EYEditor: Towards On-the-Go Heads-up Text Editing Using Voice and Manual Input

Debjyoti Ghosh[1,2], Pin Sym Foong[3], Shengdong Zhao[2], Can Liu[4], Nuwan Janaka[2], Vinitha Erusu[2]

[1]NUS Graduate School for Integrative Sciences and Engineering
[2]NUS-HCI Lab, School of Computing
[3]Saw Swee Hock School of Public Health
National University of Singapore, Singapore
debjyoti@u.nus.edu, pinsym@nus.edu.sg,
{zhaosd, nuwanj, vinithar}@comp.nus.edu.sg

[4]School of Creative Media
City University of Hong Kong
Kowloon, Hong Kong
canliu@cityu.edu.hk

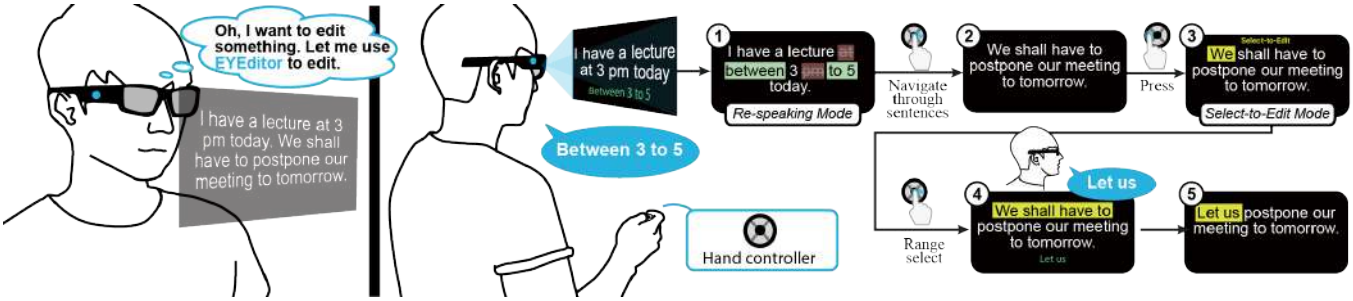


Figure 1: EYEditor interactions: User sees the text on a smart glass, sentence-by-sentence. In the *Re-speaking* mode, correction is achieved by re-speaking over the text and a hand-controller is used to navigate between sentences. Users can enter the *Select-to-Edit* mode to make fine-grained selections on the text and then speak to modify the selected text.

ABSTRACT

On-the-go text-editing is difficult, yet frequently done in everyday lives. Using smartphones for editing text forces users into a heads-down posture which can be undesirable and unsafe. We present EYEditor, a heads-up smartglass-based solution that displays the text on a see-through peripheral display and allows text-editing with voice and manual input. The choices of output modality (visual and/or audio) and content presentation were made after a controlled experiment, which showed that sentence-by-sentence visual-only presentation is best for optimizing users' editing and path-navigation capabilities. A second experiment formally evaluated EYEditor against the standard smartphone-based solution for tasks with varied editing complexities and navigation difficulties. The results showed that EYEditor outperformed smartphones as either the path OR the task became more difficult. Yet, the advantage of EYEditor became less salient when both the editing and navigation was difficult. We discuss trade-offs and insights gained for future heads-up text-editing solutions.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376173

Author Keywords

Heads-up Interaction; Smart glass; Text editing; Voice Interaction; EYEditor; Wearable Interaction; Mobile Interaction; Re-speaking; Manual-input.

CCS Concepts

•Human-centered computing → Interaction techniques; Text input; User interface design; Ubiquitous computing;

INTRODUCTION

Word processing on mobile phones has become an essential everyday task in our modern lives. According to Statistic Brain, 781 billion text messages are sent every month in the United States alone [10] which is more than 26 billion texts a day or 94 texts per person each day [11]. This astonishing number does not even count the other text-based tasks people perform on the phone, i.e., online searches, writing and replying to social media posts, emails, etc.

Although mobile phone-based word processing helps greatly with people's on-the-go information and communication needs, it also contributes to a notorious phenomena called 'smart-phone zombie' or 'heads-down tribe' [52]. As Mark Sharp, a journalist, describes it: *"The zombies are everywhere. They wander the streets, shopping malls and MTR [metro] corridors, heads down and oblivious to the world around them."* Such heads-down style has a number of undesirable consequences as it: 1) isolates the user from the environment as

the user cannot actively observe the environment and the people around them [33]; 2) text entry on the phone requires a significant amount of dedicated attention and precise motor movements, and imposes high visual, temporal, and physical demands on the user [32, 41]; and 3) forces the user into an awkward posture, which twists the spine [44] and leads to unwarranted exertion of the hand muscles; these can cause significant health problems in the long run [25, 21, 44]. Despite these serious problems, it is difficult to deny the strong information and communication needs that people have on the go [35]. Hence, it is necessary to work out alternative solutions that can enable users to perform text-based tasks, albeit avoiding the discussed problems.

One potential solution to alleviate the above problems is to perform word processing tasks on smart glasses with transparent heads-up displays — systems that allow users to access digital information in real-time while simultaneously being present in the physical world [34, 15]. In that, smart glasses allow heads-up interaction with faster attention switching between the digital display and the visual surrounding. However, due to the lack of input mechanisms, word processing on smart glasses is more difficult than on the phone [36]. Past research has explored text-entry mechanisms on smart glasses that typically allow typing text on a small touch panel on the side of the glasses and require the user's hand to be held-up to enter the text. Although some of these methods achieved relatively fast typing speeds of up to 25 wpm [65], using them still imposes significant visual, temporal and physical demands. Voice input can potentially solve this problem, as it is hands-free, natural to use, and have demonstrated input speeds that are 3 to 5 times that of touch-based input [47, 4]. Yet, while voice is convenient for generating text, it is difficult to edit text with voice. Editing requires spatial referencing to delimit *where* and *how much* of the text needs to be changed. These operations are difficult to perform using voice alone [4].

In this paper, to overcome some of the previously mentioned problems, we propose an on-the-go heads-up text editing solution, EYEditor, which allows editing text on a smart glass. We focus on text-editing as the scope of our investigation because: 1) it is more complex and demanding than text entry due to the increased number of constraints involved in error detection, localization and correction; 2) it is an essential component of word processing, as without it, a text entry method is less useful for serious use and yet, text editing on smart glasses is under-explored.

EYEditor adopts a hybrid approach of voice and manual input. Voice is used to modify the text content, while manual input through a wearable ring-mouse is used for text navigation and selection. Text content is rendered visually on the smart glass screen with a sentence-by-sentence presentation. This design choice is determined by a controlled study comparing three combinations of audio and/or visual output with the visual rendering done in two different presentations: sentence-by-sentence and block text display.

To test the feasibility, desirability, and viability of EYEditor on the go, we conducted a second study comparing it with the status-quo smartphone-based text editing technique. Participants used both our system and the phone to perform simple and difficult correction tasks while walking on three different path-types. Results showed that EYEditor offered significant benefits over the phone as the task OR the path difficulty increased — participants could correct text significantly faster while maintaining a higher average walking speed when using EYEditor. However, this performance gain over the phone narrowed when both the path AND the task demanded high visual attention, where participants faced challenges with both the techniques.

Our contribution is threefold: (1) design of EYEditor, a system to facilitate on-the-go text-editing on a heads-up display; (2) a quantitative evaluation of output modalities on the smart glass and an in-depth understanding of how each affects the user's text-correction and path-navigation abilities on the go; and (3) a comparative evaluation of our proposed technique against the smartphone baseline, based on which we discuss the trade-offs and insights gained to inspire the design of future on-the-go, heads-up text-editing solutions.

