Gaze-Assisted Typing for Smart Glasses

Sunggeun Ahn
HCI Lab, School of Computing, KAIST
Daejeon, Republic of Korea
topmaze@kaist.ac.kr

Geehyuk Lee
HCI Lab, School of Computing, KAIST
Daejeon, Republic of Korea
geehyuk@gmail.com

ABSTRACT
Text entry is expected to be a common task for smart glass users, which is generally performed using a touchpad on the temple or by a promising approach using eye tracking. However, each approach has its own limitations. For more efficient text entry, we present the concept of gaze-assisted typing (GAT), which uses both a touchpad and eye tracking. We initially examined GAT with a minimal eye input load, and demonstrated that the GAT technology was 51% faster than a two-step touch input typing method (*i.e.*,M-SwipeBoard: 5.85 words per minute (wpm) and GAT: 8.87 wpm). We also compared GAT methods with varying numbers of touch gestures. The results showed that a GAT requiring five different touch gestures was the most preferred, although all GAT techniques were equally efficient. Finally, we compared GAT with touch-only typing (SwipeZone) and eye-only typing (adjustable dwell) using an eye-trackable head-worn display. The results demonstrate that the most preferred technique, GAT, was 25.4% faster than the eye-only typing and 29.4% faster than the touch-only typing (GAT: 11.04 wpm, eye-only typing: 8.81 wpm, and touch-only typing: 8.53 wpm).

Author Keywords
Text entry; Smart glasses; Gaze and Touch, Multi-modal input

CCS Concepts
•**Human-centered computing** → **Text input;** *Empirical studies in interaction design;*

INTRODUCTION
Text entry on smart glasses is an important subject considered by several recent studies [1, 11, 14, 39, 41, 46]. It is a challenging problem owing to the limited input space of smart glasses, which is generally a small and oblong touchpad on the temple. The literature presents several text entry methods using the on-frame touchpad. One of them is 1D Handwriting (letter input of 4.67 words per minute (wpm)), which uses unistroke gestures [46]. Methods such as SwipeBoard (7.14 wpm) and SwipeZone (8.73 wpm) [11] use a two-step scheme to enter a character. However, text entry on the on-frame touchpad is difficult and inefficient because of the narrow input space. Additionally, the on-frame touchpad is perpendicular to the display of the glasses; therefore, the mapping between the touchpad and the display is often confusing.

UIST '19, October 20-23, 2019, New Orleans, LA, USA.

ACM ISBN 978-1-4503-6816-2/19/10 ...$15.00.
http://dx.doi.org/10.1145/3332165.3347883

Considering the recent developments in mobile eye tracking technology (Pupil-labs [1] and Tobii Pro Glasses 2 [2]), eye input is a promising option to overcome the limitations of the touchpad-based typing methods.

Gaze typing can overcome the ambiguous mapping problem of the touchpad-based typing. Furthermore, it can provide a wider input space considering the recent trend of increasing the field-of-view (FoV). When compared to early smart glasses, such as Google Glass (15° of FoV), recent smart glasses or head-worn displays (HWDs) have a wider display, such as Moverio BT-350 (23°) and Hololens2, with a twofold wider FoV when compared to its previous version (Hololens: 30°). In fact, typing with eye gaze was shown to be an effective option in virtual reality environment (9.36 wpm [33]).

However, gaze typing also has limitations. Its performance depends significantly on eye-tracking accuracy, which varies for different users [33]. Further, it may be accurate initially, but becomes inaccurate over time owing to the slippage of the HWD. State-of-the-art eye tracking technologies use a 3D eyeball model and track both eyes. It is more accurate and also robust against the slippage problem when compared to the early techniques (2D pupil position tracking method) [23]. However, according to a previous study [33], the inaccuracy of a mobile eye tracker is still problematic for text entry. In addition, the Midas touch [15] problem in smart glasses should be considered more seriously because a user always has to see the display while s/he sees the "real world" over the display.

To overcome the problems of the touchpad- and eye-gaze-based methods, we propose the gaze-assisted typing (GAT). GAT is a complementary combination of eye-gaze- and touchpad-based approaches. With this combination, we can use larger targets for the eye gaze and a set of simpler touchpad operations. In addition, the Midas touch problem of an eye-gaze-based method can be avoided.

GAT is two-step input approach for entering a character. The keyboard for GAT consists of several sections, and each section contains several keys. Initially, a user can select one of the sections with the gaze input. Then, the user can enter a key, in the section, with the touch input. It is the first adaptation of

[1] https://pupil-labs.com/
[2] https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

a harmonized usage of the gaze and touch inputs to discretely compensate for the limited expression of the input modalities.

We first conducted a preliminary study to validate the effect of introducing eye input modality to a two-step touch-only text entry with a minimal load on eye input. After we validated that GAT can improve touchpad typing, we attempted to optimize the GAT. The most basic design option of GAT is the number of sections (sub-keyboards). To determine the best choice, we conducted Experiment 1, where we varied the precision requirement of eye tracking and the touch gesture set size. Finally, to demonstrate that the GAT method with larger keys for eye selection and simpler touchpad gestures would lead to improved performance and usability benefits over eye-only and touch-only typing methods, we conducted Experiment 2.

RELATED WORK

Possible Text Entry Methods for Smart Glasses

To enter text, voice input is a natural and fast technique [3]; however, its use is limited because 1) environmental noise can hamper recognition accuracy, 2) a user sometimes wants to enter text without other people noticing, and 3) voice input can raise privacy issues, which may be dependent on the applications. In the subsequent section, we review the literature concerning possible text entry methods for smart glasses that can be used in a general mobile environment.

On glasses frame input

Previous researchers designed text entry techniques to overcome the limited input space of the touchpad on the frame of the glasses. The first approach utilized a two-step input to enter a character because of the small input space. Grossman *et al.* adapted SwipeBoard [5], which was designed for a smartwatch, to smart glasses [11]; SwipeBoard and SwipeZone, achieved 7.1 wpm and 8.7 wpm, respectively. Another approach utilized complex gestures to enter a character. Yu *et al.* designed a unistroke gesture technique [46] and showed that 1D Handwriting (4.7 wpm) outperformed the selection-based text entry technique, *i.e.,* 1Line-Keyboard (4.2 wpm).

On hand-held device input

Remote text entry techniques can be used for smart glasses. In a TV viewing environment, a user can enter text at 8.5 wpm without a learning effort by using a touchpad with a hover tracking feature [6]. Specialized hand-held keyboards are also viable options for fast text entry, such as the mini-qwerty [7] or the chord keyboard [20]. However, a hand-held device may constitute an additional input device rather than being a primary input device for the glasses users due to carriage load.

On Body input

The on-body input is usually achieved by using additional wearable devices, such as a finger-attached touch sensor for the on-finger input (DigiTouch [41]; 16.0 wpm), a wrist-worn sensor for on-palm input (PalmType [39]; 4.6 wpm), or a smartwatch for on-wrist input (SwipeBoard and HoldBoard [1]; 9.1 wpm and 10.2 wpm, respectively). The on-body touch input is a possible option for a larger and more comfortable touch input space than on-glasses input, without carrying a hand-held device. It may be more physically accessible [24] and socially acceptable [1, 34] than the on-glass touch input. Considering the on-skin touch, kinesthetic sense and haptic feedback also help a user to accurately input [18].

Mid-air input

Mid-air text entry was widely explored by early researchers using techniques such as mimicking typing on a standard keyboard [8, 14, 29, 45], key-stroke gesture input [35], and word-stroke gesture input [12, 25]. Mid-air gesture techniques have to consider arm fatigue [13] and social acceptance [14]. Nevertheless, mimicking of standard keyboard typing (ATK [45]: 29.2 wpm, an hour of usage, and VISAR [8]: 17.8 wpm, 8 days of usage) and word-stroke gestures (Vulture [25]: 20.6 wpm, 10 days of usage, and RotoSwipe [12]: 14.8 wpm, 5 days of usage) demonstrated high input speeds; however, according to Yeo *et al.* [44], in-air typing performance may deteriorate, subject to the quality of the hand tracker.

Gaze-based Interaction for Head-worn Display (HWD)

In this study, we attempted to achieve a more usable on-glass text entry, primarily in terms of speed. To achieve this goal, we focused on the gaze input modality. HWDs generally exploit the gaze input modality, not only for text entry, as these devices can sense the gaze input. Toyama *et al.* presented a gaze assisted translation system for smart glasses [37] using a mobile eye tracker from SMI. Toyama *et al.* also presented gaze assisted object recognition methods for museum guide systems [36]. Kito *et al.* investigated various multimodal two-step pointing techniques [17] by using an eye tracker from Pupil Labs. Esteves *et al.* presented a menu selection technique called Orbits, that utilizes the smooth pursuit of eye movement for a smartwatch utilizing mobile eye tracker [9].

