersen, h.gellersen@lancaster.ac.uk, Lancaster University, UK and Aarhus University, Denmark.

2022. 2573-0142/2022/5-ART141 $15.00
https://doi.org/10.1145/3530882

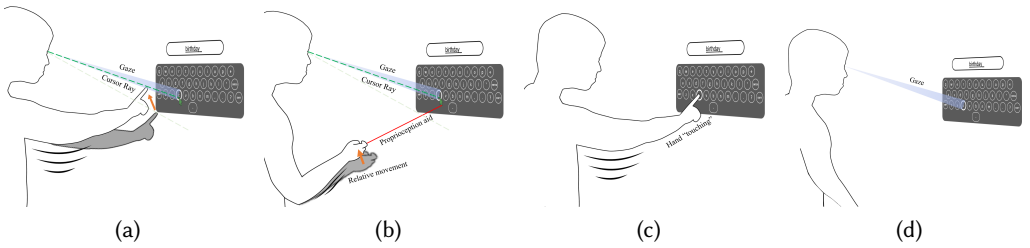


Fig. 1. Text entry using *S-Gaze&Finger* (a), *S-Gaze&Hand* (b), *AirTap* (c), and *Dwell-Typing* (d).

Inspired by this, we investigate how gaze can integrate naturally as an interaction technique for text entry. Our conceptual point of departure is the recently introduced *Gaze-Hand Alignment* [22] as an interaction principle for target selection. The idea is to use two pointers – the eye-gaze and a hand-controlled ray. When the user aligns both in their line of sight, the user selects the target that *visually* lies behind the alignment point. For example, to use the index finger for alignment with the gaze-focused key on the virtual keyboard (Figure 1a). The user (1) looks at the key of interest, and (2) aligns the tip of their index finger in line of sight with the target to immediately invoke a selection command.

An advantage is that users are free to choose the distance of the finger to the keyboard, as the concept relies on selection in the image plane. This would eliminate the finger forward movement into depth otherwise required to press virtual buttons, and with it render mid-air typing less physically demanding. However, in our first tests we found that a direct application of the prior method (that was designed for remote interaction) to the close-by virtual keyboard led to a high error rate. This is because naturally during text entry, the user's eyes are rapidly glancing over the keyboard area – often over the manual cursor. As selections trigger instantly at alignment, it led to selecting letters by accident.

We devised an advanced alignment selection algorithm, dubbed *Sustained Gaze-Hand Alignment* (S-GHA), that optimises the original *Gaze-Hand Alignment* algorithm for high-performance mid-air text entry tasks by *sustaining* the time window for alignment. This is based on a two-step algorithm, of a 150*ms* gaze dwell-time followed by an alignment dwell-time of 150*ms*. This minimises accidental selection while keeping the time short enough to avoid the typical pitfalls associated with long dwell-times. We further optimised the selection with a smarter keyboard layout, outlier removal from jittery eye or hand motion, and visual feedback to complete the interaction design.

To assess the performance of users with the proposed selection algorithm, we conducted a user study of a standard text entry task. The study includes the following four conditions, as illustrated in Figure 1. The first two are *S-Gaze&Finger* and *S-Gaze&Hand*, two techniques of [22] optimised with S-GHA for text entry. Both techniques are based on alignment with gaze but distinguish themselves in manual input usage. (1a) *S-Gaze&Finger* uses the tip of the index finger for alignment. (1b) *S-Gaze&Hand* uses a cursor for alignment that is controlled through the handheld in a more comfortable position offset from the field of view (FOV). The manual baseline is (1c) *AirTap*, a state-of-the-art text entry method in AR (e.g., standard in the Microsoft HoloLens 2) where users select by mid-air tapping with their index fingertip. The eyes-only baseline is (1d) *Dwell-Typing*, where users fixate a target for 600*ms* and that has been a baseline in recent work [9]. This allows us to set the multimodal techniques in contrast to a fully eye-based technique that eliminates physical effort, and a fully hand based technique without gaze input. 16 participants typed in text phrases

on a standard sized virtual keyboard with the techniques, while we recorded words-per-minute (WPM), error rates, and user feedback.

Our results indicate the advantage of the optimised *S-Gaze&Finger* technique. It leads to a significant reduction of effective physical motion by more than 50%, without significant performance differences in speed or error compared to *AirTap*. Surprisingly, questions on physical effort did not indicate a perceived difference in physical effort, although user feedback indicated that it is more taxing than the gaze-based techniques. We also found that both *S-Gaze&Finger* and *AirTap* lead to significantly higher WPM than *S-Gaze&Hand* and *Dwell-Typing*. No differences were revealed w.r.t. error rates. For *Dwell-Typing*, users rated significantly lower physical effort, but at the compromise of higher eye fatigue. Taken together, it suggests that *S-Gaze&Finger* is a viable alternative to mid-air text entry, that substantially reduces the physical movement requirements, without significant compromises in performance or eye fatigue. Our research informs designers and practitioners of AR interfaces on how the gaze modality can assist the predominant manual controls, and potentially render mid-air gestures easy to use by a substantial reduction of physical movement.

In sum, our contributions in this paper include the following points.

- We describe S-GHA, that represents an optimisation of the prior alignment concept of [22] for AR text entry, by implementing a novel combination of temporal and spatial parameters.
- We provide detailed empirical data and quantify user performance of a text entry user study that for the first time compares two multimodal alignment techniques with mid-air gesturing and eyes-only dwell-time techniques in AR.
- We demonstrate the advantage of one technique, *S-Gaze&Finger*, to interact with less physical movement without significant performance penalties, pointing to a novel AR text entry mechanism.

2 RELATED WORK

In this work, we are interested in the design of selection triggers, hence, we focus solely on selection-based text entry without word prediction. We provide an overview of related text entry interaction techniques in AR/VR and techniques that use gaze, and outline prior work on interaction techniques that use eyes and hands in combination.