RELATED WORK

There are three broad areas that our work relates to:

Smart glasses as an emergent platform

Smart glasses have emerged as an important platform to interact with digital content on the go, due to their unobtrusiveness and affordance of maintaining direct visual contact with the physical surrounding [5, 39]. To further promote its adoption, research needs to invest in uncovering the potentials that smart glasses bring as a new paradigm of interaction. Rauschnabel et al. [46] theorised that adoption of smart glasses would be dependent upon at least one of three factors: 1) *Effectance*: what value does it bring in making one's life more efficient? 2) *Hedonic*: its use in providing fun and entertainment; and 3) *Social*: To what degree can it maintain or foster social interactions and relationships. Much research has been done to boost the effectance of smart glasses. They have been used in industrial applications [38, 31, 62], outdoor training [58, 57], touring applications [16, 7], clinical and surgical applications [1, 40], education [28], product development and logistics [46]. In this paper, we focus on an under-explored use case of smart glasses — text editing: to support this task on the go while maintaining awareness of the path and surroundings.

Text interaction on smart glasses

Research on text entry mechanisms with smart glasses have offered many techniques to work around the limited interaction possibilities of the small screen and absence of keyboards. Strategies have included touchpad [19, 66], mid-air [3, 18, 20, 24], hand [54, 9, 2, 50], wrist, palm[61] and finger-based input mechanisms [6, 64]. More visual strategies include dwell-free techniques [27], and techniques that replace dwell operations with movement of the eye-pointer [29, 49]. Additionally, head-based text entry has been shown to achieve relatively high input rates of $\approx$25 WPM [65]. However these systems have not been tested in on-the-go scenarios and it is unclear if the interaction burden the systems impose is too high for on-the-go scenarios. Additionally, text entry is a different task from

text editing, as the latter calls on other cognitive functions for error detection, localization and correction.

For text output, previous work has explored text content presentation that optimizes users' reading experience on the go. Vadas et al. [59] compared visual and auditory displays for text comprehension on the go and found that audio output was more suitable for path-navigation. Also, text comprehension with audio output was at par with visual output. Rzayev et al. [48] explored the effect of two presentation types on text comprehension while walking: Rapid Serial Visual Presentation (RSVP) and line-by-line scrolling (3 words per line). They found that line-by-line scrolling yields higher comprehension than RSVP, while walking. However, these results are specific to reading and might not hold true for text editing as editing calls on additional cognitive functions as discussed before, and hence, needs further investigation.

Voice-based Error Correction

Azenkot and Lee [4] had found that speech was nearly 5 times as fast as keyboard-based text entry, but the efficiency was reduced by the combined difficulty of reviewing and correcting speech recognition errors in the absence of specialized voice-based editing algorithms. Editing requires spatial referencing to delimit *where* and *how much* of the text needs to be changed [12]. Voice-based dictation applications like Dragon NaturallySpeaking support a two-step process where the user says a first command to make a selection, then dictate to modify the selection. Ghosh et al. [17] found that merging the two steps was useful for eyes-free editing. To preclude the mental effort in remembering commands and in having to speak the erroneous text (which may be ungrammatical or illogical) [60], McNair and Waibel [37] had first proposed a more natural, one-step correction approach. This approach let users re-speak over erroneous parts of the text to change it. The change is effected by an alignment algorithm that tries to align the user utterance to existing parts of the text. Multiple research has explored computational models to improve the alignment accuracy [60, 13, 53]. We adapted the re-speaking approach with an in-house implementation for correcting text real-time on our smartglass-based system.

There exists a body of literature exploring multi-modality in voice-based error correction. Halverson et al. [22] studied user patterns of voice-based error correction in desktop speech systems and found that using a single modality of input might lead to spiral depths [43] and cascades [22] that slows down the correction process. They suggested switching input modalities from voice-only to voice+mouse or voice+keyboard to cut down on the error-correction time. Suhm et al. [55] also found that the use of multi-modal input improves the error-correction speed. Oviatt [42] suggested that multiple input modalities might benefit speech recognition. Based on the discussed literature, we propose to support multi-modal input in our system design by combining voice (for editing the text) with manual input (for text selection and navigation).

FIRST DESIGN OF EYEditor

Our system combines three interfaces - visual (output), auditory (input-output) and manual input (Figure 2). The voice

Figure 2: Apparatus showing all three interfaces.

and manual input are processed separately in a processing unit which applies the user's voice-based correction to the text and sends instructions to the visual/auditory output on how the edited content should be presented to the user.

Visual Interface

We used a Vuzix Blade (henceforth, just 'Blade') see-through smart glass. The Blade is regarded as one of the most commercially viable and recommended smart glasses in the market [56]. It has a 480x480 px display, vertically centered on the right glass and runs a web-server running on Android 5.1.

We developed a host server (Node.js server running on a MacBook Pro, 2017) that subscribes to the Blade server via a socket connection. The host server pushes the text content and formatting instructions to the Blade server which then renders the formatted text on an Android app, running on the Blade.

EYEditor's screen-space is divided into two parts: the *text-content space* that can render up to 8 lines of text with word-wrapping (≈21 characters per line), and the *speech transcription space* which reserves 2 lines to show a live transcription (speech-to-text) of the users' utterances. For optimum readability while rendering single sentences on the Blade screen, the text is centralized both horizontally and vertically [48].

Audio Interface

EYEditor supports voice-based correction of text through an audio interface. The audio interface supports input/output using a pair of Bose QC35 headphones (with microphone) connected via Bluetooth to the MacBook Pro. We used the MDN Web Speech API [63] for both speech-transcription of user utterances and speech-synthesis to deliver the system audio output. After pilot tests to determine the users' comfortable listening comprehension rate for text editing, the audio rate was fixed at 0.7x for text-content playback.

Voice-correction of the text can be achieved by *re-speaking* over erroneous parts of the text. For example, to correct 'quick red fox', user says, "quick brown fox". In the previous example, 'red' gets *repaired* to 'brown'; 'quick' and 'fox' are the left/right repair-contexts, i.e., matching word(s) to the left/right of the intended repair. A more detailed discussion on

correction-by-respeaking is presented in the Appendix. Additionally, our system supports command-based deletion of text using the DELETE keyword, e.g., "DELETE <*phrase*>".

Manual-Input Interface

Previous research has shown that navigation-based tasks are faster, easier and more accurate with manual input than speech [8, 51]. Therefore, we introduced a small ring-mouse (Sanwa Supply 400-MA077) hand-controller to our system for text navigation purposes. The controller has 4 buttons and a central trackpad that supports both swipe gestures and button-press. For the purpose of this study, we reprogrammed the default controls of the mouse. The first design of EYEditor supported two functions: 1) going forward/backward in the text by swiping right/left on the trackpad; and 2) undo/redo operations by pressing/long-pressing the right button.

STUDY 1: OUTPUT MODALITY AND CONTENT PRESENTATION

One essential question to answer is: How should we present the information to the users using the new smart glass platform for on-the-go text editing? Our goal is to balance the users' path-navigational needs and text editing needs on the go. Therefore, we need to find a solution that offers the optimal trade-off between these two sets of requirements. In particular, we narrow down to two important factors for investigation: the output modality and the presentation of the text content to be visually presented on the smart glass screen. Note that the output modality only pertains to the display of the text content. The system status feedback for text change confirmations or error feedback are always delivered through auditory messages and is outside the scope of our study.

While the output modality can be purely audio or visual, or a combination of both, it needs exploration of the trade-offs that each modality would present while editing the text on the go. Also, for visual presentation of the text content, it is important to explore *how much* of the text should be presented so that it strikes the right balance between presenting enough context for editing the text and minimizing the visual/cognitive load of processing the presented content.