Gaze-based Text entry methods

For gaze-based text entry, several studies focused on solving the Midas touch problem [15]. The dwell-click methods were extensively used to avoid this problem. Early studies on the dwell-click method achieved low efficiency (5 to 10 wpm [22]) owing to the relatively long dwell time (typically 400 to 1000 ms). Certain researchers focused on optimizing the dwell time (adjustable dwell time: 19.9 wpm after 10 days of usage [22] and dynamic cascading: 13.7 wpm after 8 sessions of usage [28]). These techniques improved the dwell-based eye typing; however, a user can practice to create an optimal dwell time.

For dwell-free eye typing, gesture-based eye typing approaches were thoroughly explored: Examples include a keystroke gesture approach (EyeWrite [43]: 4.9 wpm), continuous gesture approach (Dasher: 17.26 wpm [40], and word-stroke gesture approach (EyeSwipe [16]: 11.7 wpm).

Another approach to avoid dwell time is to use the click-alternative rather than the dwell-click. The click-alternative can be classified into two categories: monomodal (context switching [27] and eye blinking [2]) and multimodal click-alternatives (foot input [32], facial action [38], button click [33], *etc.*) Pfeuffer *et al.* [31] also presented the concept of complementary combination of eye and touch text entry by adapting the manual and gaze inputted cascading (MAGIC) pointing technique [47].

In fact, Gaze + Gesture interaction have been used not only for text entry but also for many other interactions [30, 31, 47], *e.g.,* pointing, object manipulation, menu selection, *etc*. As far as text entry methods are considered, however, gaze-assisted typing (GAT) is distinguished from the prior works as we consider a taxonomy for Gaze + Gesture combinations that was presented by Chatterjee *et al.* [4]. According to the taxonomy, the gaze input is used in the target acquisition phase and the gesture input is used in the target action phase. The target acquisition (approaching a target) may be above accuracy (AA) or below accuracy (BA), and the target action (selecting a target) may be a discrete action (DA) or a continuous manipulation (CM). In this 2 x 2 taxonomy, the existing Gaze + Gesture text entry methods belong either to the AA + DA category (the click-alternative methods) or the BA + CM category (MAGIC-based method). Conversely, GAT belongs to the BA + DA category because it chose BA to overcome the inaccuracy of a mobile eye tracker, and DA to overcome the limited gesture space of the oblong touchpad.

GAZE-ASSISTED TYPING, "GAT"

Our proposed approach is a two-step method using a combination of gaze and touch gesture for text entry. To enter a key using GAT, initially, a user selects one of the sections of the keyboard, called a sub-keyboard, using a gaze input. Subsequently, s/he enters a letter on the sub-keyboard using the touch gesture input as shown in Figure 1(b) and (c).

Sub-keyboard Selection

To select a sub-keyboard, a user merely looks at it as shown in Figure 1(b). For selection, we provided a cross-hair-shaped eye cursor on the keyboard. The eye cursor follows the gaze position of the user on the x-axis. When the eye cursor selected a sub-keyboard along the x-axis, the sub-keyboard is activated as shown in Figure 2. The eye cursor can provide a sense of control for the eye input; however, it may distract a user or occlude a key. To minimize distraction and to avoid occlusion, the y position of the eye cursor is fixed below the keys.

Key Selection

To select a key, a user can perform one of the nine touch gestures, similar to SwipeBoard [11], to enter a key as shown in Figure 1(c), *i.e.*, a tap gesture to select a center-positioned key on the sub-keyboard, and one of the eight directional swipe gestures to select a key corresponding to the swipe direction from the center-positioned key as shown in Figure 1(c). In our design, the left and right directions on the screen were mapped in the backward and forward directions on the touchpad, respectively. Thus, to enter "c" in Figure 1(c), a user performed the lower-forward direction swipe gesture. The space and backspace keys were mapped to two-finger forward and backward swipe gestures. The enter key was mapped to three-finger long touch gesture (500 ms).

While selecting a key with a touch input, owing to eye jitter and target key searching behavior, the eyes of the user are not stable during a keystroke. If the eye cursor always moves to a gaze position, an activated sub-keyboard can be changed accidentally before entering a key. Moreover, in our pilot study, we also observed that users were inclined to not blink until a

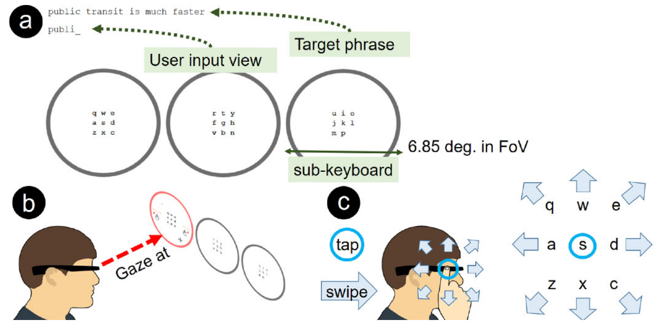


Figure 1. (a) Early GUI for gaze-assisted typing (GAT) which was used in preliminary study. (b) Eye input for sub-keyboard selection (c) Touch input for key selection on the activated sub-keyboard.

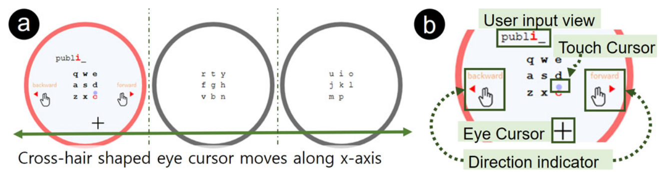


Figure 2. (a) Eye cursor movement, dotted lines indicate effective boundary of sub-keyboards (b) Magnified view, activated sub-keyboard with labeling of GUI components; The user input view on an activated sub-keyboard only show a last 6 letters of user inputted.

typing task was completed when the eye cursor continued to respond to eye movement; consequently, a user experienced more fatigue. An intentional eye movement of the user may occur before starting the touch input; therefore, we fixed the eye cursor once a user started the touch input. After inputting by touch, the eye cursor follows the gaze position again.

Further, it is easy to confuse the forward and backward directions because the display and the touchpad are placed perpendicularly. To reduce the typing error by confusing the swipe direction, we provided a visual feedback (touch cursor in Figure 2(b)) for the touch position to permit adjustments in the swipe direction before the swipe gesture is finalized. Owing to the touch cursor, a novice user can enter a key in a touch-adjust-release manner; however, an expert user can enter a key with a swipe or tap gesture. We also provided an audio feedback from the glasses for the touch input.

Thus, when a finger touches the touchpad, the touch cursor is first placed at the center of the selected sub-keyboard, regardless of the position of placement. To distinguish the touch gestures, we used a vector from the placement to release positions. If the travel length was under 107 px ($\approx$ 3 mm), the gesture was distinguished as a tap gesture. The swipe direction was distinguished by the direction of a vector.

Expected advantage of GAT

When compared to the touch-only text entry, GAT can achieve better input speed than SwipeBoard [11] owing to the higher eye movement speed. In addition, looking at a key before entering the key is a relatively natural behavior; therefore, a user is not required to spent time on accommodating the gaze input. When compared to the complex gesture approach (1D Handwriting [46]), a user is not required to memorize the number of gestures corresponding to the 26 unistroke gestures.

When compared to the eye-only text entry, the required fidelity of the eye tracker is expected to be lower than that of eye-only methods. To use the eye-only methods, at least 26 independent areas should be distinguished by eye input; conversely, in GAT, the area to be distinguished (the number of sub- keyboards) is divided by the number of touch gestures (N).

PRELIMINARY STUDY: CAN ON-FRAME TOUCH TYPING BE IMPROVED WITH EYE INPUT?

To verify GAT, initially, we demonstrate the net effect of introducing gaze modality with a minimum expression of the eye input. For the minimum expression of the eye input, we used the nine touch gestures for key selection. Owing to the target-agnostic features of swipe gestures, the nine touch gestures consisted of eight directional swipe gestures and one tap gesture, similar to SwipeBoard [11]. Thus, we choose SwipeBoard as a baseline technique in this study, to demonstrate the net effect of replacing an input step with gaze input.

To demonstrate the net effect, we modified the SwipeBoard design to be similar to our GAT design; we called this M-SwipeBoard. The original SwipeBoard [11], consists of nine sub-keyboards; moreover, the key selection utilized three easy-to-use touch gestures to reduce the input error at the key selection. However, M-SwipeBoard consists of three sub-keyboards and the key selection utilized nine touch gestures, similar to GAT. We presented a touch cursor for M-SwipeBoard to guide the finger movements of the user to reduce the input error for swipe gestures, same as GAT (Figure 2(b)).