2.1 Text Entry in Virtual and Augmented Reality

Selection-based text entry focuses on the base performance of the text entry without relying on word prediction [26, 39]. Text entry research in VR / AR have investigated controller input devices [39, 47], freehand tracking [8, 39, 47], head-tracking [21, 39, 47, 48], and combinations thereof [47]. Speicher et. al. compared freehand, controller and head based pointing techniques for text entry, and found that the best performance was achieved by ray casting using controllers onto a QWERTY keyboard [39]. While perceived as intuitive, erroneous hand tracking and physical fatigue affected the user performance for freehand typing, indicating room for improvement. Researchers also considered head-pointing based interaction techniques and found that a particular *GestureType* technique, a word-level method, performed the best by the end of the study [48]. For AR, Xu et al. assessed four text entry techniques and found, similar to Speicher et. al., that controller interaction outperformed the device-free techniques such as freehand pointing with palm-gesture for confirmation [47]. iText [21] is similarly motivated as our work, to combat arm fatigue of hand gestures. They inspected hands-free text entry via gaze interactions in imaginary keyboards, finding pointing with gaze input to be viable alternatives. In sum, these prior works focused on controller, head-pointing, or invisible keyboards. We extend the work through investigation of gaze to assist manual text entry.

2.2 Gaze-based Text Entry

Eye typing has a long history in gaze HCI research, as it is a main interaction medium for people only capable of moving their eyes [25]. Most applications are based on dwell-time, that addresses the Midas Touch problem of selecting everything one sees by only invoking a UI command when users look at a target for long enough [12]. It can be employed as a fast and comfortable user input device for the control of computers [13, 36]. Various dwell-time thresholds have been evaluated from 150ms to 1500ms [3, 12, 27, 36, 42]. While depending on the context, lower times are susceptible to false-positive selections whereas higher thresholds introduce a longer waiting time for the user. Adjustable and cascading dwell-times allow to adapt to user expertise and increase performance over time [24, 28]. The technique is often employed in studies as a baseline of an eyes only input technique [9, 18, 33, 40]. For example, Feng et al. have recently studied a head- and gaze-based typing interface through setting it in contrast to a 600ms dwell-time technique [9]. The addition of head gestures led to higher efficacy and user satisfaction in text entry. We contribute to the empirical knowledge of multimodal text entry with a study on gaze-hand typing that uncovers advantages over dwell-time and mid-air hand pointing techniques.

Beyond dwell-time, there are various other gaze-based concepts for text entry. Dasher [44] is a technique that operates on gaze direction alone and can be faster than dwell-time [35]. Context Switching duplicates the keyboard layout so that users can select via simply using the last fixated key before looking at the other keyboard [41]. Access to each character is also possible through gaze-path gestures formed by multiple saccades but comes with a high learning curve [46]. These approaches are based on adapting to a different keyboard design. Gaze gestures [7] have been inspected for text entry, by techniques that are based on saccades from the key toward a specific direction to trigger a selection [11, 28]. Complementary to selection based text entry is word-gesture typing, where users draw lines on the keyboard to enter whole words [49]. Several researchers investigated gaze paths on keyboards for fast word entry [15, 16, 29]. We extend the prior art of eye typing, with an investigation of gaze interactions to enhance selection-based text entry with mid-air gestures.

2.3 Eye-Hand Coordination and Alignment of Multiple Pointers

Humans rely on eye-hand coordination when interacting with objects in the real world; gaze is involved in the planning of hand movements by fixating on objects before reaching for them [14, 19, 37, 45]. This has been exploited in several multimodal interaction techniques, e.g., used for warping and shifting input pointers [30, 31, 50]. For example in Manual and Gaze Input cascaded (MAGIC), gaze is injected to reduce the physical effort of mouse cursor pointing, by warping the cursor implicitly to the visual area [50]. We follow a similar spirit, i.e., how gaze can assist mid-air text entry in the physical motion requirements.

Early work on using the alignment of pointers include Toolglass and magic lenses, where a bimanual interface allows the use of see-through tools laid over objects in one hand, and selection via mouse input of the other [4]. In 3D virtual environments, perspective-based pointing utilises the occlusion of objects as a pointing mechanism [2, 20], where a ray is cast from the eye position, through the hand or handheld device, outwards to the scene. However, to confirm selection, an explicit delimiter such as a pinch gesture is required to avoid the Midas Touch problem (e.g., [34]). We employ bi-modal pointer alignment as the delimiter, which reduces the reliance on physical trigger actions with the hand.

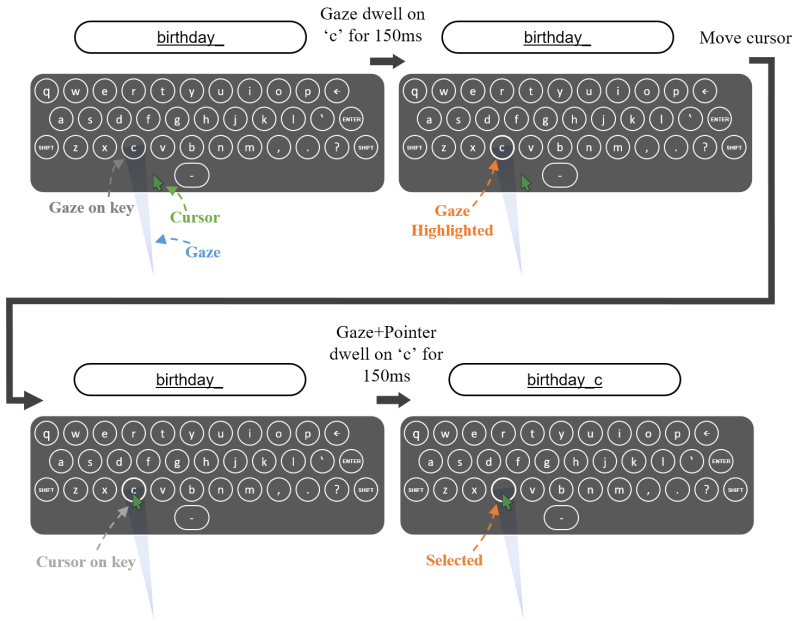


Fig. 2. The keyboard design and an example of the steps involved in S-GHA selection.

3 DESIGN OF GAZE & HAND TEXT ENTRY TECHNIQUES

In this section, we describe the operation, our process of parameter finding, and the resulting selection algorithm.

3.1 User Operation

The main sub-task of text entry is the selection of each button on the keyboard. With both techniques, a user can do this through fixating on each key with gaze and then bringing the mouse cursor on the same key (cf. Figure 2)). Re-selection of a key can be performed by a second alignment on the same key, i.e., by flicking the manual cursor to a different position on the keyboard and then returning (cf. 3). Both techniques differ in the way the manual cursor is controlled.

3.1.1 S-Gaze&Finger. The technique, that is based on *Gaze&Finger* [22], enables selection by gazing upon the key and then bringing the index fingertip into the line of sight (Figure 1a). After looking at the key, it will get highlighted. After pointing to the key while still looking at the key, it is selected and the letter is entered in the text field. The user can hold their finger at any depth level between the eyes and the keyboard to perform selections. Closer distance to the eyes decreases potential hand movement, and closer distance to the keyboard introduces less parallax of the finger when interacting over multiple depths.