Research Questions and Hypotheses

We designed our study around three main questions —

Q1. What effect does the modality of output have on the users' text-correction and path-navigation abilities?
Previous research has highlighted the trade-off between audio and visual output for on-the-go *text reading* comprehension [59]. Audio was slower, but allowed better path navigation and presented a lower task-load. *Does the same trade-off between audio-only and visual-only output apply to on-the-go text editing tasks?* We hypothesize that while visual output would allow faster editing, audio would allow better path navigation, but also present increased task-load due to the difficulty in error detection and correction without visual output. Furthermore, *does combining audio and visual output allow faster corrections and better path-navigation?* We hypothesize that the redundancy of information in the bi-modal output would allow faster corrections but present more path-navigation difficulties and increase the task-load.

Q2. How does the text presentation affect users' text-correction and path-navigation abilities?
Seeing more text on screen can allow the user to form a higher-level understanding of the text, thus making it easier to process and edit the text, but more text might also cause more distraction, thereby increasing the path-navigation challenges. We expect increasing the amount of text displayed will lead to faster corrections as it would present more context of the text and reduce the number of navigation operations needed to scan through the text. However, less visual output would be easier for path-navigation.

Q3. How does path difficulty affect the user performance in different modalities and text presentations?
We hypothesize that showing more text on the display, while might be beneficial on a simple path, would lead to a decline in performance on a more difficult path as both the text and the path would demand higher visual attention. Audio-only mode might be the least affected by variations in path difficulty.

Q4. Is correction by re-speaking sufficient?
If the user utterance contains a repair-context that has repeated occurrences in the text then it might result in an unintended alignment. However, with our current design, the user may undo and re-attempt, including more context words in their next utterance to remove possibly any ambiguity in the alignment. We wanted to understand how easy or difficult it was for users to recover from a misalignment while balancing the cognitive load of navigating their path.

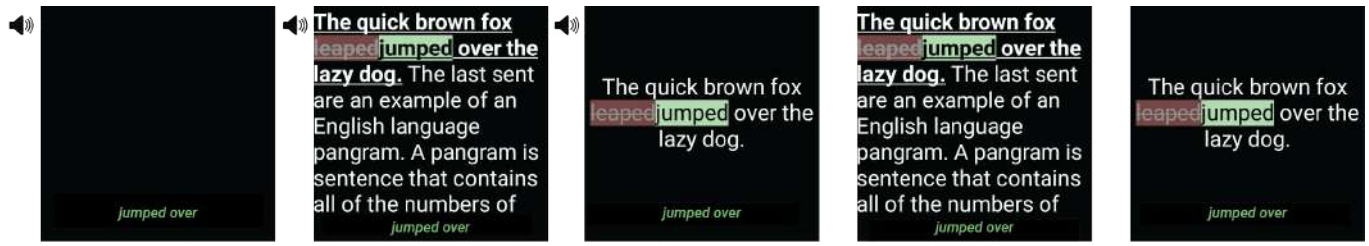
Modes of Output

To investigate our research questions, we designed 5 output modes, each exploring a combination of audio and/or visual modality. The visual output can be rendered in two different presentations: text-block and sentence-by-sentence (Table 1). A text-block in our experiment was defined as the maximum amount of text (≈33 words) that can be rendered on the *text-content space* of the Blade display. In block rendering, users can scroll through the text by swiping up/down on the hand-controller trackpad. Figure 3 shows all the output modes.

The audio-visual modes provide a reading-while-listening experience where both the modalities are active at the same time. For modes supporting audio output, playback of the text content is delivered through a Text-to-Speech (TTS) reader. To maintain consistency with the text presentation in the visual-only modes (V_s, V_b), the audio playback of the text is sentence-by-sentence in the AV_s mode and continuous in AV_b. Also, in our pilots, we noticed that editing text in an *audio-only* mode

	Audio	*Audio+Visual*	*Visual*
TEXT BLOCK	**invalid**	AV_b	V_b
SENTENCE	A	AV_s	V_s

Table 1: Output modes combining output modalities and presentations.



A: Aud | AV_b: Aud-Vis (*Block*) | AV_s: Aud-Vis (*Sentence*) | V_b: Vis (*Block*) | V_s: Vis (*Sentence*)

Figure 3: Screenshots of the Blade display showing the five output modes (Aud=Audio, Vis=Visual). Audio icons indicate support for audio-output. In green font is the transcribed user utterance.

was cognitively very challenging if a continuous stream of audio was presented. Hence, we precluded continuous audio playback in the *A* mode. The audio-output modes support barge-in interaction [17] i.e., when the user makes a change utterance (user interrupt), the TTS instantly pauses and the correction is done in real time.

The scope of the correction for modes with audio-support is always set as the interrupted sentence. Block rendering modes, AV_b and V_b, have a visual marker to indicate a sentence selection. For AV_b, the sentence being read by the TTS gets auto-selected, while for V_b, the user can select a sentence to mark the scope for correction. For both AV_b and V_b, the scope is the whole text, but prioritized by the selection, i.e., first the selected sentence is searched for a possible alignment, but if unsuccessful, rest of the text is searched.

Apparatus

The apparatus used was EYEditor as per its first design.

Participants

10 volunteers (6M, 4F, Mean Age=25.8 years, SD=3.79) took part in the study. None of the participants had prior experience of using a smart glass. All the participants had obtained at least one university degree taught in English.

Design and Procedure

A repeated-measures within-participant design was used. The independent variables were output mode *Mode* (A, AV_b, AV_s, V_b, V_s) and path-type *Path-type* (Simple, Stair). A fully crossed design resulted in 10 combinations of *Mode* and *Path-type* per participant.

Each participant performed the experiment in one session lasting approximately one hour. The session was blocked by output mode, with a participant walking on 2 path-types for each output mode. Presentation of the output modes were counterbalanced using Latin Square across all the participants. This resulted in 2 groups of 5 participants in each group. One group walked the simple path first, while the other group walked the stair-path first.

For each output mode, participants had to correct two paragraphs of text, one for each path-type. Each paragraph comprised of 5 simple sentences with each sentence having an average number of 8-9 words. The words were extracted from 3 different texts with their Flesch reading ease scores [26] fixed between 70-80. There were about 4 words per line (SD=.75) or ≈21 characters per line (Mean=20.82 characters, SD=2.48). Each sentence was embedded with two one-word errors (one each in the subject and the predicate) by adding/deleting/replacing correct words from the text. The errors served as correction opportunities for the participants. Only *semantic* (meaning) and *syntactic* (grammatical) errors [30] were embedded to ensure that the errors were identifiable without prior knowledge of the text.

We chose two different paths for the experiment, one for each path-type. Each path was 30 meters long. For the correction task, participants looped on the same path until they had completed exactly 5 of the 10 possible corrections, after which they were asked to "Stop". The first path condition was a simple straight path with no obstacles, while the second consisted of two flights of stairs. For the stair-path, participants alternated between first climbing down and then climbing up.

Before using each output mode, participants were briefed on the particulars of that mode after which they were given a single warm-up block to familiarize themselves with the mode. During the warm-up, participants corrected a sample piece of text while walking. At the start of the experiment, participants were given a short reading task on the Blade to familiarize themselves to reading text on a smart glass.

Participants filled out an unweighted NASA-TLX questionnaire [23] to report their subjective task load at the end of each output mode condition and another about their subjective preferences at the end of the experiment. We also interviewed the participants for 2 to 3 minutes at the end of the experiment.

Data Collection

We recorded 100 (= 5 *Mode* x 2 *Path-type* x 10 participants) measured trials in total. We measured task completion time, TCT, and stopping percentage, STP, defined as the percentage of TCT in which participants stopped walking during a task.