The only difference between GAT and M-SwipeBoard is the sub-keyboard selection method. When using M-SwipeBoard, a user can select a sub-keyboard with one of the three touch gestures −forward, backward swipe gestures, and tap− to select the right, left, and center positioned sub-keyboards. After the selection, the user can enter a key with the nine touch gestures similar to GAT, as described previously

Implementation

To implement a complete mobile gaze tracking system on smart glasses, we attempted to use an eye tracker from Pupil Labs and Tobii Pro Glasses 2. However, there were slippage problems (Pupil Labs) and latency issues (Tobii, over 70 ms). Different wearable eye trackers have different problems; consequently, for studies conducted with a particular tracker, the results would be dependent on the specific problems of that tracker. To avoid the issues, we simulated the environment instead of using a mobile eye tracker. Thus, we used the GP3 eye tracker [3] with a laptop as shown in figure 3. The GP3 eye tracker captures both eyes of the user at 150 Hz. We used an average filter to reduce jitter. We also used a Google Glass Enterprise Edition to use the built-in touchpad on the glasses.

For accurate and less exhausting eye input, the keyboard size was considered. By considering the study of perceived eye exertion [26], a range of ±15° round of gaze direction was determined to be tolerable for use. Thus, we decided to use 21° as the length of the square bound, fitted in the ±15° round. We also followed the recommendation of Kito *et al.* [17] for the target size for eye pointing on HWD, *i.e.*, 7° of width. The actual viewing angle varies with distance from the monitor; however, we attempted to set the keyboard size as: keyboard width: 21.6° and sub-keyboard width: 6.85°, at a distance of 65 cm. A participant was seated at a distance of approximately 65 cm from the 27 inch monitor (1920 × 1080 px, viewing angle: 54° × 30° at the distance, as shown in Figures 3).

[3] https://www.gazept.com/product/gp3hd/

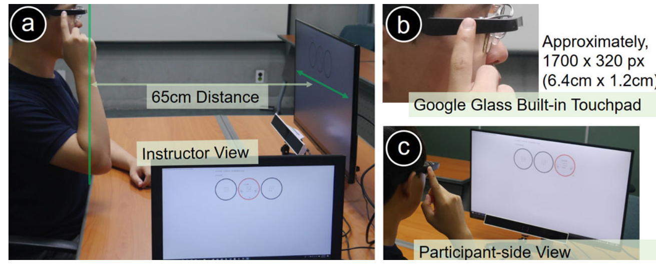


Figure 3. Simulated environment for preliminary study and Experiment 1. (a) instructor view, (b) built-in touchpad, (c) participant-side view.

Task

The task provided was a memorizing and transcription task with the Mackenzie and Soukoreff phrase set [21]. A target phrase was randomly presented at each trial. We instructed the participants to memorize a target phrase for 15 s before starting the task, but, the target phrase was not erased while performing the task. The participants were instructed to enter the text as fast and accurately as possible. We also asked the participants to correct errors only found within 3 or 4 letters.

Design and Procedure

We used a within-subject design with one factor, *i.e.*, technique: M-SwipeBoard and GAT. The experiment consisted of two sessions with a different technique for each session. The order of the sessions was counter balanced within the subjects. At the start of the session, we instructed a user on the process of entering text using a given technique and we calibrated the eye tracker with a five-point calibration. In the GAT session, if a user identified a mismatch in the position of the eye cursor, we recalibrated the eye tracker and discarded the task in progress. After that, a participant entered all 26 letters of the English alphabet in sequence for training; then, participants completed three blocks in a session, and in each block, a participant completed five tasks sequentially. There were at least 15-second breaks between tasks and blocks, and at least 3-minute breaks between sessions. After completing the experiment, we asked the user to compare the eye fatigue for both techniques. The experiment required approximately 60 min per participant. Finally, we gathered 12 participants × 2 techniques × 3 blocks × 5 tasks = 360 phrases.

Participants

We recruited 12 participants (4 females and 8 males, mean age: 22.58, from 19 to 30 years) from our university. All of them are right-handed and non-native English speakers. Three participants had previous experience in using a smart glass and/or an eye tracker; however, none of them were regular users. Five participants were wearing glasses. They received $9 USD as compensation.

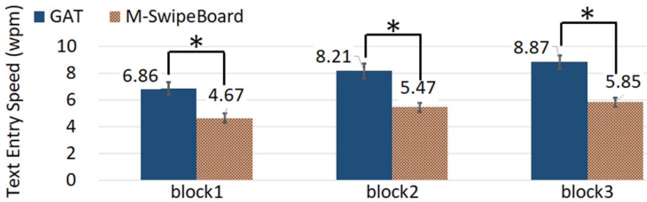


Figure 4. Text entry speed of each technique, per block. Error bars mean standard errors. Asterisks mean significant difference.

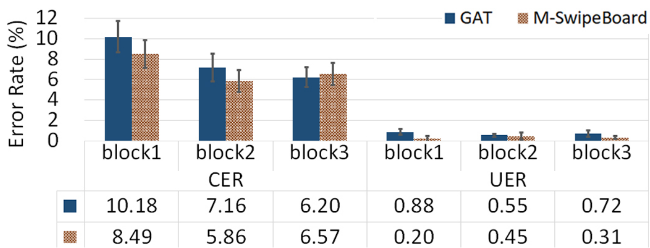


Figure 5. Corrected error rate (CER) and uncorrected error rate (UER) of each technique/block. Error bars indicate standard errors.

Results

We analyzed the text entry speed, corrected error rate (CER), and uncorrected error rate (UER) statistically using a two-way (2 techniques × 3 blocks) repeated measure ANOVA (RM-ANOVA) on each. The UER violated normality assumption; therefore, we performed an aligned rank transform (ART) [42] on the UER before conducting RM-ANOVA. We also conducted post-hoc tests to compare both techniques in each block; paired sample t-tests on WPM and CER with a Bonferroni correction.

Text entry speed

Figure 4 shows the text entry speed of GAT and M-SwipeBoard. RM-ANOVA demonstrated significant effects on the technique ($F_{(1,11)}$= 42.763, $p< 0.01$) and the block ($F_{(2,22)}$= 53.962, $p< 0.01$); however, the interaction was not significant. GAT was significantly faster than M-SwipeBoard in every block, *i.e.*, t_{11}= 4.700, $p< 0.01$; t_{11}= 5.894, $p< 0.01$; and t_{11}= 6.740, $p< 0.01$ for block 1, 2, and 3, respectively. At third block, GAT was 51% faster than M-SwipeBoard.

Error rate

Figure 5 illustrates the CER and UER. For the CER, RM-ANOVA analysis showed a significant effect of the block ($F_{(2,22)}$= 5.854, $p< 0.01$). The effect of the technique and the interaction were not significant. For the UER, RM-ANOVA analysis with ART showed a significant effect of technique ($F_{(1,11)}$= 5.309, $p< 0.05$). The effect of the block and the interaction were not significant. For both CER and UER, the post-hoc tests revealed that all differences between the techniques on each block were not significant.

Eye fatigue

Five of twelve experienced more eye fatigue while using GAT because they had to pay more attention to the eye movement; however, another one said that M-SwipeBoard was more tiring owing to the long usage time. The others said that they had to observe the keyboard in the same manner for both techniques; therefore, they could not feel any difference in the eye fatigue.

EXPERIMENT 1: CAN MORE EXPRESSIVE GAZE INPUT MAKE "GAT" BETTER?

We thought that GAT with a lesser number of touch gestures may require less effort of touch input by eliminating difficult gestures, *i.e.*, diagonal swipe gestures [11] and vertical swipe gestures. Conversely, the number of sub-keyboards must be increased with the reducing number of touch gestures. Thus, explicit eye movement to select a sub-keyboard will occur more frequently. This is expected to cause more eye fatigue. To examine our expectation, we designed and evaluated three variations of GAT with varying number of touch gestures.

GAT Variations: Different Number of Gestures

We designed three variations of GAT: GAT3, GAT6, and GAT9. GAT3 has three sub-keyboards with nine keys for each, similar to the preliminary study. GAT6 was designed to eliminate the four diagonal swipe gestures; thus, each sub-keyboard has five keys for the five touch gestures, and the keyboard has six sub-keyboards (Figure 6(b)). GAT9 was designed to eliminate both the diagonal and vertical swipe gestures; therefore, each sub-keyboard has three keys for the three touch gestures, and the keyboard has nine sub-keyboards as shown in Figure 6(c).

Because we varied the number of touch gestures, the keyboard layouts were correspondingly varied. Thus, we placed the keys in an alphabetical order to minimize the memorability effect of the key placement. In addition, considering the usage environment of smart glasses, it may require minimized visual representation of the keyboard. Thus, we removed the boundary of sub-keyboards, and displayed the visualization to notify the sub-keyboard activation as shown in Figure 6.