3.1.2 S-Gaze&Hand. Extended from *Gaze&Finger* [22], this technique allows users to control the manual cursor via spatial hand motion in a comfortable area outside the FOV as illustrated Figure 1b. The cursor is translated by computing the change in the user's hand position between frames, which is then mapped to move the cursor in a 2:1 ratio, i.e. the cursor moves twice the speed of the user's hand. As the HoloLens 2 hand-tracking has some amount of noise, a simple filter of hand motion below 0.0001 meters worked well. When the user's hand is out of tracking range, the mouse cursor is set to default to a position in the middle above the keys. The mouse cursor is

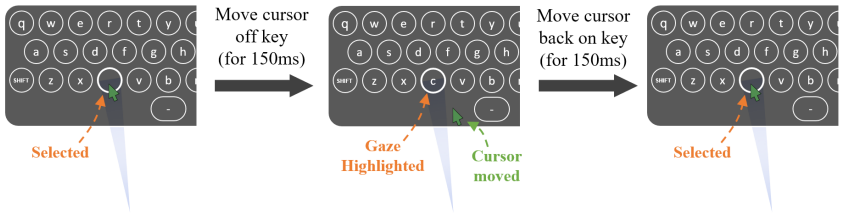


Fig. 3. Illustrating the special case of key re-selection.

constrained to only move within the border of the keyboard, so it does not fall outside its use case. We added a line drawn from the user's palm to the cursor, see Figure 1b, adding feedback in the user's periphery and contributing to understanding the spatial arrangement of the hand.

3.2 From *Gaze-Hand Alignment* to S-GHA

We conducted an iterative design method that involved rapid prototyping and pilot testing with the Microsoft HoloLens 2 headset. The goal is to find spatial and temporal parameters of the selection algorithm, that strive for a balance between speed *vs.* error rate, that is comfortable to use with the user's hands and eyes.

First tests with the direct implementation of *Gaze-Hand Alignment*'s *Gaze&Finger* and *Gaze&Hand* techniques on a virtual keyboard led to accidental typing when used in such a button-heavy interface. The original algorithm required only one frame of alignment of both pointers over an object for confirmation. The eyes often glanced over the keyboard, moving between text field and keyboard areas. In those instances, accidental activations happened when the eyes co-located with the manual pointer.

We tested several dwell-times to see whether we can reduce the erroneous selections through requiring users to fixate on a target for longer. For example, a time window of 300*ms* where both modalities need to be in alignment to trigger a selection command. This made the operation easy to use and error free, making the usability potential of the method apparent. But the added dwell-time also decreased the text entry rate. The eyes are quickly on target, but the timer only begins when the manual pointer arrives.

We considered a better fit for the actual dwell-time, i.e., the total time that was needed from gaze on a key to the selection of the key. We attempted to make the dwell-timer only count on the gaze modality so that the timer already begins on gaze onset. Meaning, after 300*ms* the user simply had to align their manual input in one moment to select. However, this ended up resulting in manual errors, when the hand pointer overshot the target as of too much speed. When you overshot the target with your mid-air aim, and then return to select it. As the manual pointer was already on-target the first time and selected. As users returned to the target, they triggered a second (unwanted) selection. We found a good compromise is provided by using two dwell-times for the eye and hand separately. Dwell-time on the key for 150*ms* for it to be highlighted, and another 150*ms* bi-modal alignment dwell-time afterwards, before selection occurs. This supports a lower net time, as gaze selection can commence before manual input arrives.

3.3 Key Selection Algorithm: S-GHA

The final key selection algorithm S-GHA extends the prior method of *Gaze-Hand Alignment* [22] in several ways. Particularly, in S-GHA, we sustain the time window for alignment by providing a 150*ms* gaze dwell-time followed by a gaze-hand alignment dwell-time of 150*ms*. The basic algorithm for *S-Gaze&Finger* and *S-Gaze&Hand* is the same, whereas they only differ in terms of how the user's hand input is incorporated. Algorithm 1 provides a pseudocode implementation.

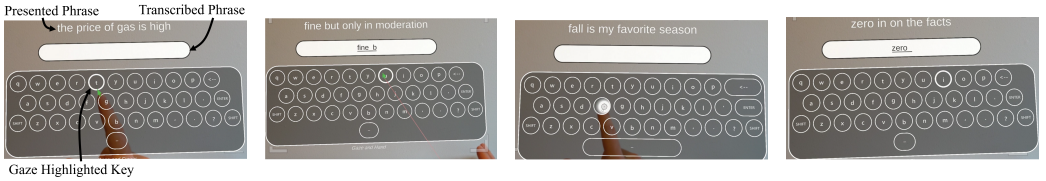
Algorithm 1: Sustained GHA (S-GHA)

```
1  while True do                                          // Represents the Update loop of Unity
2      if GazeOnTarget then
3          GazeDwellTime += FrameTime                     // The time to render the previous frame
4          OffTargetDwellTime ← 0
5      else
6          OffTargetDwellTime += FrameTime
7          if OffTargetDwellTime ≥ 150ms then             // Reset dwell-time after 150ms off target
8              GazeDwellTime ← 0
9              AlignmentDwellTime ← 0
10         end
11     end
12
13     if TimeSinceLastSelection < 150ms then             // Skip if last selection was 150ms ago
14         continue
15     end
16     if GazeDwellTime ≥ 150ms then                      // Wait for gaze dwell-time
17         if PointerOnTarget then
18             AlignmentDwellTime += FrameTime
19         end
20         if AlignmentDwellTime ≥ 150ms then             // Wait for alignment dwell-time
21             Fire Selection Event
22         end
23     end
24 end
```

Temporal Parameters. A gaze dwell-time of 150*ms* is used as a prerequisite and to highlight keys that are fixated upon (Algorithm 1, line 16) as keys in-between two letters would otherwise all highlight. A 150*ms* alignment dwell-time (Algorithm 1 follows, line 20) that starts after the gaze dwell-time is finished. We wait 150*ms* before resetting the dwell-timers (Algorithm 1, lines 7), to tolerate gaze and hand tracking outliers. Furthermore, when moving the manual pointer across the keyboard, users can overshoot to the next target and then return to the intended target, which could potentially lead to accidental selections. For this, we added a rule that the next key can only be selected 150*ms* after entering the previous key (Algorithm 1, line 13). Thus, re-selection of keys can happen by moving the manual pointer out of alignment for 150*ms* and back into alignment for 150*ms*.

