GlassMail: Towards Personalised Wearable Assistant for On-the-Go Email Creation on Smart Glasses

Chen Zhou
Synteraction lab
National University of Singapore
Singapore, Singapore
zhouchen@comp.nus.edu.sg

Zihan Yan
University of Illinois
Urbana-Champaign
Champaign, Illinois, United States
zihan25@illinois.edu

Ashwin Ram
Synteraction lab
National University of Singapore
Singapore, Singapore
ashwinram10@gmail.com

Yu Gu
Synteraction lab
National University of Singapore
Singapore, Singapore
e1314607@u.nus.edu

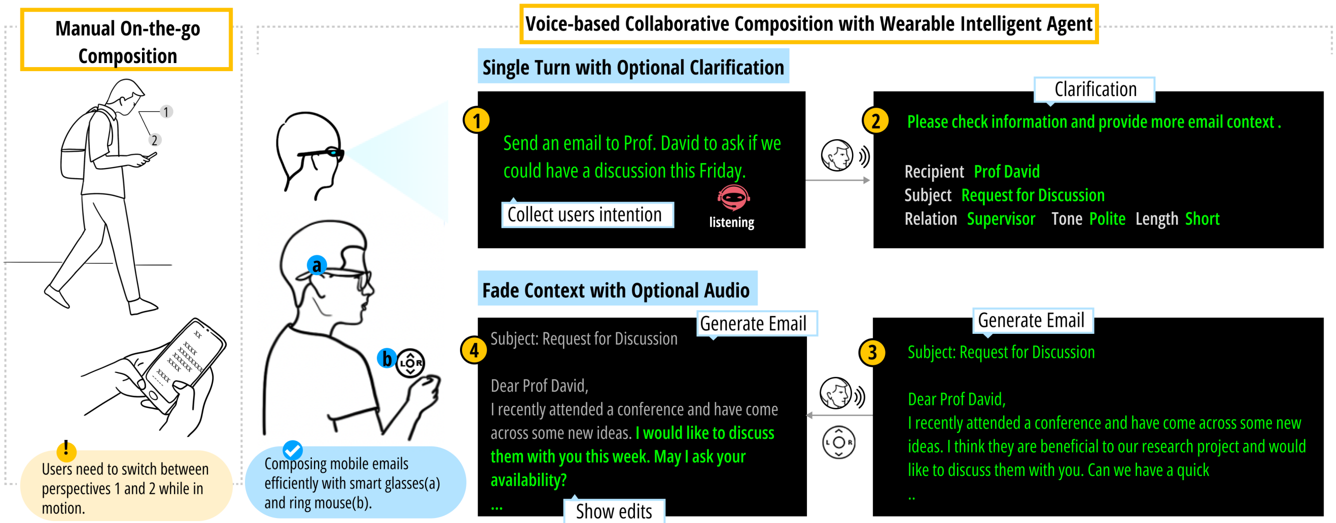
Yan Xiang
Synteraction lab
National University of Singapore
Singapore, Singapore
yanxiang.sjtu@gmail.com

Can Liu
School of Creative Media
City University of Hong Kong
Hong Kong, China
canliu@cityu.edu.hk

Yun Huang
School of Information Sciences
University of Illinois
Urbana-Champaign
Champaign, Illinois, United States
yunhuang@illinois.edu

Ooi Wei Tsang
Department of Computer Science
National University of Singapore
Singapore, Singapore
ooiwt@comp.nus.edu.sg

Shengdong Zhao*
School of Creative Media
City University of Hong Kong
Hong Kong, China
shengdong.zhao@cityu.edu.hk


Manual On-the-go Composition
1
2
Users need to switch between perspectives 1 and 2 while in motion.
Voice-based Collaborative Composition with Wearable Intelligent Agent
a
b
Composing mobile emails efficiently with smart glasses(a) and ring mouse(b).
Single Turn with Optional Clarification
1
Collect users intention
listening
2
Clarification
Fade Context with Optional Audio
4
Generate Email
Show edits
3
Generate Email

1 Send an email to Prof. David to ask if we could have a discussion this Friday.

2 Please check information and provide more email context .

Recipient Prof David
Subject Request for Discussion
Relation Supervisor Tone Polite Length Short

3 Subject: Request for Discussion

Dear Prof David,
I recently attended a conference and have come across some new ideas. I think they are beneficial to our research project and would like to discuss them with you. Can we have a quick
..

4 Subject: Request for Discussion

Dear Prof David,
I recently attended a conference and have come across some new ideas. I would like to discuss them with you this week. May I ask your availability?
...

Figure 1: *GlassMail* interactions: 1) The user initiates voice-based communication with *GlassMail* and simply expresses their email requirements in a single turn. 2) *GlassMail* then displays its understanding of the email with word-level chunking and provides the user with optional opportunities for clarification. 3) Once the user confirms, *GlassMail* generates an email. During the post-editing, 4) *GlassMail* utilizes a fading context with an optional audio output mode for efficiently displaying edits in mobile multitasking settings.

ABSTRACT

Optical See-through Head-Mounted Displays (OHMDs) offer new opportunities for completing complex information processing tasks on the go. We introduce *GlassMail*, a Large Language Models

*Corresponding author

DIS '24, July 01–05, 2024, IT University of Copenhagen, Denmark

© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0583-0/24/07
https://doi.org/10.1145/3643834.3660683

(LLMs)-based wearable assistant on OHMDs for mobile email creation. Our formative study identified two challenges of the LLM-based wearable email assistant: (i) achieving efficient and accurate understanding of user intentions, and (ii) ensuring effective information presentation for email processes. Through two empirical studies, we developed a "Single Turn with Optional Clarification" approach for accurate user intention recognition and a "Fade Context with Optional Audio" mode for effective email processing. An observation study then evaluated *GlassMail*'s feasibility in composing formal and semi-formal emails, supporting the usefulness and effectiveness of *GlassMail* in simple scenarios and yielding insights into potential future improvements for complex scenarios. We further discuss the design implications for the future development of wearable AI-enabled assistants.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**; **Empirical studies in interaction design**.

KEYWORDS

Optical See-Through Head-Mounted Displays, Smart Glasses, Large Language Model, Wearable LLM-based assistant, Complex Information Processing, Voice Assistant, Mobile Email Creation, Heads-Up Computing

ACM Reference Format:
Chen Zhou, Zihan Yan, Ashwin Ram, Yu Gu, Yan Xiang, Can Liu, Yun Huang, Ooi Wei Tsang, and Shengdong Zhao. 2024. GlassMail: Towards Personalised Wearable Assistant for On-the-Go Email Creation on Smart Glasses. In *Designing Interactive Systems Conference (DIS '24), July 01–05, 2024, IT University of Copenhagen, Denmark.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3643834.3660683

1 INTRODUCTION

A skilled human assistant can proficiently transform high-level verbal instructions received over a phone call, while the individual is commuting home, into a finely crafted message that accurately reflects the individual's intent and writing style. However, in today's context, individuals who don't have the luxury of a personal assistant, and attempt to compose such a message on their mobile phones while commuting often face challenges due to various constraints, such as limited screen sizes and restricted interaction capabilities. Consequently, using phones to compose a relatively complex email in such situations can become a laborious, error-prone, and unsatisfying experience [34], while also potentially compromising safety [25, 49].

The emergence of Large Language Models (LLMs) has revolutionized complex information processing, rendering it remarkably simpler. Users can now succinctly express their intention in a text field within a chatbot, and in return, receive a complete email draft [44]. This new capability opens up the possibility of having a digital version of a well-trained human assistant integrated into a wearable device, such as Optical See-through Head-Mounted Displays (OST-HMDs, OHMDs), to assist users in composing emails while on the go. The choice of wearable smart glasses over mobile phones stems from the desire to minimize the interference of interacting with the assistant with the user's ongoing activities. Mobile phone usage often demands excessive attention and leads to a "heads-down" posture [3], while OHMDs have proven to be superior for multitasking while maintaining environmental awareness [29, 51, 76].

However, current LLM-based interfaces are primarily optimized for desktop use, requiring users' undivided attention and relying on traditional input methods like keyboards and mice, which are less accessible in mobile settings [1, 24, 37, 38]. Moreover, on-the-go information processing involves multitasking, where users must concurrently perform various cognitive and physical tasks in dynamic and complex environments [11, 27]. Consequently, a straightforward adaptation of existing LLM capabilities with OHMDs is unlikely to suffice.

In this study, we aim to investigate the integration of current Large Language Models (LLMs) with OHMDs to develop an AI-powered wearable assistant for mobile email composition. This exploration represents a stride towards realizing the potential of AI-enabled wearable assistants for managing personal information tasks while mobile. Recognizing the diverse types and uses of email, we specifically focus on everyday email categories such as personal, social, professional, and academic, which are prevalent in daily communication.

To understand the potential and viability of the LLM-based wearable email assistant and the challenges faced, we conducted a formative study ($N = 12$). We identified two main challenges of the LLM-based wearable email assistant: (1) achieving efficient and accurate understanding of user intentions, and (2) ensuring effective information presentation for email processes. We then conducted two empirical studies (both $N = 12$) to identify the best approach to interact with LLMs and to determine the optimal modality and visual output mode for facilitating effective information presentation through OHMDs. Our findings indicated that the "Single Turn with Optional Clarification" (abbreviated as *Single+*) approach was the most efficient interaction style with LLMs, and the "Fade Context with Optional Audio" emerged as the optimal output mode for OHMD mobile settings.

Integrating the results from these two studies, we developed *GlassMail* (see Fig. 6) and conducted an observation study ($N = 12$) to evaluate *GlassMail*'s feasibility for composing formal and semi-formal emails, The study showed that *GlassMail* was useful and effective for mobile email composition in simple scenarios but further improvement is needed to support email composition in complex on-the-go scenarios. From these insights, we distilled key actionable design recommendations: 1) Users should have the opportunity to clarify AI's understanding of their intention during user-agent interactions, and the method to display AI's understanding can involve word-level chunking, which is fragmented attention-friendly and minimizes cognitive workload. 2) While reducing information overload as much as possible for mobile email editing, context is still important; Consider utilizing an optional simultaneous hybrid of visual with audio output for voice-based text editing on OHMD during visually demanding tasks. and 3) Carefully designed editing schemes and learning personalization are needed for efficient mobile post-editing with LLM-based wearable email assistants to achieve final personalization.

Our main contributions are threefold:

- A formative study to provide an understanding of the potential and viability of the LLM-based wearable email assistant and the challenges faced.
- Two empirical studies to explore the optimal design solution, and accordingly develop the *GlassMail* prototype.
- An observation study to evaluate the effectiveness of *GlassMail* in simple scenarios and yield insights for complex scenarios. We then provide design implications for future wearable AI-enabled assistants.

2 RELATED WORK

To provide more context for our contributions, we discuss two areas of highly relevant work: the challenges of mobile information processing, and the potential of integrating Large Language Models for mobile information processing.

2.1 The Challenges of Mobile Information Processing

Despite the popularity of advanced smartphone devices, processing information on the go poses significant challenges. The limited input/output capabilities and the pervasive heads-down interaction posture enforced by smartphones hinder users' mobile multitasking ability to switch between different tasks and engage with their surroundings effectively [3]. Moreover, this heads-down posture compromises situational awareness and poses safety concerns, particularly in dynamic or crowded settings.

Previous studies have looked into using wearable devices like smartwatches to improve mobile information processing, such as WearMail [57] and WearWrite [43]. WearMail explored extracting information from emails via a privacy-preserving human computation workflow, without addressing the challenges of mobile content creation. WearWrite enables mobile users to create documents by leveraging the help of crowd workers, but users have limited control over final content quality. WearWrite shares similarities with our idea of building wearable human-like AI assistants to handle complex tasks in mobile scenarios, as we both aim to overcome the limitations of the mobile environment by leveraging external resources, whether it be a crowd of human workers or an AI assistant. However, one striking limitation is the physical constraints of smartwatches, which will limit the range and complexity of tasks that can be completed. This is also corroborated by previous works, where the limitations of smartwatches, such as restricted text input capabilities, lower adoption rates, and poorer multitasking performance compared with mobile phones [12, 30], are found to leave the promise of enabling productive work remains largely unrealized.

The Optical See-through Head-Mounted Displays (OHMDs) offer a promising wearable solution by facilitating a heads-up information processing [76]. We chose OHMDs because they have demonstrated superiority in on-the-go multitasking situations compared to mobile phones [22, 51]. However, several challenges need to be addressed to leverage their potential fully.

2.1.1 Transparent Display. OHMDs' transparent screens introduce challenges in text readability under varying environmental conditions, including lighting and background texture [18, 19]. Moreover, they typically exhibit lower resolution than that of average mobile phones, which further reduces text clarity. Researchers have investigated various aspects of visual text representation to enhance the reading experience on OHMDs, such as text colour [19, 45], font type [14, 39, 52], text position [54], and text spacing layout [78]. Adhering to established best practices, our study employs green text against a black background which appears transparent for optimal readability on mobile OHMDs [46, 54, 78].