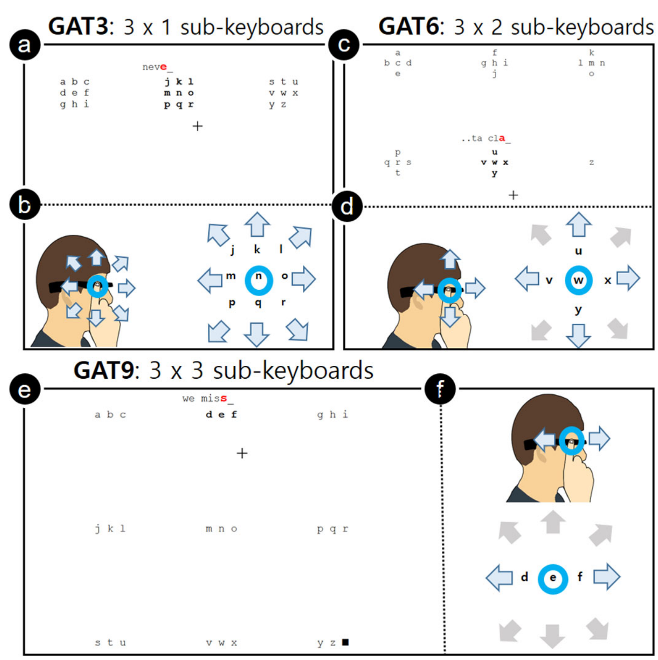


Figure 6. The sub-keyboards layouts of the GAT variations in Experiment 1. Keys are placed in an alphabetical order. (a), (c), and (e) are the layouts of GAT3, GAT6, and GAT9, respectively. (b), (d), and (f) show the sets of touch gestures used for GAT3, GAT6, and GAT9, respectively, with the corresponding keys of the activated sub-keyboards.

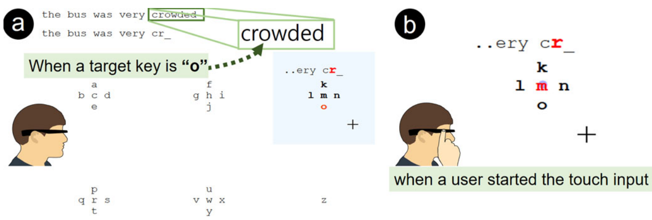


Figure 7. (a) Visual guidance condition example with GAT6, when a target key is "o". (b) Guidance is removed during the touch input mode.

The sub-keyboard size for all variations was 6.85° × 6.85°; however, if the eye cursor out of boundary of the keyboard, a closest sub-keyboard will be activated. For both GAT6 and GAT9, the eye cursor movement along the y-axis moved discretely at a fixed position below the keys.

Design and Procedure

We used the within-subject design, with the layout as a factor, *i.e.*, GAT3, GAT6, and GAT9. The experiment consisted of three sessions. In each session, by using one of the three layouts, a participant completed three blocks in NC; then, completed a block in VGC. Consequently, in each block, a participant completed five tasks. There was a break of at least a minute between blocks and 15-second breaks between tasks. At the start of each session, we calibrated the eye tracker as same to preliminary study. Each session lasted approximately 30 min. The order of the sessions was counter-balanced with a full factorial within subjects. The sessions were split across three days. We controlled the interval of sessions to approximately 24 h as much as possible (from 15 to 41 h; mean = 23.58, σ = 4.12). At the end of the last session, we surveyed a questionnaire with four questions −1) **Easy to learn**; 2) **Easy to use**; 3) **Prefer to use**; and 4) Feel natural eye movements while entering text (**Eye feels natural**)− using a five-point Likert scale. We gathered 18 participants × 3 layouts × (3 blocks of NC + 1 block of VGC) × 5 tasks = 1,080 phrases of text entry results.

Participants

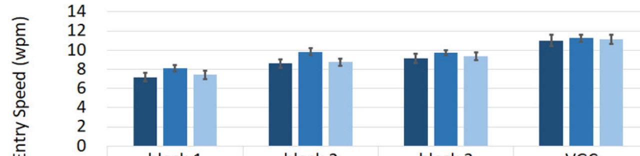
We recruited 18 participants (8 females, 10 males, mean age: 20.72, from 17 to 24 years) from our university. All participants were different from those at the preliminary study, were right-handed, non-native English speakers, with no experience in using a smart glass or an eye tracker. Five participants were wearing glasses. They received $22 USD as compensation.

Apparatus and Task

The apparatus was the same as that of the preliminary study. A user completed tasks in two different conditions: normal condition (NC) and visual guidance condition (VGC). Tasks for the NC and VGC were the same as in the preliminary study.

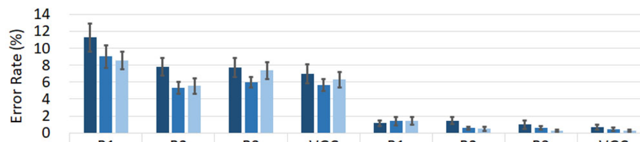
Visual Guidance Condition

Because the keyboard layouts were not typical, a novice user may spend a lot of time to discover a key. We were also uncertain of how fast a user will get used to the layout. This study was conducted to demonstrate the effect of trade-off between the number of touch gestures and the number of sub-keyboards, rather than the keyboard layout familiarity. Thus, we conducted one additional block with VGC to simulate users who are familiar with each keyboard layout. At the start of VGC, we explained the reason to the participants. Given instruction for the task was also the same as NC.



	block 1	block 2	block 3	VGC
GAT3	7.16	8.59	9.15	10.99
GAT6	8.08	9.82	9.75	11.26
GAT9	7.44	8.75	9.34	11.09

Figure 8. Text entry speed of each technique, per block. Error bars indicate standard errors.



	CER				UER			
	B1	B2	B3	VGC	B1	B2	B3	VGC
GAT3	11.27	7.80	7.75	7.01	1.16	1.45	1.00	0.70
GAT6	9.02	5.33	6.00	5.65	1.40	0.57	0.57	0.46
GAT9	8.56	5.56	7.36	6.29	1.43	0.50	0.27	0.30

Figure 9. Corrected error rate (CER) and uncorrected error rate (UER) of each technique/block. Error bars indicate standard errors.

Figure 7 shows a visual guidance. The background of a sub-keyboard, which contains a target key, was highlighted in blue, and the target key was highlighted in orange. The guide disappeared when a user started the touch input, and reappeared for the next target key once the key input was performed.

Results

We analyzed the results statistically for both NC and VGC, separately. The analysis methods for NC were almost similar to those of the preliminary study. For VGC, we used one-way RM-ANOVA with a layout factor, and also conducted three post-hoc tests, similar to the NC.

Text entry speed and corrected error rate

For NC, two-way (the technique and the block) RM-ANOVA analysis on text entry speed and on CER showed significant effects of the block (Speed: $F_{(2,34)}$= 156.903, $p< 0.01$, CER: $F_{(2,34)}$= 20.775, $p< 0.01$); however, the effects of layout and interaction were not significant. For VGC, the layout effects were not significant in either measurement.

Uncorected Error rate

For NC, a two-way RM-ANOVA with ART analysis indicated that all the effects and interaction were not significant. For VGC, the layout effect also was not significant.

Questionnaire

Friedman tests showed significant effects of the layouts on all of the four questions: **Easy to learn**, **Easy to use**, **Prefer to use**, and **Eye feels natural**: $p< 0.05$, 0.01, 0.01, and 0.01; and, $\chi^2_{(2)} = 8.561$, 15.524, 9.660, and 13.661, respectively.

Wilcoxon-signed rank tests with Bonferroni correction showed that GAT6 got significantly higher ratings than GAT3 for the different options; *i.e.*, **Easy to learn** ($Z = -2.449$, $p < 0.05$), **Easy to use** ($Z = -3.372$, $p < 0.01$), **Prefer to use** ($Z = -2.864$, $p < 0.05$). GAT9 also got a significantly higher rating than GAT3 for the **Easy to use** option ($Z = -2.967$, $p < 0.01$); however, for the option of **Eye feels natural**, GAT9 got a lower rating than GAT3 ($Z = -2.423$, $p < 0.05$).

Comments from the participants

Fifteen of the eighteen participants faced difficulty in using the diagonal gestures and six out of these fifteen also faced difficulty in using the vertical swipe for GAT3. However, four out of these six participants mentioned that the vertical swipe was easy to use in GAT6 owing to zero conflict between diagonal and vertical swipe gestures. Seven out of eighteen participants reported eye fatigue for GAT9, because they needed to move their eyes more to select or to find a key, than the other layouts.

Micro Analysis on Key Input Time

We analyzed the key input time to better understand the typing skills of users utilizing GAT, with only characters that were not erased and typed correctly (27,383 key entries).