Spatial Parameter. The size of the keys was fixed to be similar to that of the system's native virtual keyboard. To minimise potential gaze inaccuracy issues, we employ a snapping mechanism to always target the closest key to the user's gaze point [32]. This avoids potential selection error when a gaze sample falls in the small margins between keys. The button size of the space, backspace, and enter keys are decreased to be consistent with standard keys, such that an alignment selection only happens if the manual trigger is close to the user's gaze ray.

Feedback. We settled on doubling the border width of the keys when the first gaze dwell-time is completed to indicate a soon selection. Next, when the alignment dwell-time is complete, the key is turned white then quickly fully transparent, the letter is hidden, and double border width remains until dwell-time is reset – this closely resembles the feedback of mid-air taps with the finger. Additionally, a mouse cursor is shown for both *S-Gaze&Finger* and *S-Gaze&Hand* to visualise the manual pointer, which helps users to immediately grasp the manual input point. Audio feedback



(a) *S-Gaze&Finger*. (b) *S-Gaze&Hand*. (c) *AirTap*. (d) *Dwell-Typing*.

Fig. 4. Examples of how the techniques looked in the user study. (a) shows *S-Gaze&Finger* with the green cursor at the tip of the index finger. (b) shows *S-Gaze&Hand* with the green cursor at the end of the proprioception aid ray and hand in button right corner. (c) shows *AirTap* with finger pressing into a key. (d) shows *Dwell-Typing* with no hand visible as only gaze was used.

as the system's default click sound was added to signal successful selection in addition to the visual feedback.

4 USER STUDY

The goal of this study is to better understand user performance of typing in AR with *Gaze-Hand Alignment*, and compare our novel algorithm S-GHA to state-of-the-art selection-based text entry for HMDs. We compare both direct (*S-Gaze&Finger*) and indirect (*S-Gaze&Hand*) input variants against two baseline interaction techniques implemented for typing in a standard text entry task.

4.1 Apparatus

The techniques and study are implemented with Unity v2020.3.13f1 with the Mixed Reality ToolKit (MRTK) on the Microsoft HoloLens 2. This AR HMD has a resolution of 1440×936 per eye, a $43° \times 29°$ degree FOV, and supports hand- and eye-tracking ($1.5°$ visual angle accuracy). The keyboard layout is replicated from the native keyboard that the system provides with the HoloLens 2, with some changes to better fit the keyboard inside the FOV. The keys are 0.025 meters in size and the distance between the keyboard and the user is approximately 0.47 meters, similar to that of the native HoloLens 2 keyboard.

4.2 Task

The user was tasked to transcribe phrases with each of the text entry techniques. Users were instructed to perform as fast and as accurately as possible. During the task, users were allowed to correct sentences via the backspace button in the keyboard. Phrases were randomly selected from a subset of McKenzie's phrase set [23]. As in [39], we filtered the phrase set to those with length within 20 to 28 characters.

4.3 Study Design

The study used a within-subject protocol with one independent variable *technique* (*S-Gaze&Finger*, *S-Gaze&Hand AirTap*, *Dwell-Typing*) and six dependent variables (Words Per Minute (WPM), Total Error Rate, Hand Movement, TLX, Borg CR10 (Arm), and an adapted Borg CR10 for eye fatigue). The order of the techniques was counterbalanced through a Latin square. In sum, this results in: 16 participants $\times$ 4 input methods $\times$ 8 phrases = 512 trials

4.4 Conditions

We evaluated the following four techniques for typing phrases in AR using a virtual keyboard designed to resemble the HoloLens 2 UI.

- *S-Gaze&Finger*: select keys by fixating on a key and afterwards aligning their gaze and fingertip on the key.
- *S-Gaze&Hand*: select keys by fixating on a key and afterwards aligning their gaze and a cursor on the key, moving the cursor with relative hand motion.
- *AirTap*: select keys by "physically" pressing into them, as if pressing a button with depth in real life.
- *Dwell-Typing*: select keys by fixating on them for 600*ms*. Other timers imply different speed-error trade-offs. We decided to use the same timer as employed by Feng et al., as a best-paper award paper published at ETRA'21 represents a solid baseline [9]. We also improved it with the same snapping mechanism of S-GHA, i.e. snapping to the nearest key to limit eye-tracking inaccuracy.

4.5 Participants

16 volunteers participated in the experiment with ages ranging from 20 to 35 years old ($M = 27.5, SD = 4.82$). 8 were female, two wore glasses, and one wore contacts. Their background was mostly technical, including technical students and faculty members of the local university. On a scale between 1 (no experience) to 5 (expert), participants rated themselves as low-medium experienced with VR/AR ($M = 2.19, SD = 1.11$), low experienced with gaze interaction ($M = 1.81, SD = 1.05$), low experienced with 3D hand gestures ($M = 1.88, SD = 0.89$), and low experienced with 3D text entry ($M = 1.38, SD = 0.72$).

4.6 Procedure

Users first filled out a consent form and a demographics questionnaire and were briefed about the study. Users then wore the HMD and calibrated the eye tracker to begin the study. Before each condition, users watched a short video of typing with the technique and subsequently entered two phrases for training purposes. During the training, users were assisted by the experimenter who explained how the technique works. After completing the training session, the users continued with the study session using the given technique for another 8 phrases while performance was logged. After each condition, users filled out the questionnaires. The study concluded with a ranking questionnaire and an interview.

4.7 Evaluation Metrics

We collected data through performance logging and subjective questionnaires. We logged the following three measures for performance.

- Words Per Minute (WPM): measured as $\frac{(|T|-1)}{S} \cdot 60 \cdot \frac{1}{5}$ (see [47]), where one word consists of five entered characters, T and S denoting the overall number of letters and time as measured from first to last letter entry.
- Total Error Rate (TER): this is the sum of the not corrected error rate (NCER) and the corrected error rate (CER) [38].
- Hand Movement: the average hand movement per technique to measure how much the user moved their hand during the study. Computed as the difference in hand position between two frames, averaged over phrases.

We further used the following four questionnaires for subjective measures.