2.1.2 Design for Seamless Interaction. Mobile OHMD information processing also requires users to multitask, dividing their visual attention between the OHMD display, their engaged tasks and their physical environment [41]. The limited attention span and visual focus of users may make it difficult to facilitate seamless interaction and more prone to producing content of lower quality or accuracy [65]. Challenges arise from the inherent limitations of multitasking, as well as situational demands originating from the surrounding environment, such as unexpected interruptions or ambient distractions [32, 62, 65]. Thus effective integration of OHMDs for mobile information processing requires careful interaction design that considers users' multitasking demands and limited attention spans.

2.1.3 Voice-based Editing Constraints. Text editing on the go remains challenging despite the speed advantage of speech input over typing or writing [5]. Recent studies have introduced voice editing strategies with and without explicit command keywords [21, 22]. While explicit command words (e.g., insert, replace, and delete) can enhance editing precision, they impose cognitive burdens on users as they require users to remember the original text and specific command syntax[21]. Alternatively, re-speaking a part of the text with corrections, without command words, accommodates complex edits but demands sufficient context to avoid alignment errors, leading to multiple attempts for accuracy [22, 60]. A promising approach, "Just speak it" combines both methods and supports editing by inducing commands from existing context and the edit command [16]. This method opens the promise of using the user's natural language instruction editing, yet primarily focuses on editing one or a few words. Leveraging LLM's capability for auto-correction and context-aware text processing, integrated with OHMDs, could offer an alternative interaction paradigm for natural language-based voice editing, potentially transforming the way users interact with text on the go.

2.2 Integrating Large Language Models for Mobile Information Processing

Large Language Models (LLMs) have revolutionized information processing tasks, offering immense potential [44]. LLMs have significantly simplified tasks such as writing, summarizing, and text generation [35, 40, 68, 68, 74]. However, challenges persist particularly in more complex tasks like composing personal emails that require personalization and structure. Existing LLM-based email assistants, such as LaMPost [24], designed for dyslexia users, struggle to meet accuracy and quality needs due to limitations inherent in LLMs like hallucination, content inconsistency and style, repetition, mediocrity and ethical concerns [17, 20, 50].

Integrating LLMs into mobile information processing presents additional challenges, primarily stemming from the unique demands of mobile settings. Current LLM-based assistants are optimized for desktop use, relying on traditional input methods like keyboards and mice, which are less accessible in mobile settings. One key reason is their demand for physical interaction and visual focus, which detracts from multitasking ability and hinders situational awareness in mobile settings [37, 38, 64]. While LLMs hold immense potential for enhancing mobile information processing, determining how to effectively integrate LLMs to work seamlessly with mobile information processing remains an open research question worth investigating.

3 FORMATIVE STUDY

While complex mobile information processing is challenging, the introduction of LLMs' capabilities brings new hope. However, the LLM-based intelligent assistant has not been designed or evaluated for heads-up mobile usage, thus we conducted a formative study to understand the potential and viability of the AI-enabled wearable intelligent assistant as well as the challenges faced in integrating it into OHMD mobile settings.

3.1 Methods

3.1.1 Participants. We recruited 12 university-affiliated participants (6 females, 6 males, age range: 18-28, $M = 22.8, SD = 3.54$ years). All participants had standard or corrected vision, were fluent in English, and had experience using ChatGPT for emails on desktops to ensure participants had a baseline familiarity and reduced the learning curve in the study. Half of them had used OHMDs before. Each participant was compensated at the standard rate of US$7.50 per hour.

3.1.2 Tasks. Our tasks were designed to understand the use of LLM-based wearable email assistants in everyday mobile scenarios. Specifically, the tasks involve composing emails with the assistance of LLMs while utilizing Optical Head-Mounted Displays (OHMDs) and simultaneously engaging in real-world mobility tasks, as listed below:

- **Creating Emails with LLM-based Wearable Assistant**: Participants were equipped with smart glasses to compose emails. They utilized a ChatGPT web page and controlled the process through voice commands, facilitated by a Chrome plugin.
- **Real-world Mobility Tasks**: While composing emails, participants also performed various mobility tasks, simulating scenarios that are common in daily life. We designed a route that is on the way to or from work/lecture while using public transportation, and shopping at the supermarket. The route included common mobility tasks such as walking indoors/outdoors, using stairs indoors, riding a bus and shopping indoors as identified in previous studies [22, 47], as shown in Figur 2.

3.1.3 Apparatus. In this study, we selected the Nreal Light glasses[1] for their lightweight design (106 grams) and high-resolution stereoscopic display (1920×1080 pixels), essential for clear and comfortable text reading. Participants were equipped with these smart glasses throughout the study, especially during indoor walking tasks. An experimenter accompanied them, carrying a MacBook Air (M1, 2020)[2] connected to the glasses. This setup allowed participants to interact with the default ChatGPT web page displayed on the glasses using voice control, facilitated by a Chrome plugin[3].

3.1.4 Procedures. The experimental procedure consisted of multiple steps, taking approximately 60-80 minutes per participant.

- Study Briefing & Training (20 mins): Participants received an overview of the study's procedures and were trained on using the devices involved.
- Task Execution (20-30 mins): Participants composed their pre-submitted emails using smart glasses equipped with the default LLM, assessing the practicality of the LLM-assisted method. The pre-submitted emails have to be representative of their everyday communication as a baseline for final comparison.
- Semi-structured Interviews (20-30 mins): Participants provided detailed feedback on the challenges encountered during the study and compared their experiences of composing emails with LLM-based wearable assistants.

3.2 Findings: While the Viability of LLM-Based Wearable Email Assistant is not Ideal Due to Two Main Challenges, It Holds Promising Potential.

Interviews were audio-recorded, transcribed, and coded using Thematic Analysis [7] to identify key themes related to the challenges of LLM-based wearable assistants for mobile email creation. An inductive approach was applied to derive themes based on their frequency and perceived significance in the data [58].

Our findings indicate the potential of LLM-based wearable email assistants, as a majority of participants (8 out of 12) were able to create useful initial email drafts utilizing this technology. Participants are impressed by the speed at which LLM generated emails with simple voice instructions. However, our study also revealed significant challenges associated with this approach. While the initial drafts are "useful", they are rarely exactly what the participants want. Every time, participants need to perform additional edits to finalize the content, however, the editing process, particularly in mobile settings, using the voice-only input approach, is difficult (for example, location and selection editing scope, correcting errors caused by speech artefacts). Frequently, even after repeatedly providing editing instructions, the LLM fails to change the content to match the exact needs of the user, which leads to frustration, etc. We expect if the content generated in the initial phase closely matches the user's intentions and expectations, subsequent editing

[1] https://www.nreal.ai/light/

[2] https://support.apple.com/kb/SP825?locale=en_US

[3] https://voicecontrol.chat

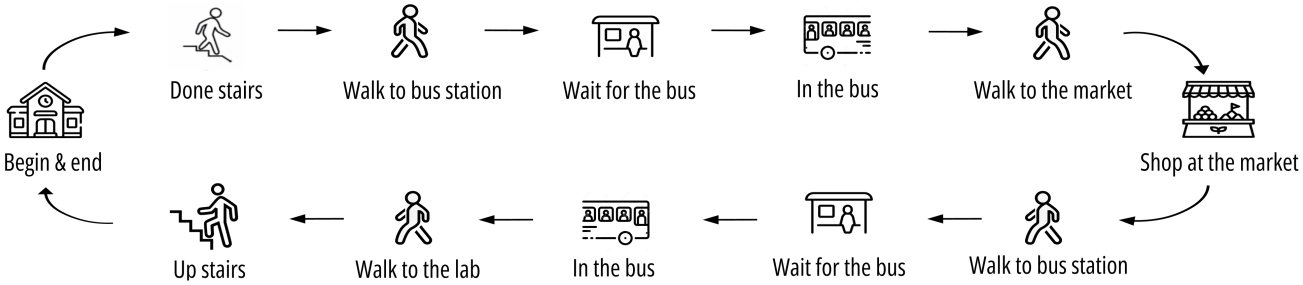


Figure 2: Details of realistic mobility tasks: Participants start by taking a bus to a supermarket to purchase items and then go back. These mobility tasks include: (1) walking indoors, (2) using stairs indoors, (3) walking outdoors, (4) riding a bus, and (5) shopping indoors.

effort can be significantly reduced, thereby enhancing user experience and overall efficiency. With careful observations, we discover that such difficulties can be attributed to two main challenges.

3.2.1 The challenge of efficiently and accurately understanding user intentions before generating initial content [C1]. Unlike keyboards and mice, which allow for quick editing of user intentions, voice-based inputs interact with LLMs in mobile contexts facing challenges steaming from both voice recognition (audio-to-text) and NLP (text-to-intent) processing. These challenges are further complicated by the presence of background noise, variations in speech patterns, and colloquial language. Additionally, mobile users frequently have limited attention to interact with LLMs on OHMDs due to multitasking constraints. This divided attention can also lead to fragmented or incomplete voice instructions to the LLM, which can be misunderstood. For instance, four participants encountered initial difficulties due to the LLM's inadequate understanding of their needs, leading to a need for further clarification. Persisting in the dialogue did not timely resolve these misunderstandings, causing two participants (P2, P7) to consider abandoning the process in favour of manual composition after 2-3 attempts. P7 specifically reported, *"I found it quite troublesome to make LLMs understand my needs [using voice-based dialogues]. It frequently failed to make any edits despite my repeated instructions, forcing me to quit and restart to save effort"*.

3.2.2 The challenge of presenting LLM responses on OHMDs for efficient email editing [C2]. In addition to difficulties associated with inputting their ideas, participants also encountered difficulties in understanding email content on OHMDs while on the go, primarily due to the challenge of dividing their visual attention between the display and their surroundings. When presenting the email in full in front of the participants, they (P2 and P8) found it *"overwhelming at times*, especially when they are multitasking. P6 highlighted the inefficiency, stating: *"Balancing attention between the external environment and my glasses' display makes it very difficult to quickly review and identify changes in the content"*. Prior research recommended displaying text sentence-by-sentence for mobile OHMD voice-based editing of a piece of text [22], yet this may not be suitable for email writing as it requires careful attention to the structure and overall style. Displaying sentences one by one may disrupt the flow of email composition, which is especially tedious when they want to edit some content. Without seeing the overall context of the email, editing particular sentences feels much more challenging.

While participants also criticise the initial content for being overly polite, lengthy, and containing irrelevant material leading to hesitancy in sending such emails, the above two challenges are crucial for facilitating email processes. Thus we decided to tackle these two challenges first: [C1] Effective and efficient understanding of users' intentions with fragmented attention and [C2] Effectively presenting information to facilitate the complex email process.

3.3 Design Goals

We establish two design goals based on the identified challenges [C1] & [C2] for building efficient LLM-based wearable assistants for mobile email creation:

- D1: Improving LLM accuracy in understanding user intentions with minimal the user's workload and attention.
- D2: Allow mobile users with fragmented attention to efficiently and effectively understand LLM response.

4 PROPOSED DESIGN PROBES FOR EACH DESIGN OBJECTIVE

Following the establishment of our design goals, we proposed design probes, based on a comprehensive overview of potential solutions aligned with our objectives.

4.1 D1: Improve the Accuracy of LLM's Understanding of User Intentions with Minimal User's Workload and Attention

4.1.1 Design Selections. The interactions between LLMs and users can be of a single-turn or multiple-turn dialogue [8]. Based on our formative study, it became evident that despite being the most convenient option, a single interaction with LLM often fails to accurately comprehend users' intentions. Prior research has demonstrated that improved response accuracy from LLMs can be attained through multi-turn dialogue interactions [10, 44]. Yet this can compromise efficiency and user workload. How to achieve the right balance between efficiency and accuracy is the focus of this investigation.

Table 1: Comparison of Single Turn, Multiple Turns and Single Turn with Optional Clarification (abbreviated as *Single+*) Approaches used in Study 1. User: 👤, LLM: 🤖.

Single Turn Approach	Multiple Turns Approach	Single+ Approach
👤: Send an email to my project member Kayla about user feedback. 🤖: [Generated Email Content]	🤖: What is the email's main topic? 👤: Email my project member Kayla about user feedback 🤖: What tone would you like to convey in your email? 👤: Casual 🤖: What is your relationship with the email recipient? 👤: Colleague 🤖: How long would you like the email to be? 👤: Short 🤖: What is the subject line of the email? 👤: Update User Feedback 🤖: What name would you like to use as the sender of the email? 👤: Liz 🤖: What is the name of the email recipient? 👤: Kayla 🤖: [Generated Email Content]	👤: Send an email to my project member Kayla about user feedback. 🤖: Main Topic: User Feedback Tone: formal Relationship: Project Member Subject: Update User Feedback Length: N.A. Sender: N.A. Recipient: Kayla Can you specify the email length and the sender's name of the email? 👤: My name is Liz and I prefer a short length. 🤖:[Generated Email Content]

4.1.2 Interaction Approaches. We present three approaches with different emphases between efficiency and accuracy, as detailed in Table 1. We are curious to find out which method is preferred by users and can achieve satisfying email composition results.