Results

Task Completion Time

A repeated measures analysis of variance (ANOVA) was performed on TCT $\sim$ *Mode* x *Path-type*. There was a significant main effect for *Mode* ($F_{4,36} = 261.80$, $p < .001$) and *Path-type* ($F_{1,9} = 26.63$, $p < .001$) on TCT. Post hoc multiple means comparison tests between the 5 output modes showed that V_s was significantly faster than others for both path-types, both

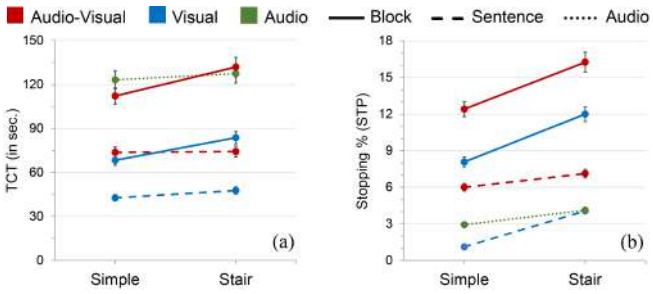


Figure 4: Comparative evaluation of the output modes.

at the p < .001 level. Text-editing was faster when visual output was available. Yet, block presentation reduced the performance steeply, more so when the higher visual demand was coupled with demand on the audio-channel (Figure 4a). Between the path-types, participants were significantly faster on the simple-path (84.02s ± 30.96) than the stair-path (92.99s ± 33.92).

There was also a significant interaction effect *Mode* x *Path-type* ($F_{4,36} = 4.07$, $p < .01$) on TCT. As mentioned earlier, V_s was unaffected by this interaction. Among the other paths, there was interaction between AV_s and V_b, and between AV_b and A. Each of these interactions showed that block-level visual performed faster (non-significant) than audio-supported modes on the simple path; on the difficult path audio-supported modes were faster. Furthermore, for modes with block-level visual, participants slowed down significantly on the stair path than on the simple path. Both AV_b-stair and V_b-stair were slower than AV_b-simple (p < .001) and V_b-simple (p < .05), respectively.

Stopping Percentage

We measured Stopping Percentage (STP) in our experiment to investigate how different mode and path combinations affect participants' natural walking. The assumption is that the more distraction an output mode induces, the more the participants stop and vice-versa. Hence, STP is an indicator of how well an output mode allows path-navigation.

A repeated measures ANOVA showed a significant main effect of *Mode* ($F_{4,36} = 112.099$, $p < .001$) and *Path-type* ($F_{1,9} = 48.102$, $p < .001$) on STP. Post-hoc analysis with Bonferroni corrections revealed that with V_s and A, participants stopped significantly less than with AV_s, V_b, and AV_b, all at the p < .001 level. Although participants stopped less in V_s than A for both path-types, there was no significant difference between the two modes. AV_b presented significantly more challenges than the other output modes, all at the p < .001 level. These results showed that simultaneous audio and visual output or visual output with higher visual load caused more disruption in the editing process while walking.

There was significant *Mode* x *Path-type* interaction ($F_{4,36} = 3.249$, $p < .05$) on STP. As Figure 4b shows, for all output modes, STP on the stair-path was higher than the simple-path. However, the difference was non-significant for A and AV_s. Thus, modes with audio-output were less affected by the change in path-type provided the visual load was low. If the visual load was high or there was no audio output, participants found the difficult path significantly more challenging as compared to the simple path.

Subjective Task Load

A repeated measures ANOVA showed a significant main effect of *Mode* on the overall unweighted NASA-TLX score ($F_{4,36} = 264.72$, $p < .001$). Post hoc multiple means comparison test showed that on the overall score, the sorted order of modes from lower to higher task-load was: $V_s < V_b$ (p < .001) < AV_s < A (p < .01) < AV_b (p < .001). Results for individual indices are given in the Appendix.

Subjective Preference and Feedback

In a post-study questionnaire, we asked participants to indicate their preferred output mode for each path-type. 100% and 80% of the participants chose V_s (visual-only, sentence-by-sentence) as their preferred mode for the simple and the difficult path, respectively. When interviewed, all 10 participants mentioned that V_s was easy to use, while AV_b was "difficult and frustrating". Also, there was a general consensus that reading text-blocks was more difficult than reading single sentences. Moreover, correcting text was difficult with audio-only output, but navigating paths, especially stairs, felt easier.

Discussion

Q1. What effect does the modality of output have on the users' text-correction and path-navigation abilities?

The results confirmed our hypothesis that overall, participants were faster and more comfortable with the editing task when they had visual output of the text. Participants corrected the text almost 3 times as fast with sentence-level output as compared to audio. Also, visual output had lower task load than audio. However, whether audio offered advantage over visual output for path-navigation depended on the text presentation—while audio was better than block text presentation, it had no significant advantage over sentence-by-sentence presentation.

Bi-modal output, as we expected, presented high cognitive load, but contrary to our hypothesis, was ≈54% slower than the visual-only modes. Hence, we recommend avoiding simultaneous bi-modal output for future designs exploring smartglass-based text editing.

Q2. How does the text presentation affect users' text-correction and path-navigation abilities?

The amount of text rendered on screen significantly affected the user performance. Hence, our hypothesis that more context of the text would result in faster correction did not hold, irrespective of the path conditions. However, as we expected, block-level output did present more path-challenges and increased task-load over sentence-level presentation.

Q3. How does path difficulty affect the user performance in different modalities and text presentations?

Among all the modes, V_s was the only mode that performed equally well on both path-types. In fact, V_s was the best mode for both editing and path-navigation, irrespective of the path difficulty. Moreover, modes that required less visual attention to the text (A, AV_s, V_s) were less susceptible to changes in path difficulty than modes that required more visual attention (AV_b, V_b), thus, confirming our hypothesis.

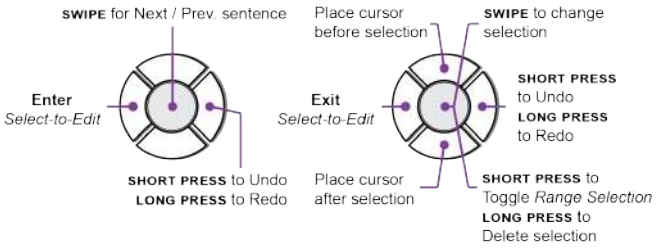


Figure 5: Hand-controller functions in the *Re-speaking Mode* (left) and the *Select-to-Edit Mode* (right).

Q4. Is correction by re-speaking sufficient?
We observed that if re-speaking resulted in a misalignment, it was difficult for first time users to strategize how much context to include in their next correction attempt. Sometimes, users made the exact same utterance on their second or third attempt despite failing to get the intended result in the previous attempt(s). Furthermore, some users, in their repeated attempts, re-spoke the whole sentence, thereby increasing the chances of a recognition error in some part of the utterance. Thus, setting the exact scope for the correction was desirable, but difficult when limited to only re-speaking based correction. This reflected in the participants' subjective feedback where 40% of the participants mentioned that if re-speaking resulted in a misalignment, they wanted finer control over the text so that they could precisely select the intended repair region.

IMPROVED DESIGN OF EYEditor

Based on our findings from Study 1, we included only V_s (visual-only with sentence-by-sentence rendering) as part of the updated design of EYEditor. The base framework used in EYEditor's final design inherited all the components directly from the V_s mode. Additionally, we added new functionalities to the hand-controller (Figure 5) to let the user have manual control over the text selection process. The updated system now operates in two modes, each allowing a different granularity of control over the text. The left controller button toggles between the two modes. The modes are (Figure 1):

Re-speaking Mode: By default, the system starts up in this mode. Users can change the text by re-speaking and has sentence-level control over the text, i.e., they can navigate between the sentences by swiping on the hand-controller.

Select-to-Edit Mode: This mode gives the user word-level control over the text. Swiping on the trackpad moves a yellow marker (indicates selection) over the words in the current sentence. While on a selection, pressing the trackpad button toggles between word-selection and range-selection. Selected text can be replaced with a voice-utterance. A selection can be deleted either by saying the DELETE keyword or by long-pressing the trackpad button. Also, pressing the top/down button places the cursor before/after a selection and allows the user to insert spoken content at the cursor location.