Technique	WPM	TER	Hand Movement	Mental Demand	Physical Demand	Arm Fatigue	Eye Fatigue
S-Gaze&Finger	II: 10.66	I: 2.9%	I: 2.07m	II: 2.88	III: 4.06	III: 2.94	II: 2.0
S-Gaze&Hand	IV: 9.49	II: 2.98%	II: 2.41m	III: 3.63	II: 3.63	II: 2.28	III: 2.38
AirTap	I: 11.37	IV: 3.54%	III: 5.66m	I: 2.13	IV: 4.56	IV: 3.75	I: 1.59
Dwell-Typing	III: 9.5	III: 3.68%	N/A	IV: 4.25	I: 2.25	I: 0.0	IV: 4.22
	$F(1.9, 28.3) = 9.8$		$\chi^2(3) =$ 44.4	20.07	19.57	33.85	27.35
	$p = .001$		$p < .001$	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	$\eta_p^2 = .395$						

Table 1. An overview of the logged performance (WPM, TER, Hand Movement) and subjective questionnaire data (Mental Demand, Physical Demand, Arm Fatigue, Eye Fatigue). The techniques are ranked for each measure based on mean with best and second best being highlighted. Significance values are given when applicable.

- NASA TLX: to measure cognitive workload, answered immediately after each technique [6].
- Borg CR10 [5] (Arm): to measure arm fatigue (including hand, shoulder, etc.), filled out after each technique.
- Borg CR10 (Eyes): to measure eye fatigue, answered after each technique.
- Ranking: ranking of the techniques from most to least preferred happened after all conditions, followed with a text field and interview on the reasons for the preferences.

4.8 Data Analysis

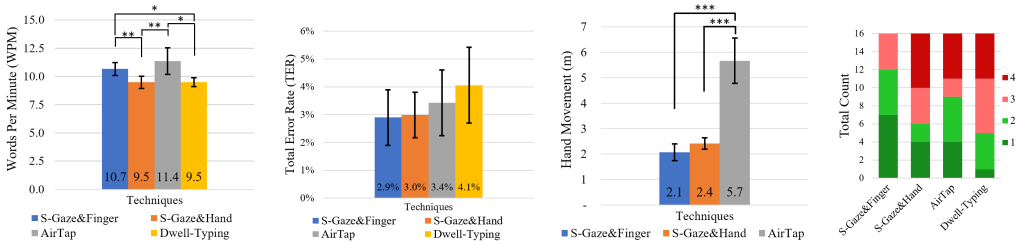
For statistical analysis of WPM and TER, we ran a repeated measures ANOVA (Greenhouse Geiser corrections if sphericity was violated) with post-hoc pairwise comparisons (Bonferroni corrected). For survey data and non-normal data, a Friedman test was employed with post-hoc Wilcoxon signed-rank tests (Bonferroni corrected).

5 RESULTS

In the following, we provide details on the results from the data analysis regarding task performance and subjective experiences. An overview of the results are given in Table 1.

5.1 Task Performance

We measured participants' task performance by logging their speed (WPM) and accuracy (error rate). Figure 5a shows the mean WPM per technique. WPM ranged from 9.49 ($SD = .97$) for *S-Gaze&Hand* to 11.37 ($SD = 2.11$) for *AirTap*. Univariate ANOVA found statistical significance in WPM ($F(1.889, 28.340) = 9.802, p = .001, \eta_p^2 = .395$). Post-hoc analysis showed significant difference in WPM between *S-Gaze&Hand* and *S-Gaze&Finger* ($p = .002$), *S-Gaze&Hand* and *AirTap* ($p = .006$), *Dwell-Typing* and *S-Gaze&Finger* ($p = .021$), and *Dwell-Typing* and *AirTap* ($p = .034$). No other significance was found. Figure 5b shows the mean TER per technique. Overall, TER was low across the techniques, ranging from 2.9% ($SD = 1.79\%$) for *S-Gaze&Finger* to 4.06% ($SD = 2.45\%$) for *Dwell-Typing*. No statistical significance was found between the techniques. Figure 5c shows the mean Hand Movement in meters per technique. Recorded Hand Movement ranged from 0 ($SD = 0$) for *Dwell-Typing* to 5.66 ($SD = 1.6$) for *AirTap*. The Hand Movement data for *S-Gaze&Hand* was found to not be normally distributed. The Friedman test indicated statistical significance in recorded Hand Movement ($\chi^2(3) = 44.4, p < .001$). Post-hoc analysis showed significant difference in recorded Hand Movement between *S-Gaze&Finger* and *AirTap* ($p < .001$), and *S-Gaze&Hand* and *AirTap* ($p < .001$). No other significance was found.



(a) Mean WPM achieved. (b) Mean TER achieved. (c) Mean Hand Movement performed per phrase. (d) Ranking from 4 (least) to 1 (most preferred).

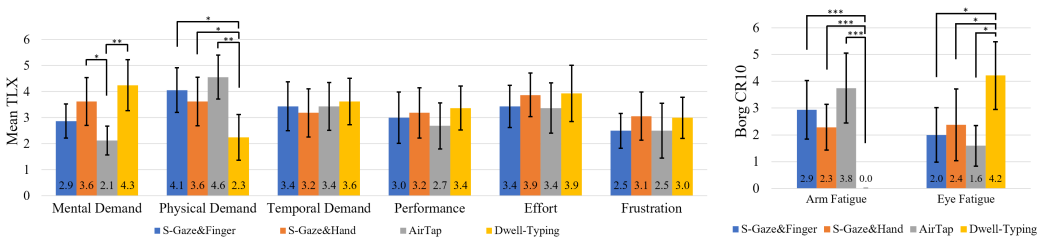
Fig. 5. Results on WPM (a), TER (b), and Hand Movement (c), and Rankings (d). Note that no Hand Movement was recorded for *Dwell-Typing*. Statistical significance shown as * for $p < .05$, ** for $p < .01$, *** for $p < .001$.

5.2 Subjective Questionnaires

Figure 6a shows the mean TLX sub-scales per technique. Reported Mental Demand ranged from 2.13 ($SD = .96$) for *AirTap* to 4.25 ($SD = 1.69$) for *Dwell-Typing*. The Friedman test indicated statistical significance in reported Mental Demand ($\chi^2(3) = 20.067, p < .001$). Post-hoc analysis showed significant difference in reported Mental Demand between *AirTap* and *S-Gaze&Hand* ($p = .048$), and from *AirTap* to *Dwell-Typing* ($p = .006$). Additionally, reported Physical Demand ranged from 2.25 ($SD = 1.53$) for *Dwell-Typing* to 4.56 ($SD = 1.46$) for *AirTap*. The Friedman test indicated statistical significance in reported Physical Demand ($\chi^2(3) = 19.57, p < .001$). Post-hoc analysis showed significant difference in reported Physical Demand between *Dwell-Typing* and *S-Gaze&Finger* ($p = .042$), *Dwell-Typing* and *S-Gaze&Hand* ($p = .048$), and *Dwell-Typing* and *AirTap* ($p = .006$). No other significance was found for the reported TLX.