- **Single Turn Approach**. This approach asks the user to provide all necessary information in a single interaction with the LLM. This method is the most efficient, but may not always capture the user's intentions accurately, especially in complex scenarios.
- **Multiple Turns Approach**. This approach involves a series of single-question interactions, each focusing on a specific aspect of the email's content, determined by its structural features. These questions guide the user in forming a comprehensive email request. This method is more likely to accurately understand and reflect user intentions through iterative questioning, while it can be time-consuming and may increase cognitive workload for the user, especially in mobile settings.
- **Single Turn with Optional Clarification Approach (abbreviated as *Single+*)**. This approach allows users to express their needs in a single-turn input, followed by one optional turn with LLMs for user clarification. This method takes the middle ground between the first two approaches.

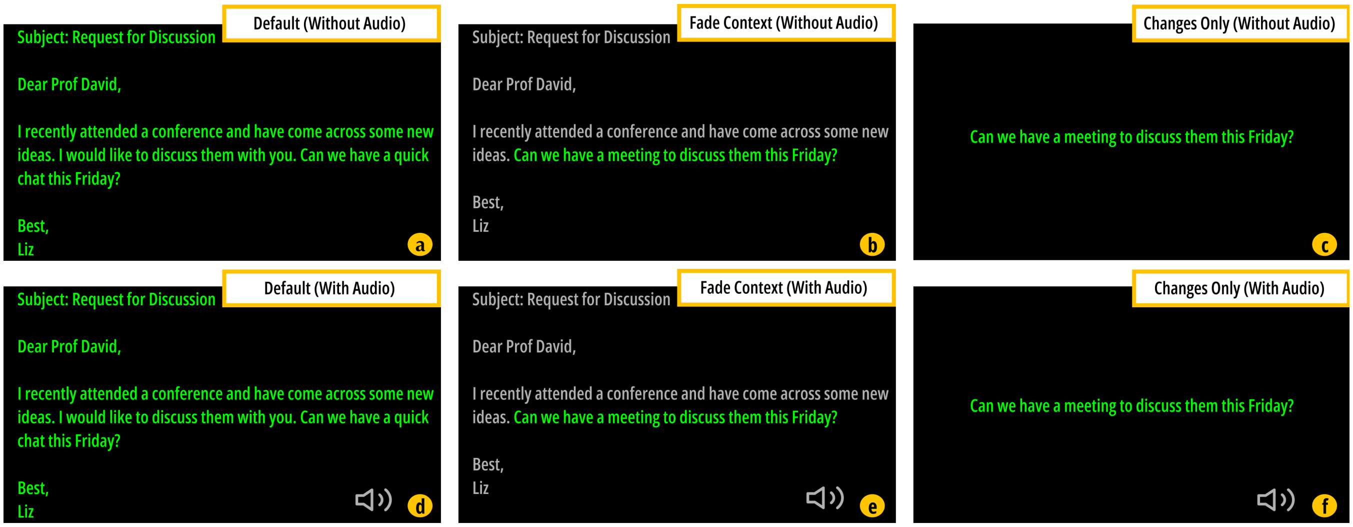
4.2 D2: Allow Mobile Users with Fragmented Attention to Efficiently and Quickly Understand LLM Edits

While the previous investigation focused on input, this investigation seeks to find a harmonious balance in presenting LLM edits to facilitate quick comprehension while minimizing cognitive load. We focus on two key elements as prior work suggested [22]: output modality and the visual output mode.

4.2.1 Design selections.

Output Modality. While the output modality can be audio only, visual only, or a hybrid. Previous work revealed that editing text on OHMD in an audio-only mode was cognitively very challenging if a continuous stream of audio was presented [23]. Thus we focus on the remaining two (see Fig.3): visual only and a hybrid of visual with audio.

Visual Output Mode. Given both visual-only and hybrid approaches involve visual presentation, we further look into how to effectively design the visual output mode. For text editing, increasing the amount of displayed text can lead to faster error correction as it provides more text context, reducing the number of navigation operations needed to scan through the text. Yet displaying more text might also cause more distraction, thereby increasing users' cognitive workload and path-navigation challenges. This issue is particularly sensitive in mobile settings, where users' attention is split between reading the text on OHMDs, engaging in other tasks, and maintaining awareness of their environment. A previous study suggests that sentence-level presentation that doesn't include context information on OHMDs is suitable for voice-based on-the-go text editing [22]. Yet, our formative study has ruled out this option as users found it difficult to obtain the context information, which is crucial for email composition and editing. Concerning the presentation of email edits, the question arises: Is the changes only" approach (i.e., sentence-level) also the optimal visual mode for email editing? Is context important when editing email content, which often requires a specific structure and careful wording? We hypothesize that context is indeed important and propose exploring



Subject: Request for Discussion

Dear Prof David,

I recently attended a conference and have come across some new ideas. I would like to discuss them with you. Can we have a quick chat this Friday?

Best,
Liz

Subject: Request for Discussion

Dear Prof David,

I recently attended a conference and have come across some new ideas. Can we have a meeting to discuss them this Friday?

Best,
Liz

Can we have a meeting to discuss them this Friday?

Subject: Request for Discussion

Dear Prof David,

I recently attended a conference and have come across some new ideas. I would like to discuss them with you. Can we have a quick chat this Friday?

Best,
Liz

Subject: Request for Discussion

Dear Prof David,

I recently attended a conference and have come across some new ideas. Can we have a meeting to discuss them this Friday?

Best,
Liz

Can we have a meeting to discuss them this Friday?

Figure 3: All conditions used in Study 2: a) Default (Without Audio), b) Fade Context (Without Audio), c) Changes Only (Without Audio), d) Default (With Audio), e) Fade Context (With Audio) and f) Changes Only (With Audio).

different methods to reduce the workload caused by adding context, such as using an unnoticeable colour to display unchanged text, to balance the need for context against the risk of cognitive overload.

4.2.2 Three Visual Output Modes. We designed three visual output modes for our study: Default, Fade Context, and Changes Only. The primary distinction among these modes is in the presentation of the email context, apart from the edited changes. Figure 3 illustrates all these output modes.

- **Default Mode**: It displays the entire email content, including both edits and unchanged parts.
- **Fade Context Mode**: Only the edits are highlighted in green, while the unchanged context is displayed in a less noticeable colour like grey, to reduce visual prominence.
- **Changes Only Mode**: This mode displays only the edited changes, using green colour for text, thereby omitting the unchanged context entirely.

5 STUDY 1: EFFICIENT CLARIFICATION OF USER INTENTIONS

We conducted a study to compare three interaction approaches (see Table 1) for interacting with LLM-based wearable email assistants. Our goal is to explore an optimal approach that enhances LLM accuracy in understanding user intentions with minimal user workload and attention [D1].

5.1 Methods

5.1.1 Participants. We recruited 12 participants (5 females, 7 males) between 18-24 years old ($M = 20.3, SD = 1.76$) from the university community. All participants had normal or corrected-to-normal vision with no colour deficiency and were native or fluent in English. Four of them had prior experience using OHMDs. Participants were compensated at the standard rate of US$7.50 per hour.

5.1.2 Apparatus. Participants wear the Nreal Glasses connected with MacBook Air (M1, 2020) for OHMD display. They also need to wear a SANWA ring mouse (400-MABT156BK, Bluetooth)[4] which is the best practice for OHMD mobile seamless interaction usage from prior work [56, 76] (Apparatus see 4(c)). React, Typescript, Ionic, GPT-3.5-turbo-16k model, and Node.js were used to develop the *GlassMail* application hosted on the MacBook Air. The MacBook Air was used to host the node email server to send out emails as its screen mirroring was most similar to the Nreal glasses and offered users better readability and simplicity. The laptop was in a lightweight bag that participants could easily carry.

5.1.3 Tasks. We chose email tasks from email analysis pilots according to their utility and usage in daily life and private concerns. The given email tasks (see Appendix A.3) are event coordination email tasks, such as sending invitations, reminders, or updates about upcoming events to three different social ties (i.e., supervisor, sister, friend) [53]. The order of materials was counterbalanced with the Latin Square design. As for the mobility task, participants composed emails while walking back and forth on a 30m long straight path consisting of objects like dustbins, tables, and chairs along the sides, simulating common daily environments.

5.1.4 Design and Procedure. A repeated measure within-subject design was used. The independent variable was Interaction Approaches: 1) Single Turn Approach; 2) Multiple Turns Approach; 3) Single Turn with Optional Clarification (abbreviated as *Single+*) Approach. The order of interaction approaches was counterbalanced with the Latin Square design

Each participant was required to compose emails according to the given email tasks using three interaction approaches. For each interaction approach, they need to compose three different emails. Before proceeding to the next condition, we ask participants to

[4] https://direct.sanwa.co.jp/ItemPage/400-MABT156BK

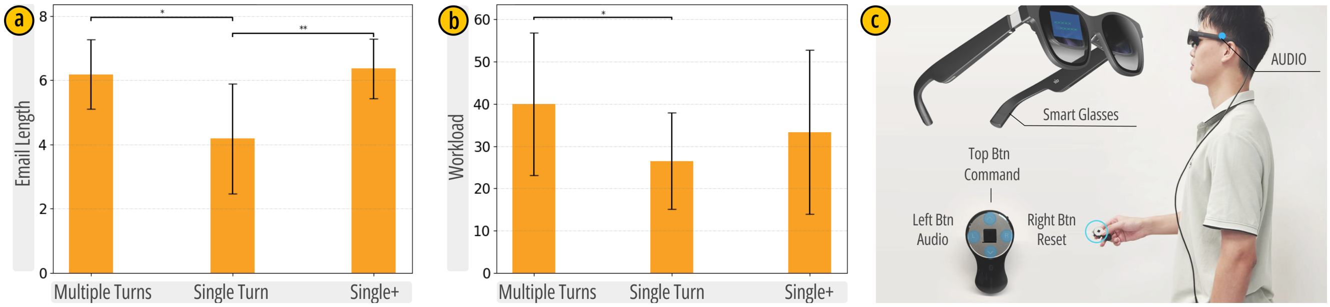


Figure 4: Study 1 results: Means and standard deviations (SD) of significant measures: a)User appreciation of generated email length b) NASA-TLX workload. The apparatus (c) shows all devices including the NReal OHMD and Ring Mouse.

complete a questionnaire relating to perceived task load and evaluate the quality of the generated email content. We also offer an optional 2-minute break. On completion of the study, they need to fill in a post-questionnaire survey to collect their overall preference ranking across all conditions, followed by a semi-structured interview to understand their issues and requirements.

5.1.5 Measures. The evaluation of each interaction approach in this study included three key components: 1) NASA-TLX task workload [26], 2) quality of generated content regarding appreciation of tone and length, as well as satisfaction, measured on 7-point Likert scales. [31], and 3) overall user preference.

5.2 Quantitative Findings

All results met the normality assumption of ANOVA (Shapiro-Wilk tests, $p > .05$). Therefore, a one-way repeated measures ANOVA was performed, followed by Bonferroni corrected post hoc analysis.

We found the main effects of Interaction Approaches on Task Workload ($F(2, 20) = 4.94, p < .05, \eta^2 = 0.33$) and appreciation of length ($F(2, 20) = 7.37, p < .05, \eta^2 = 0.42$) (See Figure 4 (a-b)). Pairwise comparisons revealed the Multiple Turns ($M = 39.97$) Approach in the highest task workload compared with the Single Turn approach ($M = 26.49$, $p < .05$) and Single+ approaches ($M = 33.33$, $p > .05$). Also, the Single Turn approach resulted in an email length that was less appreciated ($M = 4.28$), significantly worse than both Multiple Turns ($M = 6.18$) and Single+ approaches ($M = 6.36$)(both $p < .05$). No significant differences were found between the Single+ approaches and Multiple Turns on all measurements (all $p > 0.05$). The effect of Interaction Approaches on tone and satisfaction was not statistically significant ($p > 0.05$).

While the Single Turn approach is efficient, it may not meet user expectations in terms of email length and tone. Conversely, the Multiple Turns and Single+ approaches, necessitating more interaction (with the Multiple Turns approach being particularly effort-intensive), generate content that better aligns with participants' final expectations. These methods effectively address critical aspects such as length and tone, which are often neglected in the Single Turn approach.