STUDY 2: COMPARISON WITH SMARTPHONE SOLUTION

While the smartphone (henceforth, just 'phone') demands full visual attention (*visual-exclusivity*) and is generally used heads-down, the smart glass (henceforth, just 'glass') allows the user to share their visual bandwidth between the digital screen and the path (*visual-flexibility*). In this study, our objective is to understand how the two platforms compare in handling the trade-offs between the users' text-editing and path-navigation needs for on-the-go text-editing tasks. To investigate the trade-offs in different situations, we consider two important factors: difficulty of the editing/correction task and path difficulty.

Research Questions and Hypotheses

Our exploration is based on three research questions —

Q1: How does each platform handle the visual/cognitive demands of editing on the go?
Visual-exclusivity and visual-flexibility present a trade-off between supporting the user's text-editing and path-navigation needs. While the phone's visual-exclusivity might be necessary for difficult correction tasks as it channels the users' full attention to the task, it might not be suitable for difficult paths. Similarly, the glass's visual-flexibility might be useful for simple tasks but it is unclear if it would have any added advantage over the phone and how it would affect the users' editing and path-navigation abilities for more difficult paths/tasks.

With users' prior experience in phone-based text editing, we expect that on simple paths demanding less visual attention, the phone will outperform the glass. Conversely, on difficult paths, the glass will allow better focus on the path, but with lower correction efficiency.

Q2: What role does posture play in the usability of each platform on various path-types?
Heads-down text-editing is not ideal as it diverts users' attention from the path. Yet, smartphone users frequently have their heads down while typing on the phone. Thus, we want to explore if EYEditor will perform better when a heads-up posture is required, i.e., in more visually challenging conditions.

We hypothesize that heads-down editing on the phone would be faster on simple paths due to minimal visual attention required on the path, while EYEditor would be more optimal on difficult paths for simple corrections. Yet, it is unclear how our system would compare to the phone when making difficult corrections on difficult paths.

Q3: Is our solution viable for future considerations?
So that our solution can inspire future designs, our criterion for viability is that EYEditor should offer additional benefits over using phones, especially on more challenging paths.

Study Design

We controlled the difficulty of the experimental conditions by varying the path complexity and the complexity of the correction tasks. Table 2 lists the independent variables and their levels. A repeated measures design with 2 *Technique* (Glass, Phone) x 3 *Path-type* (SimPath, ObstPath, StairPath) x 2 *Task-Complexity* (Easy, Hard) resulted in 12 conditions per participant. The experiment was performed in one session lasting approximately 90 to 100 minutes.

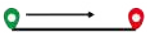
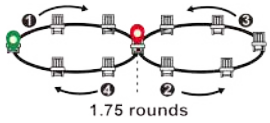
Independent Variables			Measured Variables
TECHNIQUE	PATH-TYPE	TASK-COMPLEXITY	
Glass Text was pushed to EYEditor's interface. **Phone** Participants received the text over WhatsApp and copied it to a notes application on their phone for editing. On task completion, participants sent the edited text back to the experimenter. Sending/receiving the text was not timed.	**SimPath** Straight path with no obstacles.  **ObstPath:** Curved path in the shape of figure 8. 11 obstacles (chairs) were placed symmetrically around the figure 8.  **StairPath:** 4 floors of stair-down path with 2 flights/floor. For safety, short flights were selected (10 steps/flight). Stair had hand-railings and anti-slip grooves on each step. 	**Easy** • 8 simple sentences • Avg. of 6-7 words/sentence • Flesch reading ease score between 80-90 • Two 1-word errors **Hard** • 8 sentences — realistic mix of simple, compound and complex sentences • Avg. of 11-12 words/sentence • Flesch reading ease score between 60-70 • Two 2-word errors	1) Preferred Walking Speed (PWS). 2) Task Completion Time (TCT): time (in sec.) to walk the 50m path while correcting the text. 3) Percentage of Preferred Walking Speed (PPWS) = (50m/TCT)/PWS*100. 4) Number of Corrections (#Corrections) that participants could achieve during the 50m path (reported in the appendix). 5) Corrections per second (CPS) = #Corrections/TCT. 6) Users' subjective task-load and preference for technique.

Table 2: Study 2 Design Table. The rows under an independent variable column indicate its levels. All paths measure 50m from start (green pin) to finish (red pin). For ObstPath, one round measures ≈28.5m, taking 1.75 rounds to complete the path.

We designed the paths to simulate realistic paths and obstacles encountered in on-the-go scenarios. For the correction tasks, each text comprised of 8 logically connected sentences on a given topic (selected from a diverse range of general topics) and were embedded with errors that served as correction opportunities for the participants. To avoid potential bias due to the error processing depth, the errors were randomly distributed between the first and the last word, under the following constraints: simple sentences had one error each in the subject and the predicate, compound sentences had one error in each of the two constituent simple sentences and complex sentences had one error each in the dependent and the independent clause. As in Study 1, the errors were syntactic or semantic in nature. The *Task-Complexity* levels simulate realistic general purpose correction scenarios within a body of text.

Participants

12 volunteers (6M, 6F) aged between 18 to 36 years (mean age = 24.5 years, SD = 4.78) were recruited for the study. An equal number (n=6) of native and non-native English speakers were recruited to minimize any potential bias due to speech recognition accuracy. None of the participants had prior experience of using a smart glass but all of them had been regular smartphone users for at least the past 5 years. Each participant received an equivalent of ≈11 USD as compensation for their participation. No participants from Study 1 or any of our pilot studies were repeated in this study.

Apparatus

For the glass technique, the improved EYEditor was used. For the phone technique, we let participants use their own mobile phones to allow maximum familiarity of the device. Also, we did not constrain the participants' mobile typing experience to allow for a realistic comparison of our proposed technique with the existing technique. Hence, participants could edit the text on their preferred note-taking application and were free to use any existing typing/correction aids such as auto-correct, auto-complete, swipe typing, voice-input, etc. In keeping with preserving platform familiarity, text on the mobile phone was presented as a single paragraph.

Procedure

Our study was conducted in indoor lighting conditions to ensure maximum text visibility on the smart glass. The experiment began with a briefing of the tasks. Then participants walked a 20 meter segment of each of the 3 paths twice, at their normal walking speed. The two trials were averaged to compute each participant's Preferred Walking Speed (PWS) on each path.

The glass block started with a reading exercise where the participants familiarized themselves to read text on the glass screen. This was followed by a training session and a single warm-up session for practice. The reading, training and practice sessions, combined, lasted between 20 to 25 minutes. The phone block was preceded by a warm-up session where participants corrected a sample text on their phone while walking. For both the techniques, we instructed the participants to correct as many errors as possible while walking the path at their comfortable walking speed. The entire session was video recorded for further analysis.

After each block, participants filled out an unweighted NASA-TLX questionnaire to report their subjective task load after each technique block. At the end of the study, they completed a subjective preference questionnaire and were then interviewed for about 5 to 7 minutes.