We decomposed the key input time into four parts considering eye- and touch input phase; Eye movement data was logged every 30 ms. **T1** is the time from the moment the previous key was entered to the moment the eye cursor entered a target sub-keyboard containing a target key. **T2** is the time from the end of **T1** to the start of the last fixation (< 50 °/s) before a touch gesture started. **T3** is the time from the end of **T2** to the moment a touch gesture started. **T4** is the duration of the touch gesture, from touch to release. Figure 11 shows the average of the four parts in each block. As the results show, **T2** and **T3** were not dependent on the layout. On the other hand, **T1** increased and **T4** decreased as the number of sub-keyboards increased. Considering swipe gestures are slower than tap gesture [11], **T4** may decreased as the number of swipe gestures decreased, *e.g.,* GAT3 use eight out of nine touch gestures are swipe gestures, otherwise, GAT9 use two out of three touch gestures are swipe gestures (Figure 6).

To understand the observation on **T1**, we calculated the average number of sub-keyboard switches per letter inputted (**SWITCH**, Figure 11(e)). A **SWITCH** is counted whenever the eye cursor moved from a sub-keyboard, except when there was no eye fixation on the sub-keyboard. As shown in Figure 11, **T1** and **SWITCH** are highly correlated. Thus, we concluded that **T1** increased with more sub-keyboards because with the larger number of sub-keyboards, a user might have been exploring other sub-keyboards to find a target key. Subsequently, we simulated an optimal **SWITCH** for each layout with a used phrase set [21]. For example, for a word, "the", while using GAT3 (Figure 6(a)), "t" and "h" are on different sub-keyboards; therefore, one sub-keyboard switch is required; however, "h" and "e" are on the same sub-keyboard; hence, no sub-keyboard switch is required. Thus, the required number of sub-keyboard switches is one and the word length is three; therefore, the optimal **SWITCH** for "the" using GAT3 is 1/3 = 0.3. The optimal values of **SWITCH** for each layout are as follows: GAT3: 0.19, GAT6: 0.42, and GAT9: 0.58.

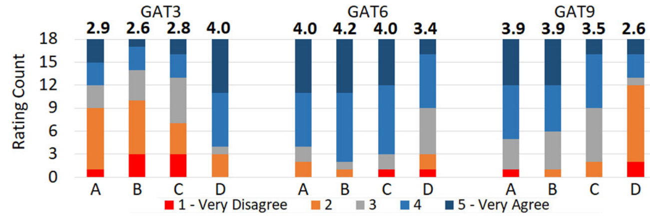


Figure 10. Cumulated rating counts on the questionnaire for each layout condition in Experiment 1. A: Easy to learn, B: Easy to use, C: Prefer to use, and D: Eye feels natural. Annotated values on the bars indicate the mean values on each option.

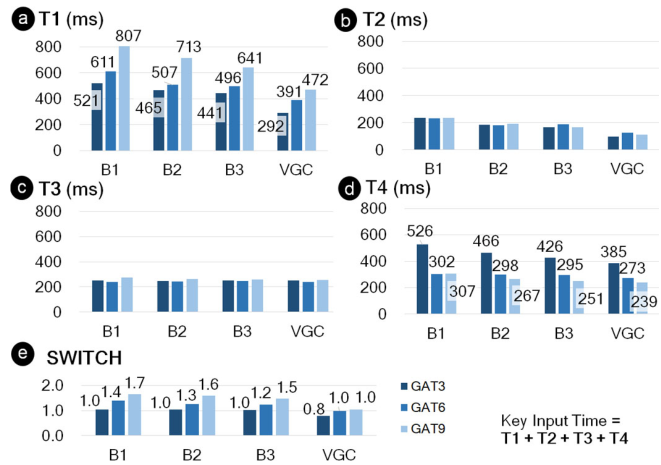


Figure 11. Four decomposed parts from the key input time −(a) T1, (b) T2, (c) T3, and (d) T4− and (e) SWITCH.

Considering these optimal values, the value of **SWITCH** is still far from an optimal level, even in the VGC (0.8, 1.0, and 1.0 for GAT3, GAT6, and GAT9, respectively). Thus, we argue that further reduction of **T1** after long-term utilization and/or incorporation of an optimized key placement that requires less **SWITCH**, can make GAT better.

EXPERIMENT 2: "GAT" EVALUATION ON "HWD"

To validate that GAT has better usability than the monomodal text entry methods, we compared GAT to eye-only typing (adjustable dwell time [22]) and touch-only typing (SwipeZone [11]) utilizing eye-trackable HWD.

Implementation

Owing to the differences in remote and wearable eye tracking environments, and the slippage and HWD wobbling problems of wearable eye trackers, we conducted Experiment 2 utilizing an eye-trackable HWD, FOVE VR[4]. In fact, the virtual reality environment and the smart glass-based environment may differ from each other. Even though it is not a smart glass, it was effective for simulating the wearable context of the smart glass.

The FOVE headset has 100° FoV and a resolution of the display is 2560×1440 pixels. The built-in eye tracker can track the gaze direction vector of both eyes at 120 Hz. The accuracy of the tracker is less than 1° in specification. However, according to a previous study that used FOVE to evaluate dwell-based eye typing [33], its mean accuracy was approximately 4°.

[4]https://www.getfove.com/

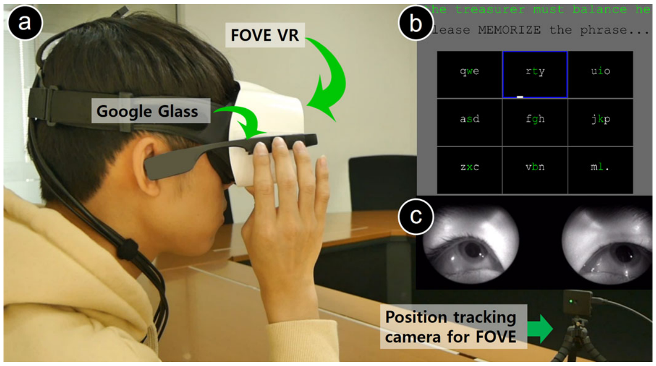


Figure 12. Experimental set-up for Experiment 2 (a) Eye tracking head-worn display (HWD) environment, FOVE VR. Google Glass is attached on the side of FOVE VR to use the touchpad. (b) Participant view in GAT session (c) Captured eye image while conducting the experiment.

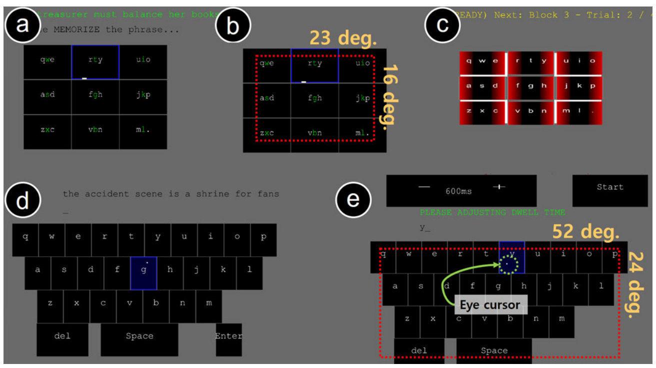


Figure 13. GUIs for the techniques (a) GAT in the phrase memorize phase. (b) Red box means required minimum field-of-view for GAT. (c) SwipeZone, touch-only, in the breaking phase inter tasks. (d) A-Dwell, eye-only, in the typing phase, and (e) in the dwell time adjustment phase.

Simulated gaze tracking smart glasses environment in VR

We constructed a VR environment that was similar to the smart glass-based environment. We set the FoV of the smart glass display to 100°, 85 × 85 cm sized display placed at a distance of 50 cm from the eyes. To calculate a gaze position, we used gaze direction vectors of both eyes, provided by FOVE software development kit. Then, we used a center position of the ray-casted positions of both eyes as the gaze position on the plane. To reduce the jitter, we also used an average filter.

Eye+Touch typing: GAT9

In Experiment 1, GAT3, GAT6, and GAT9 were equally effective; however, GAT6 was the most preferred. Regardless, we chose GAT9 to compare the other variations because GAT6 cannot use a QWERTY-like layout, which may be more familiar to users, and GAT3 requires complicated touch gestures.

Figure 13(a) shows the presented layout for GAT. We increased the keyboard size to more than that of Experiment 1 considering the mean accuracy of FOVE. Each sub-keyboard size is $10° \times 7°$ in the viewing angle. Thus, the dimensions of the GAT keyboard are $30° \times 21°$; however, the required minimum FoV is $23° \times 16°$ considering blank space as shown in Figure 13(b). It is a marginally usable size for the smart glasses, such as Moverio BT-350 (23°) and is also suitable for HWDs with large FoV such as Hololens (30°).