5.3 Qualitative Findings

Interviews were audio-recorded, transcribed, and coded using Thematic Analysis described in Braun and Clarke's methodology [7]. Then an inductive approach was used to derive themes based on frequency and perceived significance [58].

Our findings indicate a preference for the *Single+* approach for OHMD mobile use with LLMs. Most participants (8 of 12) preferred this approach because it could *"ask for clarification for important parts of email"* (P1, P2, P3, P8) while not going too far in *"asking information that is already available"* (P4, P11, P12), thereby balancing content quality and workload effectively. This approach also helped participants visualize what they wanted to write before the email was generated, offering participants the possibility of rectifying it whenever they wanted (P5, P8). Yet, the way used in the Single+ approach to display crucial parts of emails needs better design, as some participants (P2, P8, P11) complain that *"it draws too much attention and visual focus"*. Four participants favoured the swifter Single Turn approach despite its lower content quality. None opted for the Multiple Turns approach as it *"still sometimes missed details"* (P2, P6, P7) while being cumbersome to use. It was noted that essential aspects like the email's tone and length often required further queries for satisfactory content generation.

Overall, as we expect, the Single Turn input with an additional turn for clarification (i.e., the Single+ approach) strikes an optimal balance, providing a satisfactory level of accuracy without significantly increasing the workload for the user.

6 STUDY 2: EFFECTIVE EMAIL EDITING WITH MULTIMODALITY SUPPORT

We then conducted a Wizard-of-Oz (WOZ) study and focused our investigation on two factors: output modality (visual, or a visual and audio hybrid) and visual output mode (Default, Fade Context or Changes Only). We aim to seek a harmonious balance in presenting LLM edits to facilitate quick comprehension while minimizing cognitive load [D2].

6.1 Methods

6.1.1 Participants. We recruited 12 participants (5 females, 7 males) between 18-24 years old ($M = 20.3, SD = 1.76$) from the university community. All participants had normal or corrected-to-normal vision with no colour deficiency and were native or fluent in English. Three of them had prior experience using OHMDs. Participants were compensated at the standard rate of US$7.50 per hour.

6.1.2 Apparatus and Materials. We conducted a Wizard-of-Oz (WOZ) study [55] using Nreal Light glasses. The smart glasses mirrored the display of an iPad Mini tablet[5], showcasing various displays (the visual output modes) crafted using PowerPoint slides[6], hosted on a MacBook Air (M1, 2020) and shared to the iPad via Zoom [4].

We replicated the mobility task and email tasks from Study 1, focusing on three prevalent email editing actions noted in an earlier pilot: 1) removing or editing irrelevant content; 2) modifying sentence tone or phrasing; and 3) altering closings or signatures. For the hybrid visual and auditory condition, the email contents were translated into audio files through an online service[7] at a default pace and incorporated into the appropriate slides.

6.1.3 Design and Procedure. A repeated-measures within-participant design was used. The independent variables were visual output mode VMode (Default, Fade Context, Changes Only) and Audio (With, Without).

A fully crossed design resulted in 6 combinations per participant (see Fig.3). Before proceeding to the next condition, we asked participants to complete a questionnaire relating to perceived task load and information absorption and also provided an optional 2-minute break. After completing all six conditions, we collected their overall preference ranking across all conditions. We asked participants to elaborate on their choices through our semi-structured interview, before concluding the experiment.

6.1.4 Measures. We measured each output mode's task workload (i.e., NASA-TLX [26]), information absorption in easier visual search for edited changes, and support for multitasking using 7-points Likert Scales [31], as well as user overall preference for all output modes. This scale is adapted from prior works [9, 35] and we further make it more suitable for our settings.

6.2 Quantitative Findings

A 3x2 within-subjects factorial ANOVA followed by Bonferroni corrected post hoc analysis was used to analyze all data. Both the assumption of sphericity (Mauchly's test of sphericity, $p > 0.05$) and the assumption of normality (Shapiro-Wilk test, $p > 0.05$) were met for all.

There was no significant effect of the Audio ($p > 0.05$) and interaction effect of VMode x Audio ($p > 0.05$) on all measurements.

For the visual output mode (VMode), we found the main effects on Task Workload ($F(2, 22) = 11.18, p < .001, \eta^2 = 0.50$), Support Easier Visual Search ($F(2, 22) = 24.78, p < .001, \eta^2 = 0.69$), and Support Easier Multitasking ($F(2, 22) = 14.02, p < .001, \eta^2 = 0.56$) (See Figure 5). Pairwise comparisons revealed a similar trend that Fade Context ($M = 17.44$) resulted in the lowest Task Workload, easiest visual search, and multitasking (all $p < .05$) compared with Default and Changes Only. No significant differences were found between Default and Changes Only on all measurements (all $p > .05$). This suggests that Fade Context is the optimal visual output mode for presenting LLM's editing changes with efficient information absorption and minimal workload for OHMD mobile scenarios.

6.3 Qualitative Findings

Interviews were audio-recorded, transcribed, and coded using Thematic Analysis [7], followed by an inductive approach to derive themes based on the frequency and perceived significance [58].

6.3.1 Participants preferred the Fade Context for quicker visual email editing, with mixed views on audio assistance; it aided multitasking for some but seemed slow or distracting for others. Eight Participants preferred the Fade Context while four participants preferred the Fade Context with Audio. They thought the Fade Context could *"allow easier/faster visual cues on changes while keeping in context of the whole email."* The absence of contexts in the Changes Only mode might *"make it hard to determine the changes made"* (P1, P3) and *"not sure if the entire email still flows logically and the edits made the email better or worse than before"* (P12). The Default mode was slightly better than the Changes Only mode as it still kept the context of the email while still having difficulty in finding the changes made and can *"be overwhelming"* (P3, P4). In addition, for combination visuals with the audio or without audio, half of the participants preferred with audio while the rest preferred without audio. Participants preferred visuals with audio because they thought audio could help *"multitasking"* (P1, P6, P8, P9) and *"consider whether the sentence flows naturally"* (P2). Yet they also mentioned that the current audio was too slow when compared to the speed of glancing through the content (P9, P12).

6.4 Pilot Study Suggests Optional Audio Benefits for Visually Demanding Mobility Tasks.

The Fade Context mode emerged as the optimal visual output mode for displaying email contents and edits on OHMD during mobile usage, adeptly balancing the presentation of edited changes with reduced visual/cognitive load. While pairing visual displays with audio did not markedly impact performance, this may be because our testing scenarios (i.e., simple walking indoors) were not that cognitively demanding. We then conducted a pilot study with four participants. We used the same design and procedure as this study, while we redesigned the mobility task to require a higher visual load based on the setting in Zhou et al. [78]. The floors were taped to outline a rectangular path with a perimeter of 30 meters (width of 8m) that participants followed. A sign with 4 locations listed was pasted on each wall along their path. Signs were placed 2.5m away from the path, in the participants' line of sight. Participants were asked to read the signs as they passed by them. Thus participants needed to shift their visual attention between the OHMD display and environmental signs while on the move. All participants emphasized the usefulness of audio for attention switching, indicating that relying solely on visual input might be insufficient for maintaining information processing when their visual attention is occupied or limited. Furthermore, the utility of audio may vary based on individual preferences and past experiences with audio use [2, 6]. Given this variability in preference and the fact that audio may still be useful during situations that are more cognitively demanding [48, 61], we propose adopting the Fade Context mode with optional audio as the optimal output design for our system.

[5] https://www.apple.com/sg/ipad-mini/
[6] https://www.microsoft.com/en-sg/microsoft-365/powerpoint
[7] https://freetools.textmagic.com/text-to-speech

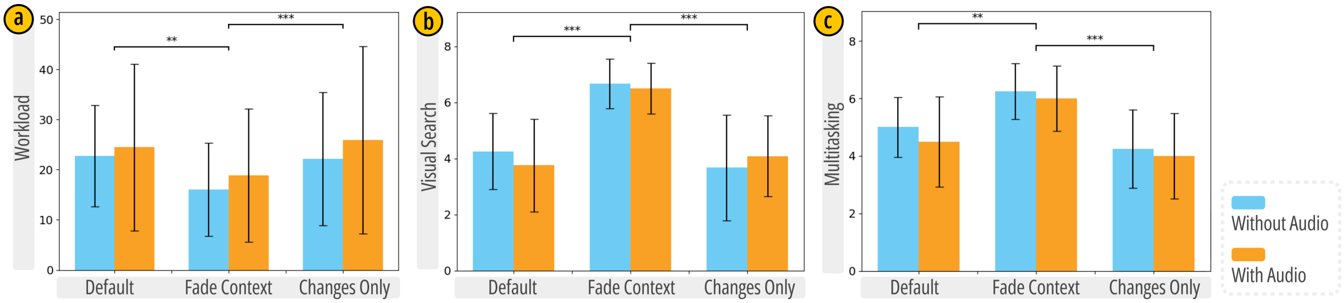


Figure 5: Study 2 Results: Means and standard deviations (SD) of significant measures: (a) NASA-TLX workload, (b) Easier Visual Search, (c) Easier Multitasking for both visual output with and without audio mode. Lower NASA-TLX workload scores suggest reduced cognitive load and increased task efficiency, while higher scores in easier visual search and multitasking tasks indicate superior perceptual and cognitive capabilities, leading to enhanced multitasking performance.

In summary, as we expect, context is important in editing emails. Using an inconspicuous colour like grey to fade the context display could facilitate easier visual search, enhance multitasking, and reduce cognitive workload. Additionally, while the results do not support that audio is particularly crucial, our participants' feedback and pilot studies suggest that optional audio assists in visually demanding tasks, helping to reduce workload and improve participants' understanding of visual information.

7 GLASSMAIL PROTOTYPE

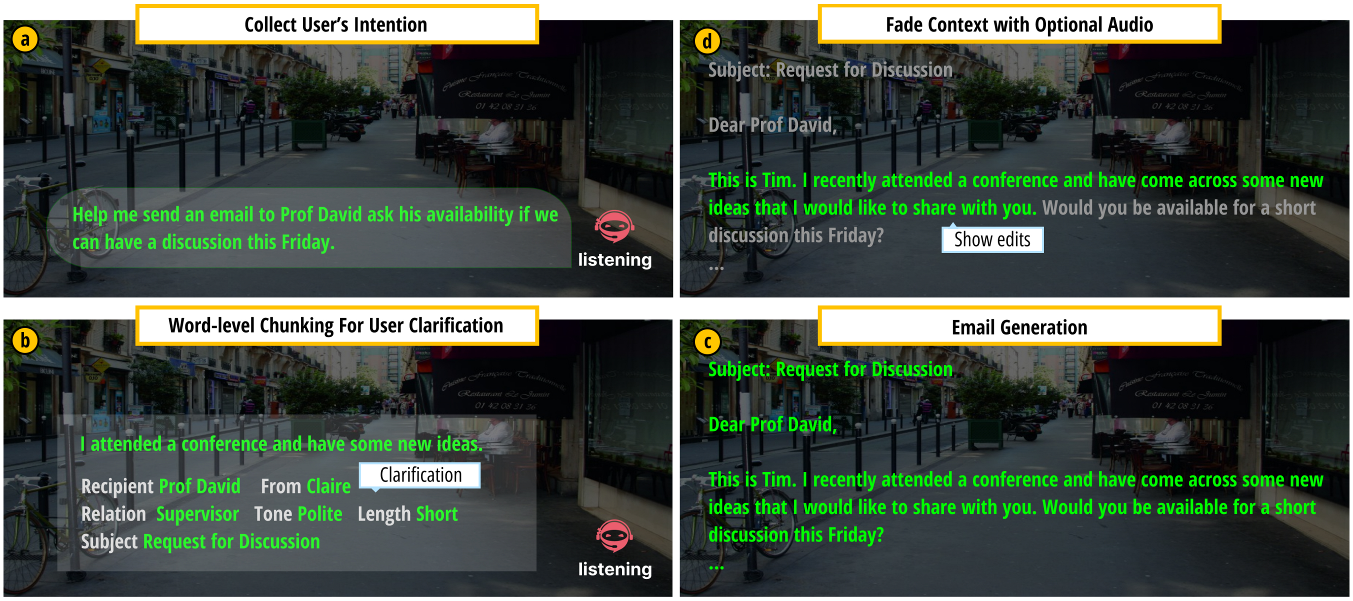
The findings from two empirical studies inform our interface design. We developed *GlassMail* features a "Single Turn with Optional Clarification" approach for accurate user intention recognition and a "Fade Context with Optional Audio" mode for effective email processing. *GlassMail* incorporates a hybrid interaction approach of voice and wearable ring-mouse input. Voice is used to interact with LLM-based wearable assistants, while ring-mouse facilitates quick and seamless confirmation tasks. The full view of the *GlassMail* interface is illustrated in Figure 6.