Figure 6b shows the mean Borg CR10 (Arm & Eyes) per technique. Reported Arm Fatigue ranged from 0.0 ($SD = 0.0$) for *Dwell-Typing* to 3.75 ($SD = 2.27$) for *AirTap*. The Friedman test indicated statistical significance in reported Arm Fatigue ($\chi^2(3) = 33.854, p < .001$). Post-hoc analysis showed a significant difference in reported Arm Fatigue between *Dwell-Typing* and all other techniques ($p < .001$). Additionally, reported Eye Fatigue ranged from 1.59 ($SD = 1.32$) for *AirTap* to 4.22 ($SD = 2.2$) for *Dwell-Typing*. The Friedman test indicated statistical significance in reported Eye Fatigue ($\chi^2(3) = 27.352, p < .001$). Post-hoc analysis showed a significant difference in reported Eye Fatigue between *Dwell-Typing* and all other techniques ($p < .012$). No other significance was found for reported Arm or Eye Fatigue.

As part of the ranking questionnaire, participants were asked to explain their reasoning for their ranking, providing insights into the pros and cons perceived by our participants. Several participants reported that *S-Gaze&Finger* gave them more control than the other gaze-based techniques and felt more natural with direct pointing as opposed to indirect cursor control and eye fixation. This is consistent with the results of [22]. On the other hand, some participants described *S-Gaze&Hand* to allow for a more comfortable hand position and more in control and less eye fatigue than *Dwell-Typing*. However, some participants also noted having issues with their hand going out of the trackable range leading to unforeseen cursor movement. *Dwell-Typing* was described as being more fatiguing and stressful than the other gaze-based techniques, with eyes drying out quicker. Finally, many participants described *AirTap* as physically demanding, some also noting it requiring high depth precision to select keys, although some noted preferring arm fatigue over eye fatigue.



(a) NASA TLX results.

(b) The results from the Borg-CR10 questionnaires.

Fig. 6. The NASA TLX (a) and BORG-CR10 (b) questionnaire results of the techniques as reported by the participants of our study. Statistical significance shown as * for $p < .05$, ** for $p < .01$, *** for $p < .001$.

6 DISCUSSION

In this work, we have optimised and studied text entry performance of two techniques based on the *Gaze-Hand Alignment* principle [22]. The following discusses the main findings in context of related work along with other insights on physical demand, dwell-time, and context switching.

The results from our study show that *S-Gaze&Finger* is comparable to the baseline of *AirTap* in terms of WPM (10.66 *vs.* 11.37) and TER (2.9% *vs.* 3.54%), while significantly reducing the need for hand movement (2.07*m* *vs.* 5.66*m*). This is an interesting result as it indicates that *S-Gaze&Finger* could reduce the effort of text entry compared to state-of-the-art techniques on the HoloLens 2 without sacrificing performance.

Regarding the perceived Mental Demand, no difference was found for *S-Gaze&Finger* compared to *AirTap*. However, the *AirTap* baseline scored better than *S-Gaze&Hand* and *Dwell-Typing*, possibly due to the indirect controls of *S-Gaze&Hand* [22]), and the cognitive demand that comes with the constant staring with *Dwell-Typing*. The techniques that exhibit hand movement scored higher on Physical Demand, similar to findings of related studies [39]. Surprisingly, however, the difference in physical movement between the three freehand techniques did not correlate with the users' perceived Physical Demand or Arm Fatigue. This disparity may be the result of the user study being short, with only 2 + 8 phrases to enter, whereas it may take longer for the arm fatigue to really manifest. It may also be that participants did not fully exploit the flexible hand positioning afforded by *S-Gaze&Finger* and *S-Gaze&Hand* yet, e.g., with *S-Gaze&Finger* users could in principle use their hand closer to themselves to further reduce physical movement. Some participants did, however, report in the qualitative feedback that *S-Gaze&Hand* allowed for more comfortable hand positions. Our work studied the standard keyboard design used with a modern AR HMD, but future studies could explore how distance of hand and keyboard location may add novel insights into this.

Regarding Eye Fatigue, no significant difference was found between the three techniques utilising hand gestures, while the gaze only *Dwell-Typing* technique exhibited significantly higher rated Eye Fatigue. This suggests that the users did not feel the addition of the gaze modality for *S-Gaze&Finger* and *S-Gaze&Hand*. Some users mention having less opportunity to pause with *Dwell-Typing*, while for the *S-Gaze&Finger* and *S-Gaze&Hand* the users would be able to look around.

The most essential difference between the techniques proposed in [22] and ours, is the addition of dwell-time for the gaze and manual pointer. This enabled a more reliable detection of the user's selection intent and provide a well-balanced speed-error trade-off for text entry. Importantly, the dwell-time enables users to perform natural eye movements that coincidentally align gaze and hand pointer, without accidentally triggering a selection. Future work can consider dynamic dwell-time

adjustment [28] for improving trigger timing as users get more familiar with the workings of the techniques. Dwell-time was not the only issue of the prior *Gaze&Hand* technique that caused accidental selections. It also sometimes occurred when the participants searched for the hand cursor when they lost track of it [22]. We resolved this through clearer visual feedback, by establishing a line from the user's palm to the cursor.

During the development of the S-GHA algorithm, it was noted that the user needs frequent context switching between reading the presented phrase, writing text, and reading the transcribed text. For *AirTap*, expert users are able to write text on the virtual keyboard while reading what they are writing, similar to a physical keyboard, although much slower, as they learn the spatial positioning of the keys. In contrast, this is not possible with the gaze-based techniques as they require the eyes to be fixated on keys. As there are yet many cases where users fixate keys before clicking, particularly for mid-air text entry novices, ours allows to potentially help out with effort over time.

Our work focuses solely on selection-based text entry. There are various techniques using word prediction to speed up gaze-based typing techniques (e.g., [17]). However, it would result in additional context switching, i.e. looking at a bar with the predicted words, that can affect the techniques in our study. Moreover, word-gesture text entry and path prediction [17] could possibly be implemented to extend the *Gaze-Hand Alignment* principle, e.g, to use our selection algorithm to delimit the start and end of a word input.