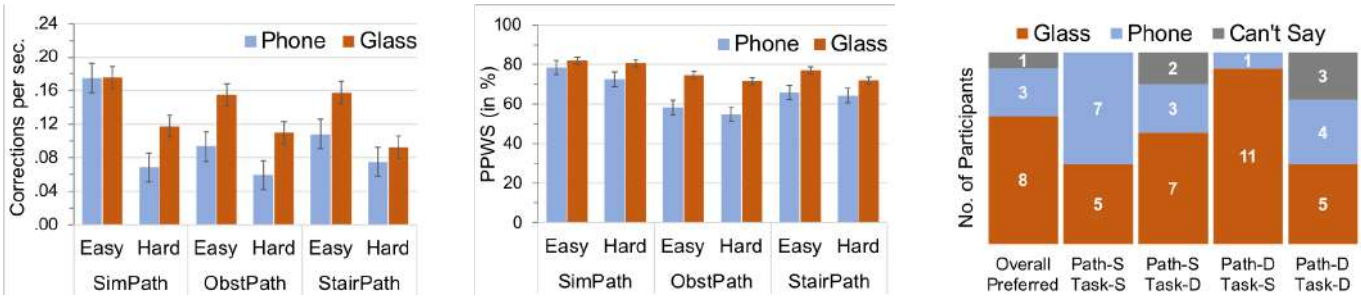
Data collection

We collected 144 (= 2 *Technique* x 3 *Path-type* x 2 *Task-Complexity* x 12 participants) measured trials in total. The measured variables are listed in Table 2.

Results

Corrections per second

A repeated measures analysis of variance was performed on CPS ∼ *Technique* x *Path-type* x *Task-Complexity*. There was



(a) (b) (c)

Figure 6: (a)-(b) Measured outcomes comparison. (c) Preferred Technique for 12 participants (S=Simple, D=Difficult).

a significant main effect for *Technique* ($F_{1,11} = 14.87$, $p < .01$), *Path-type* ($F_{2,22} = 39.65$, $p < .001$), and *Task-Complexity* ($F_{1,11} = 130.75$, $p < .001$). Participants' correction speed (in CPS) with the glass ($.135 \pm .05$) was overall faster than with the phone ($.097 \pm .046$) (p<.01). Furthermore, the correction speed was significantly lower on ObstPath (p<.001) and StairPath (p<.001) than on SimPath; however, there was no significant difference between ObstPath and StairPath (Figure 6a).

Furthermore, there was a significant *Technique* x *Path-type* ($F_{2,22} = 15.31$, $p < .001$), *Path-type* x *Task-Complexity* ($F_{2,22} = 20.55$, $p < .001$), and *Technique* x *Path-type* x *Task-Complexity* ($F_{2,22} = 17.583$, $p < .001$) interaction effect on CPS. Post hoc multiple means comparison tests with Bonferroni correction revealed that overall, glass and phone performed similarly on both SimPath and StairPath; yet, glass significantly outperformed phone on ObstPath (p<.01). On SimPath, while there was no significant difference between the glass ($.176 \pm .058$) and the phone ($.175 \pm .04$) for easy tasks, for hard tasks, glass ($.117 \pm .026$) performed significantly better (p<.05) than phone ($.068 \pm .016$). On ObstPath, glass was significantly faster for both easy (p<.01) and hard tasks (p<.05). On StairPath, glass ($.158 \pm .049$) outperformed phone ($.108 \pm .027$) (p<.05) for easy tasks, while for hard tasks, there was no significant difference between the two techniques.

Percentage of Preferred Walking Speed

PPWS can be interpreted as: the lower its value, the slower the participant is walking compared to their normal walking speed [14, 45]. Thus, PPWS quantifies the effect that a device used on the path had on the user's ability to focus on the path. We performed a repeated measures ANOVA on PPWS $\sim$ *Technique* x *Path-type* x *Task-Complexity*. There was a significant main effect of *Technique* ($F_{1,11} = 10.776$, $p < .01$), *Path-type* ($F_{2,22} = 5.14$, $p < .05$), and *Task-Complexity* ($F_{1,11} = 21.958$, $p < .001$) on PPWS. PPWS with the phone (65.76 ± 15.12) was significantly lower than the with the glass (76.33 ± 16.43) (p<.01). Also, unsurprisingly, there was a significant decrease in walking speed while making difficult corrections than simple ones (p<.001). Furthermore, there was a significant *Technique* x *Path-type* ($F_{2,22} = 7.825$, $p < .01$), and *Technique* x *Path-type* x *Task-Complexity* ($F_{2,22} = 5.012$, $p < .05$) interaction effect on PPWS (Figure 6b).

Post hoc comparisons with Bonferroni corrections showed that the glass (73.14 ± 11.43) allowed participants to maintain a significantly higher (p<.01) PPWS as compared to the phone (56.6 ± 10.21) on ObstPath. The higher PPWS of glass on ObstPath was true for both easy and hard tasks (both at the p < .05 level). For the other paths and task-difficulties, there was no significant difference between the glass and the phone.

Subjective Results

A paired sample t-test was conducted on the NASA-TLX scores for the glass and the phone to compare their subjective task loads. In the overall unweighted score, the task load with the glass (37.92 ± 12.21) was significantly lower (p<.01) than with the phone (56.58 ± 11.34). Results for individual indices are given in the Appendix.

In the post-study questionnaire, $\approx$92% of participants reported that smart glasses can be a viable alternative to the phone for on-the-go text-editing. In another question, they indicated their preferred technique for various task-path combinations. The results are reported in Figure 6c.

Discussion

Q1. How does each platform handle the visual/cognitive demands of editing on the go?

Overall, participants could both walk and correct the text faster with the glass than with the phone. The glass also presented a lower task-load. The results were surprising given that all our participants were first time users of the smart glass. In particular, the glass had significant advantage over the phone when either the path or the task presented a high cognitive load. Hence, our hypothesis that the glass will allow participants to walk faster on difficult paths was validated; however, our hypothesis that phone's correction efficiency on simple and difficult paths would exceed that of the glass's did not hold.

Moreover, there was a general consensus among participants that alternating attention between the text on the glass screen and the visual surrounding felt much easier and more seamless as compared to the phone. The availability of peripheral vision was key to this outcome and was particularly useful to have when visual attention was needed both on the task AND on the path. Yet, for difficult paths, participants had mixed feedback about the benefits of the glass's visual-flexibility. 25% of the participants believed that the glass's flexibility can instill a false sense of security while in effect drawing their attention away from hazards; however, the other 75% agreed that the glass would be more suitable for navigating difficult paths.

Q2. What role does posture play in the usability of each platform on various path-types?
Confirming our hypothesis, the glass did perform better in visually challenging conditions when a heads-up posture was required to navigate path challenges. Yet, although there was no significant difference between ObstPath and StairPath in terms of participants' average performance measures, the glass outperformed the phone on ObstPath but not on StairPath. From analysis of the participants' video logs and interview data, it was revealed that the path challenges of StairPath were more easily detected heads-down. Since, participants were correcting on the phone heads-down, it was "just a matter of glancing sideways" to be sure of their footing on the stairs, whereas with the glass, the participants had to shift posture from heads-up to heads-down, which created some discomfort and delay. However, despite the path-navigation advantage of using the phone heads-down on the stairs, performance of the glass was on par with the phone.

That even on the simple path participants performed better with the glass when the task was difficult (high visual load), proved that heads-up posture was better for visually challenging situations. Furthermore, this finding implied that for visually intensive tasks, if the participants lost view of their surrounding (as with the phone), they slowed down even if they had prior knowledge that the path was free of obstacles.

Q3. Is our solution viable for future considerations?
EYEditor indeed satisfied our criterion for viability as it offered significant task performance and path-navigation benefits over the phone for visually challenging conditions. In addition, generally, participants felt comfortable with the system. 75% of the participants mentioned that the learning curve for using our system felt short and they could easily and quickly adapt to the system. A key component of the acceptance came from the ability to correct by re-speaking. While the number of corrections done by re-speaking was about 3.5 times the number using the Select-to-Edit mode, the time spent on re-speaking was only 1.6 times of that spent in Select-to-Edit. Thus, using re-speaking was easier and faster. In general, Select-to-Edit was used as a fall-back when re-speaking failed due to either limitation of the algorithm or speech recognition errors. 42% of the participants likened Select-to-Edit to phone-based editing, while 75% preferred re-speaking to even typing on the phone. Hence, there was general consensus that making complex corrections on the glass was easier than on the phone because of the ability to correct by re-speaking on the glass. On the other hand, 25% of the participants had an ongoing preference for the phone. They reported that they were more confident with the phone as they were familiar with it.