7.1 Key Features

7.1.1 Input Interaction Style with LLMs is Single Turn with Optional Clarification Approach. Our study1 suggests that a Single Turn with Optional Clarification interaction strategy for LLM-based assistants can strike a balance between efficiency and usability in engagements with an LLM-based assistant, typically requiring no more than two turns of conversation. However, mobile environments introduce unpredictability due to inaccurate capture of voice-based input prompts and users' limited attention, leading to misinterpretations. Rectifying these misinterpretations needs further effort from users thereby defeating the simplicity of the Single Turn with Optional Clarification approach. To mitigate this, we propose a design with real-time voice interactions and "Fragmented attention-friendly chunking" to break information into manageable parts for divided attention scenarios. This aims to improve error detection and correction in voice inputs, enhancing user experience and reducing cognitive load. We provide detailed explanations of each element of this design strategy.

- **Real-time Voice Recognition to Collect User's Intention**: Real-time voice recognition is crucial. While it demands more visual and cognitive attention from the user, the effort is justified compared to the significant workload caused by correcting content generated by misinterpreted intents. Echoing Myers et al.'s [42] findings, users often opt to quit or restart upon encountering misinterpreted intent. To enhance *GlassMail*'s usability, we implemented real-time voice recognition visible to users as they speak (see Figure 6 (1)), coupled with a straightforward method for making corrections by repeating words or phrases.
- **Word-level Fragmented Attention-Friendly Chunking Display for User Clarification**: *GlassMail* distills key elements from users' voice input. These elements are then displayed in compact word-level segments (see Figure 6 (2)), simplifying the task of identifying and rectifying inaccuracies, missing parts, or necessary additions for composing emails. This method could effectively handle the issue of revising voice prompts post-entry. Contrary to real-time voice interaction corrections, modifying entered prompts is significantly more challenging and time-consuming, a situation compounded in mobile settings where conventional input tools like keyboards and mice are absent, making sentence navigation and selection difficult. Especially in scenarios where users are multitasking, physically engaged, or cognitively constrained, their capacity for patience and attention is naturally limited.

7.1.2 The Optimal OHMD's Output Mode is Fade Context with Optional Audio Mode. *GlassMail* features the "Fade Context with Optional Audio", specifically designed for the OHMD mobile environment (see Figure 6 (4)). This feature utilizes an unnoticeable colour, such as grey, to diminish the visibility of unchanged text, thus highlighting edits. This approach effectively eases the identification of edits made by the LLM, aiding users in quickly understanding these modifications. Furthermore, it enables users to efficiently process the LLM's output with the aid of optional audio, even when engaged in multitasking or facing situational constraints. This strategy aims to minimize visual and cognitive strain without affecting the primary task.



Help me send an email to Prof David ask his availability if we can have a discussion this Friday.

Subject: Request for Discussion

Dear Prof David,

This is Tim. I recently attended a conference and have come across some new ideas that I would like to share with you. Would you be available for a short discussion this Friday?

...

I attended a conference and have some new ideas.

Recipient Prof David From Claire
Relation Supervisor Tone Polite Length Short
Subject Request for Discussion

Subject: Request for Discussion

Dear Prof David,

This is Tim. I recently attended a conference and have come across some new ideas that I would like to share with you. Would you be available for a short discussion this Friday?

...

Figure 6: The interface of *GlassMail* features the Single Turn with Optional Clarification approach to interact with LLM-based wearable email assistants. The Single Turn with Optional Clarification approach includes (a) Real-time recognition to collect the user's intention and (b) Word-level Fragmented Attention-Friendly Chunking display for user clarification. Once it generates email (c), it also utilizes the (d) Fade Context with Optional Audio as the output mode for displaying editing changes.

7.2 *GlassMail* Workflow

As shown in Figure 6, users begin their Single Turn with Optional Clarification interaction with *GlassMail* by activating real-time voice interaction, accomplished by pressing the upper key on the ring mouse. They proceed to dictate their email creation requirements, with the flexibility to make straightforward modifications via voice. After completing their description, they press the down key on the ring mouse to start the system's analysis of their voice input. Post-analysis, key information is presented using the "Fragmented Attention-Friendly Chunking" method, enabling users to further clarify specifics or choose to skip directly to email generation by clicking the right key of the ring mouse. Once the email is created, users can reactivate real-time voice interaction to articulate their editing needs by pressing the up key on the ring mouse again. Then they press the down key on the ring mouse to activate the *GlassMail* system to process and display these edits using the "Fade Context" feature. The *GlassMail* system also auto-plays the audio of the modified content. Users have the option to turn off this audio by clicking the left key on the ring mouse. If necessary, they can reactivate the audio playback of the current email by pressing the left button on the ring mouse once more.

7.3 Implementation

GlassMail employed the GPT-3.5-Turbo-16K model as its primary Large Language Model (LLM) to process the user's inputs. This model is tasked with extracting key elements from these inputs to enable the "Fragmented Attention-Friendly Chunking" display, as well as to create and edit email. Several prompt engineering techniques such as prompt-chaining process [24, 33] and Chain-of-Thought Prompting [10, 63] were utilized to enhance the output (further details in Appendix A.2). *GlassMail*'s real-time voice interaction is powered by three core components: a Text-to-Speech (TTS) engine, an asynchronous Automatic Speech Recognition (ASR) engine, and a control (CTRL) module. *GlassMail* also leverages the jsdiff[8] library to implement the "Fade Context" feature and uses the TTS engine to support the audio modality, enhancing user interaction and understanding.

- **TTS Engine**: To support capabilities for eye-free listening to email content. We use the SpeechSynthesis API[9] from web browsers to enable TTS.
- **ASR Engine**: First, the RecordRTC[10] library is used to capture and record the user's voice input. Speaking detection is implemented through the hark library[11], monitoring when the user starts and stops speaking to initiate and terminate recording. Audio data is encoded into MP3 format using the lamejs library[12] to reduce file size. Recorded audio data can be sent in real-time to the OpenAI Whisper API[13], which returns recognition text. Robust error handling is implemented to manage exceptions, including issues with acquiring media streams, recording failures, or recognition errors.

[8] https://github.com/kpdecker/jsdiff
[9] https://developer.mozilla.org/en-US/docs/Web/API/SpeechSynthesis
[10] https://github.com/muaz-khan/RecordRTC
[11] https://github.com/otalk/hark
[12] https://github.com/zhuker/lamejs
[13] https://openai.com/research/whisper

- **CTRL Module**: This module integrates ASR and TTS engine into a closed-loop design and manages control operations such as start, stop, and reset.

8 OBSERVATION STUDY: EVALUATING THE FEASIBILITY AND EFFECTIVENESS OF *GLASSMAIL*

While *GlassMail*, was developed based on the optimal solutions derived from two previous design studies, its effectiveness in mitigating two main challenges C1& C2 in mobile contexts remains uncertain. Additionally, it is unclear how significantly *GlassMail* can reduce the extensive post-editing efforts required for email creation. Thus, we conducted an in-lab observational study to assess *GlassMail*'s feasibility and to understand the challenges and potential strategies for editing interactions with *GlassMail*, aiming to achieve final user personalization. While an in-situ exploration would have provided valuable insights into the system's performance and user experience in diverse mobile contexts, we believed that conducting an initial in-lab observational study was a necessary first step. This approach allowed us to establish a solid foundation for understanding *GlassMail*'s feasibility and to identify potential challenges and strategies for editing interactions in a controlled setting.

8.1 Methods

8.1.1 Participants. We recruited 12 participants (7 males, 5 females) between 19-23 years old ($M = 21.3, SD = 1.5$) from the university community. All participants had a normal or corrected-to-normal vision with no colour deficiency and were native or fluent in English at the university level. Three of them had prior experience using OHMDs. Participants were compensated at the end of the experiment with the standard rate of US$7.50 per hour.

8.1.2 Apparatus and Materials. We used *GlassMail* as described in Section 7. For a realistic composition and editing experience, participants were asked to provide three email scenarios and the emails they had written in each scenario involving supervisors, friends, and family members they often mailed in reality. These scenarios were used as the email composition and editing tasks in the study. The provided emails served as baselines for evaluating the usefulness and effectiveness of the *GlassMail*. We replicated the mobility task from Study 1.

8.1.3 Procedures. Each participant experimented in one session lasting approximately one hour. The session was blocked by the real email scenario provided by the participant. Participants were tasked with composing and editing emails using our system while walking until they felt that the content was similar to their writing style. All levels of editorial precision required to achieve their desired personal email content when collaborating with *GlassMail* have been recorded and analysed. After completing all three email scenarios, detailed semi-structured interviews were conducted to gather insights about their overall experience of *GlassMail*, their editing needs and challenges they faced, and strategies they adopted to address the issues, as well as suggestions for improving the system.

8.2 Overall Findings

To better understand users' experiences with creating mobile emails using *GlassMail*, we analyzed a sample user journey map of Participant 8 (P8), illustrated in Figure 7.

8.2.1 Users' Editing Process with GlassMail: From Global Adjustments to Detailed Personalization. Overall, all participants began with high-level editing adjustments, with a particular focus on email length and tone. Most of them visually skimmed through the email to get a sense of the tone, while some participants, like P8, preferred to listen to the audio to understand whether the tone was perfect and also identify any potential misunderstandings. Participants then started deleting, replacing, or adding content by making a sentence (all participants) or paragraph-level edits (P3, P8, P10). Until the desired length and tone are reached. Consequently, they will proceed to make detailed adjustments to modify specific sentences or words. For example, P7 noted that during his reading of a draft, he consistently removed sections of text that appeared overly formal or unnecessary. Additionally, he incorporated brief sentences between paragraphs to align with his writing style.

8.2.2 Efficient and Accurate Interpretation: Ensured understanding of user intentions with style-aligned, comprehensible final emails. The successful interaction of all participants with *GlassMail* to generate email content represents a notable improvement over the previous formative study, where only 8 out of 12 participants were able to initiate their first interaction with LLMs. Moreover, two participants even abandoned the process in the formative study. This improvement indicates that *GlassMail*'s interface and functionality are more user-friendly and intuitive, allowing for smoother initial interactions with LLMs. Furthermore, *GlassMail*'s ability to enable participants to align the final emails with their desired styles, with no reported challenges in understanding *GlassMail*'s edits, highlights the platform's effectiveness in facilitating effective communication. P8 found it did a very good job in *"presenting my scenarios in great detail and just needed a very few edits before confirming the mails"*. P3 and P9 appreciated its efficiency in *"performing well in formal and semi-formal contexts without any challenges"*. This suggests that *GlassMail* is effectively translating the participants' input and intent into coherent and stylistically appropriate emails. This seamless integration of user preferences and LLM-generated content demonstrates the platform's potential to enhance productivity and streamline email communication for participants.

8.2.3 Usability and Efficiency: Streamlined creation of neutral-tone, straightforward emails for time savings. More than half of the participants appreciated *GlassMail*'s usability. P2 found it useful for neutral-tone office emails, stating, *"AI assistance can save time if I don't need to maintain my character and feelings, it would be useful, such as in an office setting, where it is good to keep my tone neutral."*. P11 thought the time taken was the same for typing the email but believed *"it will save time when I want to draft some emails which are not important."*. P5 appreciated its convenience for simple emails: *"Just telling the AI to convey my simple message is quite convenient."*. P6 noted its time-saving aspect: *"I can have the first version of a grammatically correct and natural email within several minutes."*. P7 found *"a decent balance for formal emails"*, and P9 acknowledged