Touch-only typing: SwipeZone

There were few options for the touch-only typing baseline for smart glasses. We chose the fastest letter-wise text entry method, *i.e.*, SwipeZone [11].

SwipeZone is a two-step key selection method, similar to SwipeBoard, but using different sub-keyboard selection gestures. It uses three touch gestures −upper and lower directional swipe gestures, and tap gesture− on the discretized region of the touchpad, which is divided into three regions to select one of the nine sub-keyboards. If a user performed a tap gesture on the middle-positioned region, the centrally-positioned sub-keyboard, which includes "f g h", is activated. Conversely, if a user taps on the backward-positioned region, the sub-keyboard including "a s d" is activated. To help a user distinguish between the regions, we attached two strips of tape on the backward and forward regions of the touchpad to provide a tactile feedback, similar to the original Swipezone.

As shown in Figure 13(c), SwipeZone is nearly the same as the original. However, the original SwipeZone did not have the delete, space, and enter keys. Thus, we used the multifinger swipe gestures for the keys, which are same as GAT. In addition, the backward and forward directions of the touchpad matched the display in reverse from the original, to prevent confusion regarding the gestures for deletion and adding a space; thus, the matched direction same as GAT.

Eye-only typing: Adjustable dwell time selection

We chose the Adjustable dwell time selection method [22] as the eye only typing baseline, *i.e.*, A-Dwell. Because it is commonly considered outperformed than fixed dwell time technique. We follow dwell time adjust fomula used in the original [22], $Dwell = 300 \times EXP(X/12) - 150$, where as X, from 1 to 20, can be adjust in a digit (figure 13 (e)). We set a X to 11 ($Dwell = 600ms$) as a default value.

We used plain QWERTY layout for the eye only typing as shown in figure 13 (d). The size of each key is 5.5° x 7 °, and the gap between the key is 0.5°. Thus the dimension of the keyboard is 60° x 29°. We used dwell time feedback as similar to in the implementation of Dynamic cascading [28]. After a users̈ eye dwelling on a key over 30% of the dwell time, boundary of the key change its color to blue. After dwelling time over the dwell, the key is inputted and the color of the entire key set to blue. We also provided audio feedback for key entered, it is same to the other techniques.

Task and Participants

The task was same as that of the previous studies. We recruited 12 participants (3 females, 9 males, mean age: 22.33, from 18 to 28 years) from our university. All of them were different from the previous studies, right-handed, non-native English speakers, with no experience in using a smart glasses or eye tracker. Ten out of twelve had at least one experience in using a VR HWD. Five out of twelve were wearing contact lenses. All participants received $44 USD as compensation.

Design and Procedure

We used the within-subject design with the technique as a factor; touch-only (SwipeZone), eye-only (A-Dwell), and GAT.

The experiment consisted of three sessions per day; the participants participated for two days in sequence. In each session, three blocks were performed using one of the three techniques. In each block, a participant completed four tasks sequentially.

At the start of each session, we calibrated the eye tracker with the default calibration method of FOVE. In addition, a nine-point calibration process was performed again if required. If a user experienced a mismatch in the eye cursor position, we recalibrated the eye tracker and discarded the task in progress.

The order of sessions was counter balanced with a full factorial within subjects. There were at least 3-minute breaks between sessions, 1-minute breaks between blocks, and 15-second breaks between tasks. At the end of the last session, we surveyed a questionnaire with five questions using a five-point Likert scale; four of the questions were same to Experiment 1; and, the fifth option was **Eye fatigue**. The experiment was conducted for approximately two hours per day, and participants utilized 53 min, 49 min, and 41 min for eye-only, touch-only, and GAT, respectively, which included the break times between tasks and blocks.

Finally, we gathered 12 participants $\times$ 2 days $\times$ 3 sessions (techniques) $\times$ 3 blocks $\times$ 4 tasks $=$ 864 phrases. However, four participants could not complete all tasks for the eye-only technique owing to a time constraint. A total of 56 tasks for the eye-only technique was either discarded or not completed.

Results

The analysis methods were nearly the same as those for Experiment 1; however, we conducted three post-hoc tests only in the last block, and we used a one-tailed assumption for the tests because we assumed that GAT outperforms the others.

In the eye-only technique, four participants did not complete all the blocks. To conduct statistical tests, we simulated the missing data. The simulated data were generated by normal distribution of the mean and standard deviation from the top three performers, in terms of typing speed of the eye-only technique for each missing block.

Text entry speed

The RM-ANOVA analysis on text entry speed showed significant effects of the technique ($F_{(2,22)}$= 11.020, $p< 0.01$) and block ($F_{(2.696,29.654)}$ = 48.822, $p< 0.01$). The interaction effect was also significant ($F_{(4.351,47.865)}$= 3.055, $p< 0.05$).

The post-hoc tests showed that GAT is significantly faster than both touch-only (t_{11}= 6.919, $p< 0.01$) and eye-only (t_{11}= 3.882, $p< 0.01$) techniques. GAT (mean: 11.04 wpm, from 9.10 to 14.13 wpm, individually) was 29.4% faster than touch-only (mean: 8.53 wpm, from 5.24 to 11.01 wpm) and 25.5% faster than eye-only (mean: 8.81 wpm, from 6.27 to 11.85 wpm) techniques at the last block.

Error rate

The RM-ANOVA analysis on CER showed significant effects of block ($F_{(5,55)}$= 3.260, $p< 0.05$); however, the effects of technique and interaction were not significant.

The RM-ANOVA with ART analysis on UER showed that both the effects and the interaction were not significant.

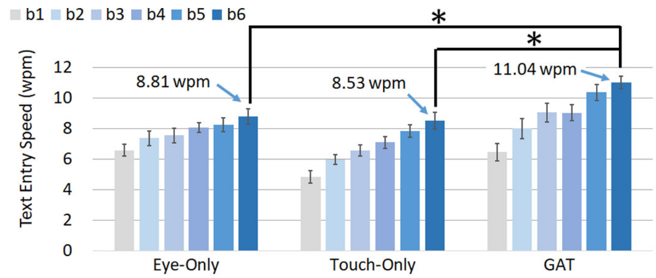


Figure 14. Text entry speed of each technique/block. Error bars indicate standard errors. The annotated values indicate average text entry speeds at the last block of each techniques.

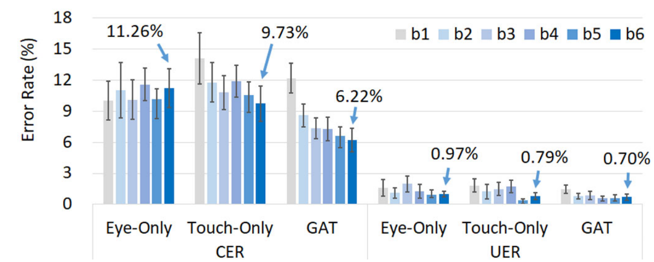


Figure 15. CER and UER of each technique/block. Error bars indicate standard errors. The annotated values indicate average CERs and UERs at the last block of each techniques.

Questionnaire

Friedman tests showed significant effects of the technique on **Easy to learn**, **Prefer to use**, and **Eye fatigue**: $\chi^2_{(2)}$ = 8.667, 6.826, and 18.050, $p< 0.05$, 0.05, and 0.01, respectively. Therefore, we conducted post-hoc Wilcoxon signed-rank tests with Bonferroni correction.

GAT got significantly better ratings than touch-only for **Easy to learn** (Z= -2.565, $p< 0.05$), and better ratings than eye-only for **Prefer to use** (Z= -2.280, $p< 0.05$) and **Eye fatigue** (Z= -2.980, $p< 0.01$). Touch-only technique got significantly better ratings than both eye-only and GAT for **Eye fatigue**: Z= -2.850 and -2.165, $p< 0.01$ and 0.05, respectively.

Friedman tests did not show significant effects on the other two questions. However, three out of six Wilcoxon signed-rank tests represent medium effect sizes ($r> 0.3$) as following: GAT received more positive ratings on **Easy to use** than both touch-only (Z= -2.280, r= 0.33) and eye-only (Z= -2.095, r= 0.30). Touch only technique received more positive ratings on **Eye feels natural** than eye-only: Z= -2.420, r= 0.35.

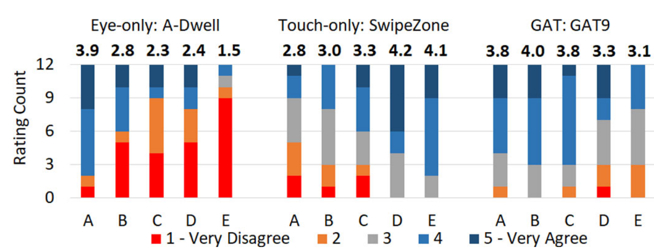


Figure 16. Cumulated rating counts on the questionnaire for each technique in Experiment 2. A: Easy to learn, B: Easy to use, C: Prefer to use, D: Eye feels natural, E: Eye Fatigue (Red means increasing fatigue). Annotated values on the bars indicate the mean values on each option.