Based on our results and user feedback, we believe that the smartglass-based display and the support for respeaking-based correction were the key contributors to the effectiveness of our approach. Yet, the other design considerations enhance the viability of our solution—while our proposed content presentation style allows optimal utilization of the display, the Select-to-Edit mode allows the user a finer-grained control over the editing process, and the manual input is efficient for text navigation and selection.

Overall, Study 2 shows that our smart glass solution, EYEditor, offered certain advantages over the phone and helped maintain better path awareness. Hence, EYEditor might potentially be safer to use on the go. Yet, we found there is a cognitive bottleneck when both the editing and navigation were more challenging, where the advantages of EYEditor becomes less salient especially while walking down stairs.

Limitations and Future Work

Although we tested on-the-go scenarios with realistic path challenges, we could not extend the study outdoor to preserve maximum text visibility on the smart glass display. How ambient light and ambient noise of outdoor conditions would affect our system's performance remains to be seen. Also, voice interaction can sometimes be undesirable in public for security/social reasons. An in-depth study of safety and social factors warrants further investigation in future work. One potential design consideration to make the user experience safer might be to interleave audio output with visual output. As Study 1 had shown that while audio-only was difficult to use for editing text, it did offer path-navigation benefits.

Furthermore, our system can benefit from a more intelligent and adaptive re-speaking algorithm with language analysis and predictive text. Finally, our speak-to-edit mode functionality can benefit from more intuitive ways to invoke/exit the mode (or, through modeless operation) and support for quicker, non-sequential text selection, for example, by dividing the text into zones, or by allowing both horizontal and vertical selection as with mouse/touch-based input.

CONCLUSION

We presented EYEditor, a novel heads-up, smartglass-based text-editing solution optimized for on-the-go use cases with a combination of voice and manual input. An iterative design process with two controlled experiments gained us the following insights as take-away messages: 1) text content should be presented visually, sentence-by-sentence to optimize users' text-correction and path-navigation capabilities on the go; 2) overlapping audio and visual output (reading-while-listening) or overloading the screen space with text is highly distracting for on-the-go text-editing and should be avoided; 3) our hybrid solution supports mobility while text-editing better than typing on a smartphone for more complex paths/tasks, until users' attention span reaches a limit. In conclusion, our paper takes a significant step forward in understanding how to design heads-up interactions for on-the-go text-editing. This is also our first step to establish the feasibility, desirability and viability of using smart glasses as an interactive platform in on-the-go scenarios.

ACKNOWLEDGMENTS

This work was supported in part by the NUS School of Computing Strategic Initiatives. We thank Yang Chen for her generous help with designing some of the figures in the paper and Divyanshu Mandowara for reviewing a few related works.

REFERENCES

[1] Urs-Vito Albrecht, Ute von Jan, Joachim Kuebler, Christoph Zoeller, Martin Lacher, Oliver J Muensterer,

Max Ettinger, Michael Klintschar, and Lars Hagemeier. 2014. Google Glass for documentation of medical findings: evaluation in forensic medicine. *Journal of medical Internet research* 16, 2 (2014), e53.

[2] Ouais Alsharif, Tom Ouyang, Françoise Beaufays, Shumin Zhai, Thomas Breuel, and Johan Schalkwyk. 2015. Long short term memory neural network for keyboard gesture decoding. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2076–2080.

[3] Ahmed Sabbir Arif and Ali Mazalek. 2016. A survey of text entry techniques for smartwatches. In *International Conference on Human-Computer Interaction*. Springer, 255–267.

[4] Shiri Azenkot and Nicole B Lee. 2013. Exploring the use of speech input by blind people on mobile devices. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 11.

[5] Ronald T Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* 6, 4 (1997), 355–385.

[6] Bartosz Bajer, I Scott MacKenzie, and Melanie Baljko. 2012. Huffman base-4 text entry glove (H4 TEG). In *2012 16th International Symposium on Wearable Computers*. IEEE, 41–47.

[7] Mafkereseb Kassahun Bekele, Roberto Pierdicca, Emanuele Frontoni, Eva Savina Malinverni, and James Gain. 2018. A Survey of Augmented, Virtual, and Mixed Reality for Cultural Heritage. *Journal on Computing and Cultural Heritage* 11, 2 (March 2018), 1–36. DOI:http://dx.doi.org/10.1145/3145534

[8] Mathilde M Bekker, Floris L van Nes, and James F Juola. 1995. A comparison of mouse and speech input control of a text-annotation system. *Behaviour & Information Technology* 14, 1 (1995), 14–22.

[9] Doug A Bowman, Vinh Q Ly, and Joshua M Campbell. 2001. Pinch keyboard: Natural text input for immersive virtual environments. (2001).

[10] Statistic Brain. 2017. Text Message Statistics — United States. (2017). Retrieved September 14, 2019 from https://www.statisticbrain.com/text-message-statistics/.

[11] Kenneth Burke. 2018. How Many Texts Do People Send Every Day (2018)? (Nov 2018). Retrieved September 14, 2019 from https://www.textrequest.com/blog/how-many-texts-people-send-per-day/.

[12] Stuart K Card, Thomas P Moran, and Allen Newell. 1980. Computer text-editing: An information-processing analysis of a routine cognitive skill. *Cognitive psychology* 12, 1 (1980), 32–74.

[13] Junhwi Choi, Kyungduk Kim, Sungjin Lee, Seokhwan Kim, Donghyeon Lee, Injae Lee, and Gary Geunbae Lee. 2012. Seamless error correction interface for voice word processor. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4973–4976.

[14] DD Clark-Carter, AD Heyes, and CI Howarth. 1986. The efficiency and walking speed of visually impaired people. *Ergonomics* 29, 6 (1986), 779–789.

[15] Brian Lystgaard Due. 2014. The future of smart glasses: An essay about challenges and possibilities with smart glasses. *Working papers on interaction and communication* 1, 2 (2014), 1–21.

[16] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. 1997. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal Technologies* 1, 4 (01 Dec 1997), 208–217. DOI: http://dx.doi.org/10.1007/BF01682023

[17] Debjyoti Ghosh, Pin Sym Foong, Shengdong Zhao, Di Chen, and Morten Fjeld. 2018. EDITalk: towards designing eyes-free interactions for mobile word processing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 403.

[18] Mitchell Gordon, Tom Ouyang, and Shumin Zhai. 2016. WatchWriter: Tap and gesture typing on a smartwatch miniature keyboard with statistical decoding. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3817–3821.

[19] Tovi Grossman, Xiang Anthony Chen, and George Fitzmaurice. 2015. Typing on glasses: Adapting text entry to smart eyewear. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 144–152.

[20] Aakar Gupta and Ravin Balakrishnan. 2016. DualKey: miniature screen text entry via finger identification. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 59–70.

[21] Ewa Gustafsson, Sara Thomée, Anna Grimby-Ekman, and Mats Hagberg. 2017. Texting on mobile phones and musculoskeletal disorders in young adults: a five-year cohort study. *Applied ergonomics* 58 (2017), 208–214.

[22] Christine A Halverson, Daniel B Horn, Clare-Marie Karat, and John Karat. 1999. The beauty of errors: Patterns of error correction in desktop speech systems.. In *INTERACT*. 133–140.

[23] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[24] Jonggi Hong, Seongkook Heo, Poika Isokoski, and Geehyuk Lee. 2015. SplitBoard: A simple split soft keyboard for wristwatch-sized touch screens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1233–1236.