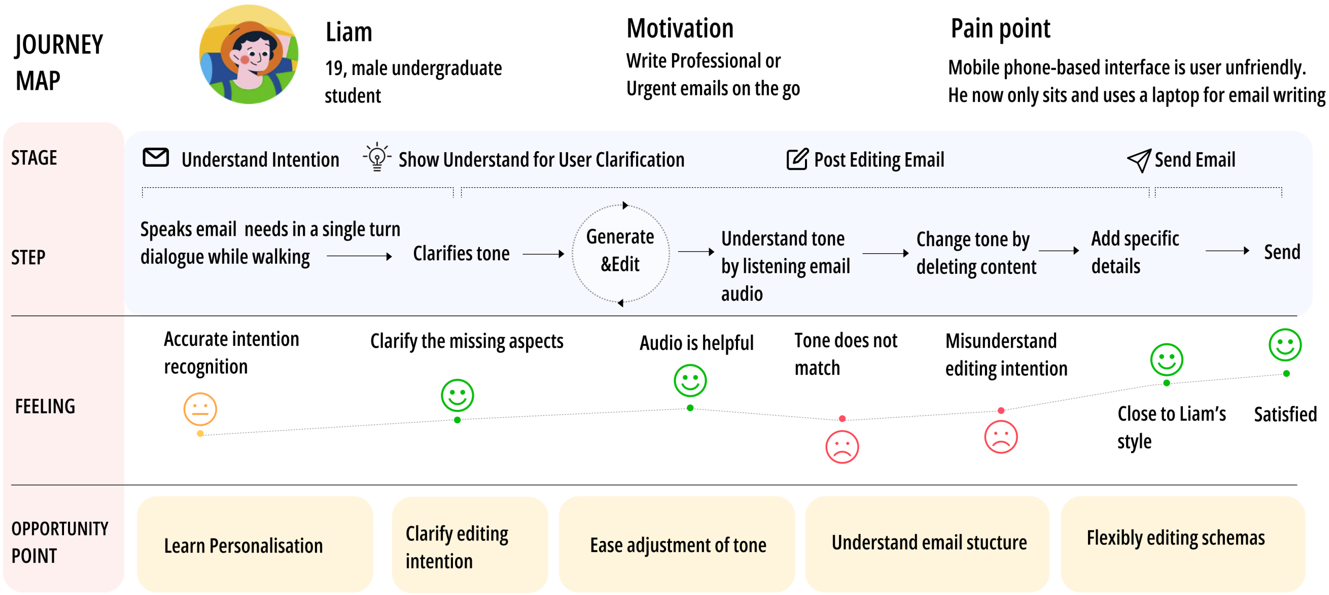

JOURNEY MAP
Liam
19, male undergraduate student
Motivation
Write Professional or Urgent emails on the go
Pain point
Mobile phone-based interface is user unfriendly.
He now only sits and uses a laptop for email writing
STAGE
Understand Intention
Show Understand for User Clarification
Post Editing Email
Send Email
STEP
Speaks email needs in a single turn dialogue while walking
Clarifies tone
Generate &Edit
Understand tone by listening email audio
Change tone by deleting content
Add specific details
Send
FEELING
Accurate intention recognition
Clarify the missing aspects
Audio is helpful
Tone does not match
Misunderstand editing intention
Close to Liam's style
Satisfied
OPPORTUNITY POINT
Learn Personalisation
Clarify editing intention
Ease adjustment of tone
Understand email stucture
Flexibly editing schemas

Figure 7: A Sample User Journey Map of Participant 8 (P8)

its basic structure formation but required personalization *"to make it more personalised will require some effort from users."*.

8.2.4 Feasibility of Formal Mobile Emails: GlassMail Showcases the Potential for Composing Formal Emails on Mobile Devices Effectively. In the observation study, the majority of participants (9 out of 12) appreciated the effectiveness of *GlassMail* in composing formal and semi-formal emails. P8 highlighted *GlassMail*'s capability in *"presenting my scenarios in great detail with very few edits needed,* while P6 remarked, *"It can understand me fairly well."*. This effectiveness is likely bolstered by features like the "Fade Context with Optional Audio" mode, which aids participants' effective comprehension of email edits. The "Single Turn with Optional Clarification" approach incorporating the concept of fragmented attention chunking display, enhances accuracy and natural expression by allowing participants to see the *GlassMail*'s understanding of their intentions with opportunities for clarification, facilitating real-time corrections.

8.3 Remaining Issues

8.3.1 Personalization Challenges: Users experienced difficulties in customizing AI-generated content to reflect personal style, raising concerns over authenticity. Conversely, P3 and P4 faced challenges in aligning AI with personal style. P3 felt *"more effort trying to get it to edit to my writing style"* was needed, and P4 found it slower than manual writing. P8 raised concerns about AI-generated content: *"Of course, AI assistance saves my time and effort, and the quality is definitely better than my own emails, but the only thing I am concerned about is how we can differentiate what's true information because the model sometimes generates hypothetical details which are not true, and this may give the recipient a wrong impression. A label of AI-generated should be added with some quantified value of user input."*. P10 preferred *"composing the email myself, but probably AI could help in generating phrases and sample sentences instead of an entire email."*

8.3.2 Context-Driven Preferences: Users highlighted the trade-offs between device speed and convenience (laptops vs. mobile devices) and how email length affects time-saving benefits. P1 acknowledges laptops as faster for email composition but points out their cumbersomeness in mobile settings. The need to physically set up a laptop ("take them out and probably stop somewhere") is seen as a disadvantage compared to the more immediate access provided by mobile devices. This observation highlights the trade-off between the efficiency of typing and processing speed offered by laptops and the convenience and portability of mobile devices. In scenarios requiring mobility or when space and time are limited, the ease of pulling out a phone and quickly typing an email outweighs the speed advantages of a laptop.

P12 contrasts the time-saving benefits of using the system for composing longer emails against the inefficiencies it introduces for shorter emails. For lengthy emails, especially those of a formal or informal nature that require more thought and organization, the system could streamline the writing process, making it quicker than traditional methods. This suggests that the AI's assistance is more valuable when dealing with complex content that would typically take a long time to compose, possibly due to its ability to generate content, organize thoughts, or even correct grammar. Conversely, for brief emails, the time taken to interact with the AI (possibly including command inputs, corrections, or navigating the AI's interface) may exceed the time it would take to simply type out a short message directly. This indicates a limitation in the efficiency of using AI for all types of email tasks, particularly when the simplicity of the task does not warrant the overhead of AI interaction.

8.3.3 Global Editing Challenges: AI-assisted email editing encounters tone and brevity discrepancies, necessitating extensive user revisions for optimal outcomes. Global editing involves adjusting the tone, writing structure, and length of emails. However, participants often encountered difficulties due to discrepancies between their intended tone and length and the AI's understanding. One issue is tone discrepancy. As per P1, P4, and P5, the *"AI often fails to capture a friendly or casual tone, defaulting to a more formal approach"*. This mismatch necessitates substantial edits to make the content sound less awkward, especially in informal settings. P1 thought *"AI is better with a formal tone but struggles with an informal tone because each person has their own way of writing it. So it's very hard for the AI to catch it, especially on the first try."* P11 noted, *"AI is hard to distinguish friends and brothers/sisters. In my opinion, they are not in a familiar relationship."*. Another issue is the difficulty in altering the global features, be it language alone or dialogue, which can be challenging due to imprecision in natural human speech and the mismatch in intention captured by LLM. Participants like P2 and P6 experience challenges *"in capturing the right tone in short emails or when trying to spell names correctly"*. The AI tends to *"add unnecessary details, making it hard to maintain the desired brevity."* Finally, despite participants instructing the AI to make the content casual or friendly, generated content often remains overly formal and wordy. When participants aim for concise emails, *"the AI tends to produce longer responses than desired, even when instructed to be casual or friendly."* as mentioned by P8 and P10. This mismatch requires participants to invest extra effort in editing, either by deleting or adding content. Therefore, participants need easy-to-use global adjustment interactions that can more accurately align with their intended tone and content length.

8.3.4 Local Editing Challenges: Voice-based editing with AI assistants can be imprecise, demanding detailed instructions for successful modifications. Editing emails through voice interaction with AI assistants presents significant challenges, particularly in fine-editing tasks such as precise placement, sentence-level modifications, word-level control, contextual understanding, and word correction (e.g., names, and places). One issue is AI's misplacement of new content. AI assistants struggle to accurately identify the email's structure, often leading to the misplacement of new content. *"AI made it very difficult to edit emails as it would not understand the structure of an email like body, header, ending"*, noted P2, P4 and P11 also mentioned that *"AI does not understand the correct position for adding new content, often leading to inaccuracies like adding sentences after the signature."* To improve editing accuracy, P1 attempted to give more specific prompts like direct commands "1st sentence, last paragraph" rather than natural language instructions. However, *"there were still some errors which forced me to edit by saying the whole sentence, which was cumbersome. I would rather have this more precise one because I think the more precise you can make it, the easier the editing process would be."*. Another issue is adjusting sentence-level order, especially within paragraphs, becomes a time-consuming process. Participants like P5 and P7 expressed a need for *"enhanced control over word choice and arrangement"*, and P6 mentioned *"improved contextual comprehension to accurately incorporate names and context"*. Challenges also arise when merging separate sentences into coherent paragraphs, as P3 said that *"AI frequently splits smell details into new paragraphs results in unnecessary line breaks"*. Participants envision a system that can deduce their editing requirements based on descriptions, allowing for a more efficient editing process with reduced workload.

9 DISCUSSION

We initiated our exploration into building LLM-based wearable assistants to handle personal information tasks on the go, enabling seamless collaboration and the creation of complex textual content while in motion. Our iterative design process revealed that simply having an intelligent agent is insufficient; user interaction design plays a critical role in ensuring a seamless computing experience. Drawing from our design process and studies, we present the following lessons to guide the development of future AI-enabled wearable assistants for complex information processing tasks.

9.1 Design Implications for Extending Two Proposed Approaches

9.1.1 "Single Turn with Optional Clarification" Design: Allows user clarifications on AI's interpretations via word-level chunking, reducing workload and suiting fragmented attention. *GlassMail* uses a "Single Turn with Optional Clarification" approach to facilitate user-agent interactions. This method acknowledges that the amount of information users provide in their initial turn can vary based on the task or user behaviour, and they may unintentionally miss details. Instead of using fixed templates like the dialogue approach which can be cognitively demanding, this approach starts with a general prompt to collect as much information as possible for the task. It then utilises "Fragmented Attention-Friendly Chunking" to display the agent's understanding to the user. This transparency serves as feedback, empowering the user, allowing them the freedom to express and initiate, and allowing users to clarify, correct errors, or provide additional information relaxedly.

The utility of this approach extends beyond OHMD mobile scenarios and applies to all contexts involving AI interactions. This strategy allows users to clarify the AI's understanding, thereby enhancing the accuracy and overall experience of the interaction. The use of word-level chunking to display the AI's interpretations is especially user-friendly in situations where the user has fragmented attention. It reduces the comprehension workload and guides users to easily provide spoken instructions for clarifications. Given that our final testing is lab-controlled indoor simple walking and did not include outdoor scenarios, where the environment is typically more complex, there remains a possibility that the Single Turn with Optional Clarification approach might still fall short in accuracy if the user's initial input is too vague or incomplete and lacks crucial details.

9.1.2 "Fade Context with Optional Audio" Design: Minimizes overload in mobile email editing, keeping essential context. Supports optional audio-visual output for easier editing in demanding visual tasks. Whether to display contextual information in voice-based text editing depends on the nature of the text. For texts that require a well-structured format or careful wording, such as emails or formal documents, displaying contextual information remains important for the editing process. Specifically, using an inconspicuous colour like grey to fade the context display could facilitate easier visual search, enhance multitasking, and reduce cognitive workload. As for the output modality, whether the simultaneous hybrid of visual with audio output offers an advantage over visual-only output depends on the text presentation and mobile tasks. With consistent visual outputs, optional audio support tends to be more beneficial during visually demanding tasks.

9.2 LLM-based Personalization

Users report that *GlassMail* has the potential to reduce a significant portion (i.e., approximately 70-80%) of the effort involved in manual editing and grammar checking when composing emails. However, the extent of this reduction may depend on the type of email being composed. During the observation study, *GlassMail* was found to require more post-editing effort for informal emails compared to formal emails. While *GlassMail* can accurately capture users' intentions for drafting emails with a casual tone using the Single Turn with Optional Clarification, there is a noticeable gap between the system's interpretation of casual language and users' actual language usage, which affects the quality of the generated content. This discrepancy often leads to increased conversational interactions with the LLMs for adjustments, resulting in higher time and effort costs during the post-editing process. Moreover, the style, tone, and words generated by *GlassMail* are frequently described by participants as robotic and unnatural, necessitating additional post-editing efforts to achieve the desired email content quality. In the future, integrating an understanding of users' personality traits and incorporating emotion recognition techniques could improve the system's performance. For example, we could explore the possibility of allowing users to control the emotional tone of their writing through facial expressions, utilizing smart-glass-based emotion recognition technology [72].

GlassMail faces challenges in efficient post-editing to achieve final personalization. While LLMs can proficiently create quality drafts from brief user descriptions and restructure drafts when additional details need to be added, our studies indicate that achieving precise editing using LLMs is still challenging. Users occasionally found themselves in iterative cycles of adding and removing content, as alterations frequently affect more than the intended segment of text. For instance, when a user only wants to inquire about "Lily's dinner time preferences," LLM may include unrelated details, such as "tell Lily I like the Italian restaurant located on XXX Street." Users then need to ask the LLM to remove these unrelated details, but the LLM may remove other unintended details. This creates an unpredictable and iterative editing process. When users do not articulate changes as a command, the LLM sometimes fails to act, indicating a deficiency in recognizing and interpreting the nuances in the way humans expect a personal assistant to help them. Moreover, the post-editing process of *GlassMail* lacks such a mechanism to confirm whether LLMs' understandings are aligned with users' editing instructions. This leads to challenges in correcting misunderstandings and necessitates higher editing efforts.