A limitation of the *S-Gaze&Hand* techniques is that the mapping of hand movement to cursor movement was not axis-aligned. With the keyboard being tilted to face the user, it would possibly have been more natural to move the hand on the same axis of tilt. Instead, users had to move their hand up in world space to make the cursor move up in the keyboard's coordinate space. However, this kind of movement may include more muscle activity, which could increase physical demand. Furthermore, the *S-Gaze&Hand* technique was limited by the tracking range of the apparatus used, with a few users noticing that their hand would go outside the tracking range and the cursor would snap back to the original position in the top of the keyboard area. Data collection was limited by the apparatus used. The use of a static dwell-time for the *Dwell-Typing* technique could potentially be a cause for the high rating of Eye Fatigue. The use of a lower static dwell-time or potentially a dynamic dwell-time could mitigate eye fatigue.

7 CONCLUSION

This paper investigated a multimodal gaze and manual text entry interface. We described a novel selection algorithm that intelligently uses dwell-timers for both modalities to strive for an efficient speed-error trade-off. Through two instances of techniques, we assessed the text entry performance to fully manual and fully eye-based baselines. We found that one of the proposed techniques, *S-Gaze&Finger*, provides a good balance between high text entry performance and reduction of physical movement. Our work provides empirical insights and quantifies the pros and cons of multimodal gaze interaction compared to uni-modal approaches and demonstrates the advantages of combining gaze and manual pointing, and enabling selection by their alignment. With increasing availability of gaze and hand tracking sensors in AR headsets, we believe that gaze can provide efficient assistance to render mid-air gesture text entry easier to perform.

ACKNOWLEDGMENTS

This work has received funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grants no. 101021229 GEMINI, and no. 740548 CIO) and by the Innovation Fund Denmark, as part of the Manufacturing Academy of Denmark (MADE) FAST project.

REFERENCES

[1] Richard A Abrams, David E Meyer, and Sylvan Kornblum. 1990. Eye-hand coordination: oculomotor control in rapid aimed limb movements. *Journal of experimental psychology: human perception and performance* 16, 2 (1990), 248.

[2] Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136.

[3] Ana M Bernardos, David Gómez, and José R Casar. 2016. A comparison of head pose and deictic pointing interaction methods for smart environments. *International Journal of Human-Computer Interaction* 32, 4 (2016), 325–351.

[4] Eric A. Bier, Maureen C. Stone, Ken Pier, Ken Fishkin, Thomas Baudel, Matt Conway, William Buxton, and Tony DeRose. 1994. Toolglass and Magic Lenses: The See-through Interface. In *Conference Companion on Human Factors in Computing Systems* (Boston, Massachusetts, USA) *(CHI '94)*. Association for Computing Machinery, New York, NY, USA, 445–446. https://doi.org/10.1145/259963.260447

[5] Gunnar Borg. 1998. *Borg's Perceived Exertion And Pain Scales*.

[6] Lacey Colligan, Henry W.W. Potts, Chelsea T. Finn, and Robert A. Sinkin. 2015. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *International Journal of Medical Informatics* 84, 7 (2015), 469–476. https://doi.org/10.1016/j.ijmedinf.2015.03.003

[7] Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the computer using gaze gestures. In *IFIP Conference on Human-Computer Interaction*. Springer, 475–488.

[8] J. Dudley, H. Benko, D. Wigdor, and P. Kristensson. 2019. Performance Envelopes of Virtual Keyboard Text Input Strategies in Virtual Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE Computer Society, Los Alamitos, CA, USA, 289–300. https://doi.org/10.1109/ISMAR.2019.00027

[9] Wenxin Feng, Jiangnan Zou, Andrew Kurauchi, Carlos H Morimoto, and Margrit Betke. 2021. HGaze Typing: Head-Gesture Assisted Gaze Typing. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. https://doi.org/10.1145/3448017.3457379

[10] Gabriel González, José P Molina, Arturo S García, Diego Martínez, and Pascual González. 2009. Evaluation of text input techniques in immersive virtual environments. In *New Trends on Human–Computer Interaction*. Springer, 109–118.

[11] Poika Isokoski. 2000. Text Input Methods for Eye Trackers Using Off-Screen Targets. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00)*. Association for Computing Machinery, New York, NY, USA, 15–21. https://doi.org/10.1145/355017.355020

[12] Robert J. K. Jacob. 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* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/97243.97246

[13] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 152–169. https://doi.org/10.1145/123078.128728

[14] Roland S Johansson, Göran Westling, Anders Bäckström, and J Randall Flanagan. 2001. Eye–hand coordination in object manipulation. *Journal of neuroscience* 21, 17 (2001), 6917–6932.

[15] Per Ola Kristensson and Keith Vertanen. 2012. The Potential of Dwell-Free Eye-Typing for Fast Assistive Gaze Communication. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 241–244. https://doi.org/10.1145/2168556.2168605

[16] Andrew Kurauchi, Wenxin Feng, Ajjen Joshi, Carlos Morimoto, and Margrit Betke. 2016. EyeSwipe: Dwell-Free Text Entry Using Gaze Paths. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1952–1956. https://doi.org/10.1145/2858036.2858335

[17] Andrew Kurauchi, Wenxin Feng, Ajjen Joshi, Carlos Morimoto, and Margrit Betke. 2016. EyeSwipe: Dwell-Free Text Entry Using Gaze Paths. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1952–1956. https://doi.org/10.1145/2858036.2858335

[18] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173655

[19] Michael Land, Neil Mennie, and Jennifer Rusted. 1999. The roles of vision and eye movements in the control of activities of daily living. *Perception* 28, 11 (1999), 1311–1328.

[20] Gun A. Lee, Mark Billinghurst, and Gerard Jounghyun Kim. 2004. Occlusion Based Interaction Methods for Tangible Augmented Reality Environments. In *Proceedings of the 2004 ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry* (Singapore) *(VRCAI '04)*. Association for Computing Machinery,

New York, NY, USA, 419–426. https://doi.org/10.1145/1044588.1044680

[21] Xueshi Lu, Difeng Yu, Hai-Ning Liang, Wenge Xu, Yuzheng Chen, Xiang Li, and Khalad Hasan. 2020. Exploration of Hands-free Text Entry Techniques For Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 344–349. https://doi.org/10.1109/ISMAR50242.2020.00061

[22] Mathias Lystbæk, Peter Rosenberg, Ken Pfueffer, Jens Emil Grønbæk, and Hans Gellersen. 2022. Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Menu-based Input in Augmented Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Seattle, USA) *(ETRA '22 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 145. https://doi.org/10.1145/3530886

[23] I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI EA '03)*. Association for Computing Machinery, New York, NY, USA, 754–755. https://doi.org/10.1145/765891.765971

[24] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast Gaze Typing with an Adjustable Dwell Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 357–360. https://doi.org/10.1145/1518701.1518758

[25] Päivi Majaranta and Kari-Jouko Räihä. 2002. Twenty Years of Eye Typing: Systems and Design Issues. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications* (New Orleans, Louisiana) *(ETRA '02)*. Association for Computing Machinery, New York, NY, USA, 15–22. https://doi.org/10.1145/507072.507076

[26] Anders Markussen, Mikkel Rønne Jakobsen, and Kasper Hornbæk. 2013. Selection-Based Mid-Air Text Entry on Large Displays. In *INTERACT*.