[25] David M Kietrys, Michael J Gerg, Jonathan Dropkin, and Judith E Gold. 2015. Mobile input device type, texting style and screen size influence upper extremity and trapezius muscle activity, and cervical posture while texting. *Applied ergonomics* 50 (2015), 98–104.

[26] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. (1975).

[27] 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*. ACM, 241–244.

[28] Jochen Kuhn, Paul Lukowicz, Michael Hirth, Andreas Poxrucker, Jens Weppner, and Junaid Younas. 2016. gPhysics—Using smart glasses for head-centered, context-aware learning in physics experiments. *IEEE Transactions on Learning Technologies* 9, 4 (2016), 304–317.

[29] 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*. ACM, 1952–1956.

[30] Pascale Larigauderie, Daniel Gaonac'h, and Natasha Lacroix. 1998. Working memory and error detection in texts: What are the roles of the central executive and the phonological loop? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition* 12, 5 (1998), 505–527.

[31] Jae Yeol Lee and Guewon Rhee. 2008. Context-aware 3D visualization and collaboration services for ubiquitous cars using augmented reality. *The International Journal of Advanced Manufacturing Technology* 37, 5 (01 May 2008), 431–442. DOI: http://dx.doi.org/10.1007/s00170-007-0996-x

[32] Min Lin, Rich Goldman, Kathleen J Price, Andrew Sears, and Julie Jacko. 2007. How do people tap when walking? An empirical investigation of nomadic data entry. *International journal of human-computer studies* 65, 9 (2007), 759–769.

[33] Ming-I Brandon Lin and Yu-Ping Huang. 2017. The impact of walking while using a smartphone on pedestrians' awareness of roadside events. *Accident Analysis & Prevention* 101 (2017), 87–96.

[34] Bob W Lord and Ray Velez. 2013. *Converge: transforming business at the intersection of marketing and technology*. John Wiley & Sons.

[35] Aliaksandr Malokin, Giovanni Circella, and Patricia L Mokhtarian. 2019. How do activities conducted while commuting influence mode choice? Using revealed preference models to inform public transportation advantage and autonomous vehicle scenarios. *Transportation Research Part A: Policy and Practice* 124 (2019), 82–114.

[36] Roderick McCall, Benoît Martin, Andrei Popleteev, Nicolas Louveton, and Thomas Engel. 2015. Text entry on smart glasses. In *2015 8th International Conference on Human System Interaction (HSI)*. IEEE, 195–200.

[37] Arthur E McNair and Alex Waibel. 1994. Improving recognizer acceptance through robust, natural speech repair. In *Third International Conference on Spoken Language Processing*.

[38] Paul Milgram. 1994. Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems* E77-D, 12 (1994), 1321–1329.

[39] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[40] Stefan Mitrasinovic, Elvis Camacho, Nirali Trivedi, Julia Logan, Colson Campbell, Robert Zilinyi, Bryan Lieber, Eliza Bruce, Blake Taylor, David Martineau, and others. 2015. Clinical and surgical applications of smart glasses. *Technology and Health Care* 23, 4 (2015), 381–401.

[41] Hugo Nicolau and Joaquim Jorge. 2012. Touch typing using thumbs: understanding the effect of mobility and hand posture. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2683–2686.

[42] Sharon Oviatt. 2000. Taming recognition errors with a multimodal interface. *Commun. ACM* 43, 9 (Sept. 2000), 45–51. DOI:http://dx.doi.org/10.1145/348941.348979

[43] Sharon Oviatt and Robert VanGent. 1996. Error resolution during multimodal human-computer interaction. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, Vol. 1. IEEE, 204–207.

[44] Physiopedia. 2019. Text Neck. (2019). Retrieved September 18, 2019 from https://www.physio-pedia.com/Text_Neck.

[45] Antti Pirhonen, Stephen Brewster, Stephen Brewster, and Christopher Holguin. 2002. Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 291–298.

[46] Philipp A Rauschnabel, Alexander Brem, and Young Ro. 2015. Augmented reality smart glasses: definition, conceptual insights, and managerial importance. *Unpublished Working Paper, The University of Michigan-Dearborn, College of Business* (2015).

[47] Sherry Ruan, Jacob O Wobbrock, Kenny Liou, Andrew Ng, and James Landay. 2016. Speech is 3x faster than typing for english and mandarin text entry on mobile devices. *arXiv preprint arXiv:1608.07323* (2016).

[48] Rufat Rzayev, Paweł W Woźniak, Tilman Dingler, and Niels Henze. 2018. Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 45.

[49] Sayan Sarcar, Prateek Panwar, and Tuhin Chakraborty. 2013. EyeK: an efficient dwell-free eye gaze-based text entry system. In *Proceedings of the 11th asia pacific conference on computer human interaction*. ACM, 215–220.

[50] Alexander Schick, Daniel Morlock, Christoph Amma, Tanja Schultz, and Rainer Stiefelhagen. 2012. Vision-based handwriting recognition for unrestricted text input in mid-air. In *Proceedings of the 14th ACM international conference on Multimodal interaction*. ACM, 217–220.

[51] Andrew Sears, Jinhuan Feng, Kwesi Oseitutu, and Claire-Marie Karat. 2003. Hands-free, speech-based navigation during dictation: difficulties, consequences, and solutions. *Human-computer interaction* 18, 3 (2003), 229–257.

[52] Mark Sharp. 2015. Beware the smartphone zombies blindly wandering around Hong Kong. (2 March 2015). Retrieved September 14, 2019 from https://www.scmp.com/lifestyle/technology/article/1725001/smartphone-zombies-are-putting-your-life-and-theirs-danger.

[53] Matthias Sperber, Graham Neubig, Christian Fügen, Satoshi Nakamura, and Alex Waibel. 2013. Efficient speech transcription through respeaking.. In *Interspeech*. 1087–1091.

[54] 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*. ACM, 3643–3652.

[55] Bernhard Suhm, Brad Myers, and Alex Waibel. 2001. Multimodal error correction for speech user interfaces. *ACM transactions on computer-human interaction (TOCHI)* 8, 1 (2001), 60–98.

[56] Husain Sumra. 2019. The best augmented reality glasses 2019: Snap, Vuzix, Microsoft, North & more. (Mar 2019). Retrieved September 14, 2019 from https://www.wareable.com/ar/the-best-smartglasses-google-glass-and-the-rest.

[57] B. Thomas, B. Close, J. Donoghue, J. Squires, P. De Bondi, M. Morris, and W. Piekarski. 2000. ARQuake: an outdoor/indoor augmented reality first person application. In *Digest of Papers. Fourth International Symposium on Wearable Computers*. 139–146. DOI: http://dx.doi.org/10.1109/ISWC.2000.888480

[58] B. Thomas, V. Demczuk, W. Piekarski, D. Hepworth, and B. Gunther. 1998. A wearable computer system with augmented reality to support terrestrial navigation. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*. 168–171. DOI: http://dx.doi.org/10.1109/ISWC.1998.729549

[59] Kristin Vadas, Nirmal Patel, Kent Lyons, Thad Starner, and Julie Jacko. 2006. Reading on-the-go: a comparison of audio and hand-held displays. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*. ACM, 219–226.

[60] Keith Vertanen and Per Ola Kristensson. 2009. Automatic selection of recognition errors by respeaking the intended text. In *2009 IEEE Workshop on Automatic Speech Recognition & Understanding*. IEEE, 130–135.

[61] 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 *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 153–160.

[62] X. Wang, S. K. Ong, and A. Y. C. Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (March 2016), 1–22. DOI:http://dx.doi.org/10.1007/s40436-015-0131-4

[63] MDN web docs. 2019. Web Speech API. (2019). Retrieved August 29, 2019 from https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API.

[64] 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 113.

[65] 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*. ACM, 4479–4488.

[66] 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 *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 71–82.