To address these challenges *GlassMail* could implement better post-editing schemes and personalization through learning to facilitate more complex email composition tasks. For instance, *GlassMail* could utilise a single sample of a user's email through few-shot learning [36, 77] to generate more personalized content, which could further reduce the post-editing efforts. The approach could centre on two fundamental elements identified in our observation study: tone (e.g., greetings, openings, closings, signatures) and the individual's email writing structure. By instructing the LLM to analyze these facets in a sample email, *GlassMail* could establish the context for drafting new emails, effectively retaining users' individual writing preferences. We anticipate that the incorporation of continual learning approaches could further enhance the preservation of personal writing styles [66].

9.3 Safety, Privacy, Accessibility and Ethical Considerations

Safety, privacy, accessibility and ethical considerations during mobile multitasking with OHMD AI-based systems are paramount. Future enhancements of *GlassMail* should focus on these critical aspects. Safety issues involve cognitive aspects and physical risks [28, 65]. Users writing emails on OHMDs while walking may be less aware of their surroundings, thus increasing the likelihood of accidents [11, 27], and further consideration needs to be given to safety-focused system design and rigorous large-scale evaluations of real-world usability if the system is to be widely used.

To provide a more natural experience in mobile scenarios, *GlassMail* utilizes voice as an input modality, allowing users to compose emails hands-free. However, privacy concerns often arise when using voice input for email drafting, particularly when dealing with personal content. In such cases, silent speech recognition may serve as a viable alternative to traditional voice input methods. Silent speech recognition technology enables users to communicate without vocalizing, reducing the risk of eavesdropping and enhancing privacy [15, 75]. Additionally, given the sensitive nature of email communications, the security and privacy of user data are of utmost importance [67]. Future versions of *GlassMail* should incorporate robust encryption protocols and clear user consent mechanisms, adhering to data protection regulations and maintaining user trust.

GlassMail possesses the potential to improve accessibility for users with motor impairments. Currently, our voice input method contributes to accessibility, and in the future, we could further enhance it by integrating voice input and gaze-based interactions [71]. This integration may enable individuals with limited hand mobility to compose emails more easily and potentially reduce cognitive load [69, 70]. Future research should explore the specific needs and preferences of this user group to optimize the system's accessibility features [13, 59].

Finally, assessing the moral responsibility of AI in crafting email content and understanding the impact of its suggestions on users' personal and professional relationships is crucial [73]. Implementing ethical guidelines and providing users with override options can ensure generative AI usage while empowering users.

9.4 Limitations

As a proof-of-concept prototype, *GlassMail* was designed for simple personal email creation in the OHMD mobile scenario. Responding to a single email or email threads that require users to contemplate and catch up on the context was beyond our study scope. Also, *GlassMail* does not support the creation of all types of emails, including those that require adding bullet points, attachments and links. Additionally, it does not support users in customizing audio speed and content for personalization functions.

Among the real-world challenges faced, a limitation is the current inadequacy of our speech recognition features in accurately extracting a user's voice in real-time scenarios. This limitation becomes particularly pronounced in environments with background noise, such as conversations or transport announcements. Additionally, the system's reliance on sending requests for real-time recognition at a frequency of once per second can lead to delays attributable to varying network conditions. These factors cumulatively contribute to elevated error rates in user interactions, potentially affecting the user's experience and trust in the system, especially when voice input is the sole mode of interaction. Currently, the success of our system in high ambient noise environments depends on the noise reduction capability of the microphones used. Integrating ambient noise cancellation into *GlassMail*'s software will provide more reliable and consistent performance across users and environments.

While our in-lab observational study provided valuable insights into the feasibility of *GlassMail* and the challenges for editing interactions, the controlled environment of the in-lab study may not fully capture the complexities and variability of real-world mobile contexts. The findings from this study may not be entirely generalizable to real-world situations. Future research should build upon the findings of this in-lab observational study and consider conducting an in-situ exploration to evaluate *GlassMail*'s effectiveness in real-world mobile contexts. This would provide a more comprehensive understanding of the system's potential to mitigate the challenges of mobile email composition and reduce post-editing efforts. Future research should explore *GlassMail*'s long-term effects on user behaviour, communication, and productivity to identify changes and evaluate its impact on email composition habits and efficiency.

10 CONCLUSION

This study serves as an initial step towards the ultimate goal of building AI-enabled wearable assistants for handling complex information tasks on the go. We conducted a formative study to understand the potential and viability of the LLM-based wearable email assistant and the challenges faced: (i) achieving efficient and accurate understanding of user intentions, and (ii) ensuring effective information presentation for email processes. Through two empirical studies, we developed *GlassMail* features a "Single Turn with Optional Clarification" approach for accurate user intention recognition and a "Fade Context with Optional Audio" mode for effective email processing. An observation study then evaluated *GlassMail*'s feasibility in composing formal and semi-formal emails, supporting the usefulness and effectiveness of *GlassMail* in simple scenarios and yielding insights into potential future improvements for complex email scenarios. We then provide design implications for future wearable AI-enabled assistants.

While we have demonstrated the potential of *GlassMail* as a viable solution for composing emails through wearable devices, we acknowledge that this work represents just one small step towards the ultimate larger goal. Future iterations of *GlassMail* will focus on developing dynamic and context-aware interfaces to better understand and address users' diverse needs and situations, as well as exploring additional features for enhanced email management. We hope that our research will contribute to the realization of the vision of heads-up computing of wearable AI-powered assistants [76], ultimately empowering users to handle complex information tasks efficiently and effectively, even when on the go.

ACKNOWLEDGMENTS

This research is supported by the National Research Foundation, Singapore, under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-016). It is also partially supported by City University of Hong Kong Academic Fund (Project Number 9610677). We extend our gratitude to the Synteraction Lab members for their valuable feedback at different stages of our research.

REFERENCES

[1] 2024. Flowrite. https://www.flowrite.com/. Accessed: 2024-02-07.

[2] Andreja Andric and Goffredo Haus. 2006. Automatic playlist generation based on tracking user's listening habits. *Multimedia Tools and Applications* 29 (2006), 127–151.

[3] Markus Appel, Nina Krisch, Jan-Philipp Stein, and Silvana Weber. 2019. Smartphone zombies! Pedestrians' distracted walking as a function of their fear of missing out. *Journal of Environmental Psychology* 63 (2019), 130–133.

[4] Mandy M Archibald, Rachel C Ambagtsheer, Mavourneen G Casey, and Michael Lawless. 2019. Using zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. *International journal of qualitative methods* 18 (2019), 1609406919874596.

[5] 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*. 1–8.

[6] Dmitry Bogdanov et al. 2013. From music similarity to music recommendation: Computational approaches based on audio features and metadata. (2013).

[7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[8] Peter Burggräf, Moritz Beyer, Jan-Philip Ganser, Tobias Adlon, Katharina Müller, Constantin Riess, Kaspar Zollner, Till Saßmannshausen, and Vincent Kammerer. 2022. Preferences for Single-Turn vs. Multiturn Voice Dialogs in Automotive Use Cases—Results of an Interactive User Survey in Germany. *IEEE Access* 10 (2022), 55020–55033.

[9] Daniel Buschek, Martin Zürn, and Malin Eiband. 2021. The impact of multiple parallel phrase suggestions on email input and composition behaviour of native and non-native english writers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

[10] Zefan Cai, Baobao Chang, and Wenjuan Han. 2023. Human-in-the-Loop through Chain-of-Thought. *arXiv preprint arXiv:2306.07932* (2023).

[11] Jeff K Caird, Kate A Johnston, Chelsea R Willness, Mark Asbridge, and Piers Steel. 2014. A meta-analysis of the effects of texting on driving. *Accident Analysis & Prevention* 71 (2014), 311–318.

[12] Xiao Chen, Wanli Chen, Kui Liu, Chunyang Chen, and Li Li. 2021. A comparative study of smartphone and smartwatch apps. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*. 1484–1493.

[13] Eric Corbett and Astrid Weber. 2016. What can I say? addressing user experience challenges of a mobile voice user interface for accessibility. In *Proceedings of the 18th international conference on human-computer interaction with mobile devices and services*. 72–82.

[14] Saverio Debernardis, Michele Fiorentino, Michele Gattullo, Giuseppe Monno, and Antonio Emmanuele Uva. 2014. Text Readability in Head-Worn Displays: Color and Style Optimization in Video versus Optical See-Through Devices. *IEEE Transactions on Visualization and Computer Graphics* 20, 1 (2014), 125–139. https://doi.org/10.1109/TVCG.2013.86

[15] Bruce Denby, Tanja Schultz, Kiyoshi Honda, Thomas Hueber, Jim M Gilbert, and Jonathan S Brumberg. 2010. Silent speech interfaces. *Speech Communication* 52, 4 (2010), 270–287.

[16] Jiayue Fan, Chenning Xu, Chun Yu, and Yuanchun Shi. 2021. Just speak it: Minimize cognitive load for eyes-free text editing with a smart voice assistant. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 910–921.

[17] Raymond Fok and Daniel S Weld. 2023. What Can't Large Language Models Do? The Future of AI-Assisted Academic Writing. In *In2Writing Workshop at CHI*.

[18] Joseph L. Gabbard, J. Edward Swan, and Deborah Hix. 2006. The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence* 15, 1 (2006), 16–32. https://doi.org/10.1162/pres.2006.15.1.16

[19] Joseph L. Gabbard, J. Edward Swan, Deborah Hix, Si-Jung Kim, and Greg Fitch. 2007. Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications. In *2007 IEEE Virtual Reality Conference*. 35–42. https://doi.org/10.1109/VR.2007.352461

[20] Maliheh Ghajargar, Jeffrey Bardzell, and Love Lagerkvist. 2022. A redhead walks into a bar: experiences of writing fiction with artificial intelligence. In *Proceedings of the 25th international academic MindTrek conference*. 230–241.

[21] 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*. 1–10.

[22] Debjyoti Ghosh, Pin Sym Foong, Shengdong Zhao, Can Liu, Nuwan Janaka, and Vinitha Erusu. 2020. Eyeditor: Towards on-the-go heads-up text editing using voice and manual input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[23] Debjyoti Ghosh, Can Liu, Shengdong Zhao, and Kotaro Hara. 2020. Commanding and re-dictation: Developing eyes-free voice-based interaction for editing dictated text. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 4 (2020), 1–31.

[24] Steven M Goodman, Erin Buehler, Patrick Clary, Andy Coenen, Aaron Donsbach, Tiffanie N Horne, Michal Lahav, Robert MacDonald, Rain Breaw Michaels, Ajit Narayanan, et al. 2022. Lampost: Design and evaluation of an ai-assisted email writing prototype for adults with dyslexia. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–18.

[25] Zheng Haolan, Isabella M Campbell, and Wayne CW Giang*. 2021. Phone-related distracted walking injuries as a function of age and walking environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 65. SAGE Publications Sage CA: Los Angeles, CA, 611–615.

[26] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[27] Rami Hashish, Megan E Toney-Bolger, Sarah S Sharpe, Benjamin D Lester, and Adam Mulliken. 2017. Texting during stair negotiation and implications for fall risk. *Gait & posture* 58 (2017), 409–414.

[28] Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2021. Towards indistinguishable augmented reality: A survey on optical see-through head-mounted displays. *ACM Computing Surveys (CSUR)* 54, 6 (2021), 1–36.

[29] Nuwan Janaka, Jie Gao, Lin Zhu, Shengdong Zhao, Lan Lyu, Peisen Xu, Maximilian Nabokow, Silang Wang, and Yanch Ong. 2023. GlassMessaging: Towards Ubiquitous Messaging Using OHMDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (2023), 1–32.

[30] Sangeun Jin, Minsung Kim, Jihyeon Park, Minsung Jang, Kyuseok Chang, and Daemin Kim. 2019. A comparison of biomechanical workload between smartphone and smartwatch while sitting and standing. *Applied ergonomics* 76 (2019), 105–112.

[31] Ankur Joshi, Saket Kale, Satish Chandel, and D Kumar Pal. 2015. Likert scale: Explored and explained. *British journal of applied science & technology* 7, 4 (2015), 396–403.

[32] Marijke Keus van de Poll and Patrik Sörqvist. 2016. Effects of task interruption and background speech on word processed writing. *Applied cognitive psychology* 30, 3 (2016), 430–439.

[33] Jeongyeon Kim, Sangho Suh, Lydia B Chilton, and Haijun Xia. 2023. Metaphorian: Leveraging Large Language Models to Support Extended Metaphor Creation for Science Writing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 115–135.

[34] Per Ola Kristensson. 2007. *Discrete and continuous shape writing for text entry and control*. Ph. D. Dissertation. Institutionen för datavetenskap.

[35] Mina Lee, Percy Liang, and Qian Yang. 2022. Coauthor: Designing a human-ai collaborative writing dataset for exploring language model capabilities. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–19.