Reasons for Preference

For the "Prefer to use" question, we asked the reason for the highest/lowest rating. If the highest/lowest score was not the only one, we asked the participants to choose the best/worst technique along with the reason for this choice.

Eye + Touch: GAT

Seven out of twelve participants chose GAT as the best technique because it was the fastest. When compared to eye-only, GAT required less eye movements and demonstrated better accuracy. When compared to touch-only, the participants experienced less physical fatigue owing to the lesser number of keystrokes per character. It was also easy to use because GAT does not use vertical swipe gestures and there was no requirement to distinguish a touched region on the touchpad for selecting a sub-keyboard. No one chose GAT as the worst.

Touch-only: SwipeZone

Three of the five participants chose touch-only as the best technique. They mentioned that eye-only and GAT techniques required a focus on eye movements; further, errors often occurred because of unintentional eye movements. They felt that the use of eye input methods was unnatural; moreover, the touch-only technique became familiar by the second day. However, three out of twelve chose touch-only as the worst. This is because it was not only slow, but also difficult to perform the touch gestures. They also experienced more wrist and/or finger fatigue than while using GAT.

Eye-only: A-Dwell

The other two participants chose eye-only as the best. They did not want to use the touchpad on the head and felt no fatigue of both the arm and eyes. However, nine out of twelve chose eye-only as the worst. The most crucial reason was eye fatigue. Some also felt stress and nervousness caused by out-of-control movement of the eye cursor due to drift problem over time. Some also suffered from inaccuracy of eye tracking when they looked at a key in the peripheral region, such as "q" and "p".

DISCUSSION

Validity of the Simulated Environment

We conducted the experiments in simulated environments. Experiment 1 (Exp 1) and 2 (Exp 2) were conducted with the stationary eye tracker in a desk-based environment and the wearable eye tracker in a VR environment, respectively.

If we consider that the eye-tracking accuracy may affect the performance of GAT variations, the results of Exp 1 may differ in the wearable context, because a wearable eye tracker may be less accurate than a stationary one. However, we determined that the performances of GAT9 in the stationary setup (Exp 1, five tasks/block) and in the wearable setup (Exp 2, four tasks/block) were consistent in terms of both speed (9.3 wpm vs. 9.0 wpm, respectively, at the 3rd block) and accuracy (CER: 6.3% vs. 7.4% and UER: 0.3% vs. 0.9%). Assuming the stationary setup might have been favorable to GAT9 than the others, the main results of Exp 1 would have remained essentially the same if a wearable setup had been used.

Although Exp 2 was conducted with a VR HWD rather than with an augmented reality (AR) glasses, it was effective for simulating the wearable computing context of the smart glasses. The main differences between VR and AR may be in the visual fidelity of the environment. Our future work will handle such environmental validity issues using an eye-tracking AR glasses which can meet the requirements of GAT.

Performance Comparison with Literature

We demonstrated that GAT (11.04 wpm, after 40 min of usage) is faster than SwipeZone (8.53 wpm, after 50 min of usage) in Experiment 2. The result of SwipeZone was comparable to a previous study (8.73 wpm, after 80 min of usage) [11].

However, when compared to eye-only typing, the typing speed may be questionable as the previous studies were conducted in a desk-based environment. In Experiment 2, GAT outperformed an eye-typing method, A-Dwell (8.81 wpm, after 50 min of usage), when using the mobile eye tracker after 40 min of use. However, in a previous study [22], A-Dwell achieved 20 wpm over 10 days of training. Dynamic cascading (cascading, 13.70 wpm after eight days training; over 12 wpm on the second day) also outperformed A-Dwell (12.08 wpm after eight days; 10 wpm on the second day) [28].

Even though the learning curve of GAT was steeper than A-Dwell, a longitudinal study was required to compare both techniques after long-term training. Consequently, Experiment 2 was conducted, which demonstrated that GAT causes less eye fatigue and is sufficiently more robust than the eye-only text entry when considering the inaccuracy of eye tracker; consequently, it can be used with a wearable eye tracker.

Scalability of GAT

In our study, we considered first the input modalities for the conventional gesture input and finalized on a touchpad. However, the gesture input for GAT can be performed by not only other touch input devices, such as smartwatch, handheld touchpad, or on-body touch input techniques, but also using freehand gestures, such as mid-air finger gesture [19] or wrist tilting gestures [10]; this incorporates more subtle gestures, considering the social context and/or utilizing more comfortable postures rather than raising a hand to the glasses.

CONCLUSION

We presented the concept of GAT and the first adaptation of a complementary combination of the gaze and touch input modalities for wearable text entry. We demonstrated that GAT can ensure faster on-glass text entry when compared to the touch-only text entry. Further, we explored the variations in the number of touch gestures. According to our micro analysis results, the gesture completion time (**T4**) increased with the number of touch gestures. Conversely, the target sub-keyboard reaching time (**T1**) decreased with the increasing number of touch gestures, corresponding to the increasing frequency of the sub-keyboard change. Consequently, the text entry efficiencies were not different from each other. Finally, we demonstrated that GAT outperformed the single input methods in the HWD environment as GAT (11.04 wpm) was 25.4% faster than the eye-only typing technique (8.81 wpm) and 29.4% faster than the touch-only typing method (8.53 wpm); moreover, it was the most preferred technique.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research-Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01058992).

REFERENCES

[1] Sunggeun Ahn, Seongkook Heo, and Geehyuk Lee. 2017. Typing on a Smartwatch for Smart Glasses. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces (ISS '17)*. ACM, New York, NY, USA, 201–209. DOI: http://dx.doi.org/10.1145/3132272.3134136

[2] Behrooz Ashtiani and I. Scott MacKenzie. 2010. BlinkWrite2: An Improved Text Entry Method Using Eye Blinks. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*. ACM, New York, NY, USA, 339–345. DOI: http://dx.doi.org/10.1145/1743666.1743742

[3] L. R. Bahl, R. Bakis, J. Bellegarda, P. F. Brown, D. Burshtein, S. K. Das, P. V. de Souza, P. S. Gopalakrishnan, F. Jelinek, D. Kanevsky, R. L. Mercer, A. J. Nadas, D. Nahamoo, and M. A. Picheny. 1989. Large vocabulary natural language continuous speech recognition. In *International Conference on Acoustics, Speech, and Signal Processing,*. 465–467 vol.1. DOI: http://dx.doi.org/10.1109/ICASSP.1989.266464

[4] Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+Gesture: Expressive, Precise and Targeted Free-Space Interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*. ACM, New York, NY, USA, 131–138. DOI: http://dx.doi.org/10.1145/2818346.2820752

[5] Xiang 'Anthony' Chen, Tovi Grossman, and George Fitzmaurice. 2014. Swipeboard: A Text Entry Technique for Ultra-small Interfaces That Supports Novice to Expert Transitions. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)*. ACM, New York, NY, USA, 615–620. DOI: http://dx.doi.org/10.1145/2642918.2647354

[6] Sangwon Choi, Jaehyun Han, Geehyuk Lee, Narae Lee, and Woohun Lee. 2011. RemoteTouch: Touch-screen-like Interaction in the Tv Viewing Environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 393–402. DOI: http://dx.doi.org/10.1145/1978942.1978999

[7] Edward Clarkson, James Clawson, Kent Lyons, and Thad Starner. 2005. An Empirical Study of Typing Rates on mini-QWERTY Keyboards. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1288–1291. DOI:http://dx.doi.org/10.1145/1056808.1056898

[8] John J. Dudley, Keith Vertanen, and Per Ola Kristensson. 2018. Fast and Precise Touch-Based Text Entry for Head-Mounted Augmented Reality with Variable Occlusion. *ACM Trans. Comput.-Hum. Interact.* 25, 6, Article 30 (Dec. 2018), 40 pages. DOI: http://dx.doi.org/10.1145/3232163

[9] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. ACM, New York, NY, USA, 457–466. DOI: http://dx.doi.org/10.1145/2807442.2807499

[10] Jun Gong, Zheer Xu, Qifan Guo, Teddy Seyed, Xiang 'Anthony' Chen, Xiaojun Bi, and Xing-Dong Yang. 2018. WrisText: One-handed Text Entry on Smartwatch Using Wrist Gestures. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 181, 14 pages. DOI: http://dx.doi.org/10.1145/3173574.3173755

[11] Tovi Grossman, Xiang Anthony Chen, and George Fitzmaurice. 2015. Typing on Glasses: Adapting Text Entry to Smart Eyewear. In *Proc. MobileHCI '15*. ACM, 144–152. DOI: http://dx.doi.org/10.1145/2785830.2785867