[27] Darius Miniotas, Oleg Špakov, Ivan Tugoy, and I. Scott MacKenzie. 2006. Speech-augmented Eye Gaze Interaction with Small Closely Spaced Targets. In *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications (ETRA '06)*. ACM, New York, NY, USA, 67–72. https://doi.org/10.1145/1117309.1117345

[28] Martez E. Mott, Shane Williams, Jacob O. Wobbrock, and Meredith Ringel Morris. 2017. Improving Dwell-Based Gaze Typing with Dynamic, Cascading Dwell Times. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2558–2570. https://doi.org/10.1145/3025453.3025517

[29] Diogo Pedrosa, Maria Da Graça Pimentel, Amy Wright, and Khai N. Truong. 2015. Filteryedping: Design Challenges and User Performance of Dwell-Free Eye Typing. *ACM Trans. Access. Comput.* 6, 1, Article 3 (mar 2015), 37 pages. https://doi.org/10.1145/2724728

[30] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, and Hans Gellersen. 2014. Gaze-Touch: Combining Gaze with Multi-Touch for Interaction on the Same Surface. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association for Computing Machinery, New York, NY, USA, 509–518. https://doi.org/10.1145/2642918.2647397

[31] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. 2015. Gaze-Shifting: Direct-Indirect Input with Pen and Touch Modulated by Gaze. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) *(UIST '15)*. Association for Computing Machinery, New York, NY, USA, 373–383. https://doi.org/10.1145/2807442.2807460

[32] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 99–108. https://doi.org/10.1145/3131277.3132180

[33] Ken Pfeuffer, Lukas Mecke, Sarah Delgado Rodriguez, Mariam Hassib, Hannah Maier, and Florian Alt. 2020. Empirical Evaluation of Gaze-Enhanced Menus in Virtual Reality. In *26th ACM Symposium on Virtual Reality Software and Technology* (Virtual Event, Canada) *(VRST '20)*. Association for Computing Machinery, New York, NY, USA, Article 20, 11 pages. https://doi.org/10.1145/3385956.3418962

[34] Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. 1997. Image Plane Interaction Techniques in 3D Immersive Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics* (Providence, Rhode Island, USA) *(I3D '97)*. Association for Computing Machinery, New York, NY, USA, 39–ff. https://doi.org/10.1145/253284.253303

[35] Daniel Rough, Keith Vertanen, and Per Ola Kristensson. 2014. An Evaluation of Dasher with a High-Performance Language Model as a Gaze Communication Method. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (Como, Italy) *(AVI '14)*. Association for Computing Machinery, New York, NY, USA, 169–176. https://doi.org/10.1145/2598153.2598157

[36] Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of Eye Gaze Interaction. In *Proc. SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) *(CHI '00)*. ACM, New York, USA, 281–288. https://doi.org/10.1145/332040.332445

[37] Ludwig Sidenmark and Anders Lundström. 2019. Gaze Behaviour on Interacted Objects during Hand Interaction in Virtual Reality for Eye Tracking Calibration. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research &*

Applications (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 6, 9 pages. https://doi.org/10.1145/3314111.3319815

[38] R. William Soukoreff and I. Scott MacKenzie. 2003. Metrics for Text Entry Research: An Evaluation of MSD and KSPC, and a New Unified Error Metric. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 113–120. https://doi.org/10.1145/642611.642632

[39] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. *Selection-Based Text Entry in Virtual Reality*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174221

[40] Sophie Stellmach and Raimund Dachselt. 2012. Look & Touch: Gaze-Supported Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2981–2990. https://doi.org/10.1145/2207676.2208709

[41] Antonio Diaz Tula, Filipe M. S. de Campos, and Carlos H. Morimoto. 2012. Dynamic Context Switching for Gaze Based Interaction. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 353–356. https://doi.org/10.1145/2168556.2168635

[42] Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. 1997. Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem". In *Human-Computer Interaction INTERACT '97*. Springer, Boston, MA, 509–516. https://doi.org/10.1007/978-0-387-35175-9_77

[43] Jean-Louis Vercher, G Magenes, C Prablanc, and GM Gauthier. 1994. Eye-head-hand coordination in pointing at visual targets: spatial and temporal analysis. *Experimental brain research* 99, 3 (1994), 507–523.

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

[45] Pierre Weill-Tessier and Hans Gellersen. 2017. Touch input and gaze correlation on tablets. In *International Conference on Intelligent Decision Technologies*. Springer, Springer International Publishing, Cham, 287–296.

[46] Jacob O. Wobbrock, James Rubinstein, Michael W. Sawyer, and Andrew T. Duchowski. 2008. Longitudinal Evaluation of Discrete Consecutive Gaze Gestures for Text Entry. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications* (Savannah, Georgia) *(ETRA '08)*. Association for Computing Machinery, New York, NY, USA, 11–18. https://doi.org/10.1145/1344471.1344475

[47] Wenge Xu, Hai-Ning Liang, Anqi He, and Zifan Wang. 2019. Pointing and Selection Methods for Text Entry in Augmented Reality Head Mounted Displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 279–288. https://doi.org/10.1109/ISMAR.2019.00026

[48] Chun Yu, Yizheng Gu, Zhican Yang, Xin Yi, Hengliang Luo, and Yuanchun Shi. 2017. 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* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4479–4488. https://doi.org/10.1145/3025453.3025964

[49] Shumin Zhai and Per Ola Kristensson. 2012. The Word-Gesture Keyboard: Reimagining Keyboard Interaction. *Commun. ACM* 55, 9 (sep 2012), 91–101. https://doi.org/10.1145/2330667.2330689

[50] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053

Received November 2021; revised January 2022; accepted April 2022