[12] Aakar Gupta, Cheng Ji, Hui-Shyong Yeo, Aaron Quigley, and Daniel Vogel. 2019. RotoSwype: Word-Gesture Typing Using a Ring. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 14, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300244

[13] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1063–1072. DOI: http://dx.doi.org/10.1145/2556288.2557130

[14] Yi-Ta Hsieh, Antti Jylhä, Valeria Orso, Luciano Gamberini, and Giulio Jacucci. 2016. Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses. In *Proc. CHI '16*. ACM, 4203–4215. DOI: http://dx.doi.org/10.1145/2858036.2858436

[15] 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. DOI: http://dx.doi.org/10.1145/123078.128728

[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 (CHI '16)*. ACM, New York, NY,

USA, 1952–1956. DOI: http://dx.doi.org/10.1145/2858036.2858335

[17] 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 (CHI '18)*. ACM, New York, NY, USA, Article 81, 14 pages. DOI:http://dx.doi.org/10.1145/3173574.3173655

[18] Shu-Yang Lin, Chao-Huai Su, Kai-Yin Cheng, Rong-Hao Liang, Tzu-Hao Kuo, and Bing-Yu Chen. 2011. Pub - Point Upon Body: Exploring Eyes-free Interaction and Methods on an Arm. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 481–488. DOI: http://dx.doi.org/10.1145/2047196.2047259

[19] Mingyu Liu, Mathieu Nancel, and Daniel Vogel. 2015. Gunslinger: Subtle Arms-down Mid-air Interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. ACM, New York, NY, USA, 63–71. DOI: http://dx.doi.org/10.1145/2807442.2807489

[20] K. Lyons, D. Plaisted, and T. Starner. 2004. Expert chording text entry on the Twiddler one-handed keyboard. In *Eighth International Symposium on Wearable Computers*, Vol. 1. 94–101. DOI: http://dx.doi.org/10.1109/ISWC.2004.19

[21] I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase Sets for Evaluating Text Entry Techniques. In *Proc. CHIEA'03 (CHI EA '03)*. ACM, 754–755. DOI: http://dx.doi.org/10.1145/765891.765971

[22] 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 (CHI '09)*. ACM, New York, NY, USA, 357–360. DOI: http://dx.doi.org/10.1145/1518701.1518758

[23] Päivi Majaranta and Andreas Bulling. 2014. Eye tracking and eye-based human–computer interaction. In *Advances in physiological computing*. Springer, 39–65.

[24] Meethu Malu and Leah Findlater. 2015. Personalized, Wearable Control of a Head-mounted Display for Users with Upper Body Motor Impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 221–230. DOI: http://dx.doi.org/10.1145/2702123.2702188

[25] Anders Markussen, Mikkel Rønne Jakobsen, and Kasper Hornbæk. 2014. Vulture: A Mid-air Word-gesture Keyboard. In *Pro. CHI '14*. ACM, 1073–1082. DOI: http://dx.doi.org/10.1145/2556288.2556964

[26] M. Menozzi, A. v. Buol, H. Krueger, and Ch. Miége. 1994. Direction of gaze and comfort: discovering the relation for the ergonomic optimization of visual tasks. *Ophthalmic and Physiological Optics* 14, 4 (1994), 393–399. DOI: http://dx.doi.org/10.1111/j.1475-1313.1994.tb00131.x

[27] Carlos H. Morimoto and Arnon Amir. 2010. Context Switching for Fast Key Selection in Text Entry Applications. In *Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications (ETRA '10)*. ACM, New York, NY, USA, 271–274. DOI: http://dx.doi.org/10.1145/1743666.1743730

[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 (CHI '17)*. ACM, New York, NY, USA, 2558–2570. DOI:http://dx.doi.org/10.1145/3025453.3025517

[29] Shahriar Nirjon, Jeremy Gummeson, Dan Gelb, and Kyu-Han Kim. 2015. TypingRing: A Wearable Ring Platform for Text Input. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '15)*. ACM, New York, NY, USA, 227–239. DOI: http://dx.doi.org/10.1145/2742647.2742665

[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 (UIST '14)*. ACM, New York, NY, USA, 509–518. DOI: http://dx.doi.org/10.1145/2642918.2647397

[31] Ken Pfeuffer and Hans Gellersen. 2016. Gaze and Touch Interaction on Tablets. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 301–311. DOI: http://dx.doi.org/10.1145/2984511.2984514

[32] Vijay Rajanna. 2016. Gaze Typing Through Foot-Operated Wearable Device. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '16)*. ACM, New York, NY, USA, 345–346. DOI: http://dx.doi.org/10.1145/2982142.2982145

[33] Vijay Rajanna and John Paulin Hansen. 2018. Gaze Typing in Virtual Reality: Impact of Keyboard Design, Selection Method, and Motion. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (ETRA '18)*. ACM, New York, NY, USA, Article 15, 10 pages. DOI: http://dx.doi.org/10.1145/3204493.3204541

[34] Julie Rico and Stephen Brewster. 2010. Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability. In *Proc. CHI'10*. ACM, 887–896. DOI: http://dx.doi.org/10.1145/1753326.1753458

[35] Srinath Sridhar, Anna Maria Feit, Christian Theobalt, and Antti Oulasvirta. 2015. Investigating the Dexterity of Multi-Finger Input for Mid-Air Text Entry. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3643–3652. DOI: http://dx.doi.org/10.1145/2702123.2702136

[36] Takumi Toyama, Thomas Kieninger, Faisal Shafait, and Andreas Dengel. 2012. Gaze Guided Object Recognition Using a Head-mounted Eye Tracker. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA '12)*. ACM, New York, NY, USA, 91–98. DOI:http://dx.doi.org/10.1145/2168556.2168570

[37] Takumi Toyama, Daniel Sonntag, Andreas Dengel, Takahiro Matsuda, Masakazu Iwamura, and Koichi Kise. 2014. A Mixed Reality Head-mounted Text Translation System Using Eye Gaze Input. In *Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI '14)*. ACM, New York, NY, USA, 329–334. DOI: http://dx.doi.org/10.1145/2557500.2557528

[38] Outi Tuisku, Veikko Surakka, Ville Rantanen, Toni Vanhala, and Jukka Lekkala. 2013. Text Entry by Gazing and Smiling. *Adv. in Hum.-Comp. Int.* 2013, Article 1 (Jan. 2013), 1 pages. DOI: http://dx.doi.org/10.1155/2013/218084

[39] Cheng-Yao Wang, Wei-Chen Chu, Po-Tsung Chiu, Min-Chieh Hsiu, Yih-Harn Chiang, and Mike Y. Chen. 2015. PalmType: Using Palms As Keyboards for Smart Glasses. In *Proc. MobileHCI'15*. ACM, 153–160. DOI: http://dx.doi.org/10.1145/2785830.2785886

[40] David J. Ward, Alan F. Blackwell, and David J. C. MacKay. 2000. 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 (UIST '00)*. ACM, New York, NY, USA, 129–137. DOI: http://dx.doi.org/10.1145/354401.354427

[41] Eric Whitmire, Mohit Jain, Divye Jain, Greg Nelson, Ravi Karkar, Shwetak Patel, and Mayank Goel. 2017. DigiTouch: Reconfigurable Thumb-to-Finger Input and Text Entry on Head-mounted Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 113 (Sept. 2017), 21 pages. DOI: http://dx.doi.org/10.1145/3130978

[42] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 143–146. DOI: http://dx.doi.org/10.1145/1978942.1978963

[43] 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 (ETRA '08)*. ACM, New York, NY, USA, 11–18. DOI: http://dx.doi.org/10.1145/1344471.1344475

[44] Hui-Shyong Yeo, Xiao-Shen Phang, Taejin Ha, Woontack Woo, and Aaron Quigley. 2017. TiTAN: Exploring Midair Text Entry Using Freehand Input. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 3041–3049. DOI:http://dx.doi.org/10.1145/3027063.3053228

[45] Xin Yi, Chun Yu, Mingrui Zhang, Sida Gao, Ke Sun, and Yuanchun Shi. 2015. ATK: Enabling Ten-Finger Freehand Typing in Air Based on 3D Hand Tracking Data. In *Proc. UIST'15*. ACM, 539–548. DOI: http://dx.doi.org/10.1145/2807442.2807504

[46] Chun Yu, Ke Sun, Mingyuan Zhong, Xincheng Li, Peijun Zhao, and Yuanchun Shi. 2016. One-Dimensional Handwriting: Inputting Letters and Words on Smart Glasses. In *Proc. CHI'16*. ACM, 71–82. DOI: http://dx.doi.org/10.1145/2858036.2858542

[47] 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 (CHI '99)*. ACM, New York, NY, USA, 246–253. DOI: http://dx.doi.org/10.1145/302979.303053