[36] Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. 2023. GPT understands, too. *AI Open* (2023).

[37] I Scott MacKenzie and R William Soukoreff. 2002. Text entry for mobile computing: Models and methods, theory and practice. *Human–Computer Interaction* 17, 2-3 (2002), 147–198.

[38] I Scott MacKenzie and Kumiko Tanaka-Ishii. 2010. *Text entry systems: Mobility, accessibility, universality*. Elsevier.

[39] Yuki Matsuura, Tsutomu Terada, Tomohiro Aoki, Susumu Sonoda, Naoya Isoyama, and Masahiko Tsukamoto. 2019. Readability and Legibility of Fonts Considering Shakiness of Head Mounted Displays. In *Proceedings of the 23rd International Symposium on Wearable Computers* (London, United Kingdom) *(ISWC '19)*. Association for Computing Machinery, New York, NY, USA, 150–159. https://doi.org/10.1145/3341163.3347748

[40] Piotr Mirowski, Kory W Mathewson, Jaylen Pittman, and Richard Evans. 2023. Co-Writing Screenplays and Theatre Scripts with Language Models: Evaluation by Industry Professionals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–34.

[41] Terhi Mustonen, Mikko Berg, Jyrki Kaistinen, Takashi Kawai, and Jukka Häkkinen. 2013. Visual Task Performance Using a Monocular See-Through Head-Mounted Display (HMD) While Walking. *Journal of experimental psychology. Applied* 19 (11 2013). https://doi.org/10.1037/a0034635

[42] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–7.

[43] Michael Nebeling, Alexandra To, Anhong Guo, Adrian A de Freitas, Jaime Teevan, Steven P Dow, and Jeffrey P Bigham. 2016. WearWrite: Crowd-assisted writing from smartwatches. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 3834–3846.

[44] R OpenAI. 2023. GPT-4 technical report. *arXiv* (2023), 2303–08774.

[45] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2013. Dynamic Text Management for See-through Wearable and Heads-up Display Systems. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces* (Santa Monica, California, USA) *(IUI '13)*. Association for Computing Machinery, New York, NY, USA, 363–370. https://doi.org/10.1145/2449396.2449443

[46] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2014. Managing Mobile Text in Head Mounted Displays: Studies on Visual Preference and Text Placement. *SIGMOBILE Mob. Comput. Commun. Rev.* 18, 2 (jun 2014), 20–31. https://doi.org/10.1145/2636242.2636246

[47] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. 2005. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 919–928.

[48] Harold Pashler, Sean HK Kang, and Renita Y Ip. 2013. Does multitasking impair studying? Depends on timing. *Applied Cognitive Psychology* 27, 5 (2013), 593–599.

[49] Paulo HS Pelicioni, Lloyd LY Chan, Shuotong Shi, Kenny Wong, Lauren Kark, Yoshiro Okubo, and Matthew A Brodie. 2023. Impact of mobile phone use on accidental falls risk in young adult pedestrians. *Heliyon* 9, 8 (2023).

[50] Dongqi Pu and Vera Demberg. 2023. ChatGPT vs Human-authored Text: Insights into Controllable Text Summarization and Sentence Style Transfer. *arXiv preprint arXiv:2306.07799* (2023).

[51] Ashwin Ram and Shengdong Zhao. 2021. LSVP: Towards Effective On-the-go Video Learning Using Optical Head-Mounted Displays. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–27.

[52] Ashwin Ram and Shengdong Zhao. 2022. Does Dynamically Drawn Text Improve Learning? Investigating the Effect of Text Presentation Styles in Video Learning. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 89, 12 pages. https://doi.org/10.1145/3491102.3517499

[53] Ronald E Robertson, Alexandra Olteanu, Fernando Diaz, Milad Shokouhi, and Peter Bailey. 2021. "I can't reply with that": Characterizing problematic email reply suggestions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[54] 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*. 1–9.

[55] Daniel Salber and Joëlle Coutaz. 1993. Applying the wizard of oz technique to the study of multimodal systems. In *Human-Computer Interaction: Third International Conference, EWHCI'93 Moscow, Russia, August 3–7, 1993 Selected Papers 3*. Springer, 219–230.

[56] Shardul Sapkota, Ashwin Ram, and Shengdong Zhao. 2021. Ubiquitous Interactions for Heads-Up Computing: Understanding Users' Preferences for Subtle Interaction Techniques in Everyday Settings. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*. 1–15.

[57] Saiganesh Swaminathan, Raymond Fok, Fanglin Chen, Ting-Hao Huang, Irene Lin, Rohan Jadvani, Walter S Lasecki, and Jeffrey P Bigham. 2017. Wearmail: On-the-go access to information in your email with a privacy-preserving human computation workflow. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 807–815.

[58] David R Thomas. 2003. A general inductive approach for qualitative data analysis. (2003).

[59] Markku Turunen, Jaakko Hakulinen, K-J Raiha, E-P Salonen, Anssi Kainulainen, and Perttu Prusi. 2005. An architecture and applications for speech-based accessibility systems. *IBM Systems Journal* 44, 3 (2005), 485–504.

[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] Zheng Wang, Prabu David, Jatin Srivastava, Stacie Powers, Christine Brady, Jonathan D'Angelo, and Jennifer Moreland. 2012. Behavioral performance and visual attention in communication multitasking: A comparison between instant messaging and online voice chat. *Computers in Human Behavior* 28, 3 (2012), 968–975.

[62] Wouter Weerkamp, Krisztian Balog, and Maarten De Rijke. 2009. Using contextual information to improve search in email archives. In *Advances in Information*

Retrieval: 31th European Conference on IR Research, ECIR 2009, Toulouse, France, April 6-9, 2009. Proceedings 31. Springer, 400–411.

[63] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems* 35 (2022), 24824–24837.

[64] Daryl Weir, Henning Pohl, Simon Rogers, Keith Vertanen, and Per Ola Kristensson. 2014. Uncertain text entry on mobile devices. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 2307–2316.

[65] Christopher D Wickens. 2017. Mental workload: assessment, prediction and consequences. In *Human Mental Workload: Models and Applications: First International Symposium, H-WORKLOAD 2017, Dublin, Ireland, June 28-30, 2017, Revised Selected Papers 1.* Springer, 18–29.

[66] Tongtong Wu, Linhao Luo, Yuan-Fang Li, Shirui Pan, Thuy-Trang Vu, and Gholamreza Haffari. 2024. Continual Learning for Large Language Models: A Survey. arXiv:2402.01364 [cs.CL]

[67] Xiaodong Wu, Ran Duan, and Jianbing Ni. 2023. Unveiling security, privacy, and ethical concerns of chatgpt. *Journal of Information and Intelligence* (2023).

[68] Yujia Xie, Xun Wang, Si-Qing Chen, Wayne Xiong, and Pengcheng He. 2023. Interactive Editing for Text Summarization. *arXiv preprint arXiv:2306.03067* (2023).

[69] Zihan Yan, Yifei Shan, Yiyang Li, Kailin Yin, and Xiangdong Li. 2021. Gender Differences of Cognitive Loads in Augmented Reality-based Warehouse. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* 500–501. https://doi.org/10.1109/VRW52623.2021.00132

[70] Zihan Yan, Yufei Wu, Yiyang Li, Yifei Shan, Xiangdong Li, and Preben Hansen. 2022. Design Eye-Tracking Augmented Reality Headset to Reduce Cognitive Load in Repetitive Parcel Scanning Task. *IEEE Transactions on Human-Machine Systems* 52, 4 (2022), 578–590. https://doi.org/10.1109/THMS.2022.3179954

[71] Zihan Yan, Yue Wu, Yifei Shan, Wenqian Chen, and Xiangdong Li. 2023. A dataset of eye gaze images for calibration-free eye tracking augmented reality headset (vol 9, 115, 2022). *SCIENTIFIC DATA* 10, 1 (2023).

[72] Zihan Yan, Yufei Wu, Yang Zhang, and Xiang'Anthony' Chen. 2022. EmoGlass: An end-to-end AI-enabled wearable platform for enhancing self-awareness of emotional health. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* 1–19.

[73] Yifan Yao, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Eric Sun, and Yue Zhang. 2023. A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. *arXiv preprint arXiv:2312.02003* (2023).

[74] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: story writing with large language models. In *27th International Conference on Intelligent User Interfaces.* 841–852.

[75] Ruidong Zhang, Mingyang Chen, Benjamin Steeper, Yaxuan Li, Zihan Yan, Yizhuo Chen, Songyun Tao, Tuochao Chen, Hyunchul Lim, and Cheng Zhang. 2022. SpeeChin: A Smart Necklace for Silent Speech Recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 192 (dec 2022), 23 pages. https://doi.org/10.1145/3494987

[76] Shengdong Zhao, Felicia Tan, and Katherine Fennedy. 2023. Heads-Up Computing: Moving Beyond the Device-Centered Paradigm. *Commun. ACM* 66 (9 2023), 56–63. https://doi.org/10.1145/3571722

[77] Zihao Zhao, Eric Wallace, Shi Feng, Dan Klein, and Sameer Singh. 2021. Calibrate before use: Improving few-shot performance of language models. In *International Conference on Machine Learning.* PMLR, 12697–12706.

[78] Chen Zhou, Katherine Fennedy, Felicia Fang-Yi Tan, Shengdong Zhao, and Yurui Shao. 2023. Not All Spacings are Created Equal: The Effect of Text Spacings in On-the-go Reading Using Optical See-Through Head-Mounted Displays. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–19.

A APPENDIX

A.1 JSON Schemas for Prompt Response

We used JSON response schema for all prompts, as it eases response parsing by eliminating the need for custom functions and reduces parsing errors.

```
{
    Tone: "Formal",
    Length: "Short",
    Subject: "Discuss User Feedback",
    Relation: "Project member",
    FromName: "Liz",
    Recipient: "Kayla"
}
```

A.2 Prompts for LLMs

Prompts were crafted for *GlassMail*'s scalability, focusing on 1) Extracting users' intentions for composing emails, including key aspects such as topic, tone, length, subject, relation, sender's name (i.e., fromName), and recipient. 2) Generate emails based on the users' intention. 3) Regenerate emails.

A.2.1 Prompts for extracting user's intentions for composing emails. The "extractEmailSettings" function extracts the user's intentions for composing emails and returns these key aspects in the settings using a JSON format.

```
export const extractEmailSettings = async(prompt,
    handleResponse) => {
    const gptMsgs = [{
            role: "system",
            content: "You are a writing expert who
     assists users in the process of composing and
     editing emails. You should strictly follow the
     requirements and output specifications provided by
     the user.",
        },
        {
            role: "user",
            content: "Given the <UserInstrcutions>:
     ${prompt}$, you should extract the following key
     aspects: Recipient, Subject, Relation, FromName,
     Tone and Length. For any aspect you cannot extract
     from <UserInstrcutions>, please infer or predict
     based on the provided <UserInstrcutions>. If you
     are unable to predict or infer it, please return
     'N.A'.
            The output should be a JSON format:
            {
                settings: {
                    recipient: "Recipient's Name",
                    subject: "Suggest one if not
     provided in <UserInstrcutions>, within 5 words",
                    relation: "If not provided, suggest
     the most suitable one based on <UserInstrcutions>,
     using one word",
                    tone: "If not provided, suggest the
     most suitable one based on <UserInstrcutions>,
     using one word",
                    length: "Short(within 200 words),
     Medium(200-250 words), or Detailed (350-500 words)",
                    fromName: "Sender's Name",
                }
            }.",
        }
    }]
}
```

A.2.2 Prompts for generating emails. The "generateEmail" function generates emails based on the email settings and returns emails using a JSON format.

```
export const generateEmail = async (emailSettings,
    prompt, handelResponse) => {
   const gptMsgs = [{
            role: "user",
            content: "Given the <UserInstrcutions>:
     ${prompt}$ and
     <EmailSettings>:${JSON.stringify(emailSettings)}$,
     you should first update the <EmailSettings>
     according to the latest <UserInstrcutions>. Then
     compose an email strictly following the
     <UserInstrcutions> and <EmailSettings>.
            The output should be a JSON format:
            {
                subject: "Suggest one if not provided
     the latest <EmailSettings>, within 5 words",
                content: "Each paragraph should be
     separated by \n\n",
                settings: "The latest <EmailSettings>",
            }.",
        }
    }]
}
```

A.2.3 *Prompts for editing emails.*

The "editEmail" function regenerates the email and returns a new email using a JSON format.

```
export const editEmail = async ({emailSettings, email} =
     preOutput, prompt, handelResponse) => {
   const gptMsgs = [{
            role: "user",
            content: "Given the <Email>: ${email}$ and
     <EmailSettings>:${JSON.stringify(emailSettings)}$,
     you should first update the <EmailSettings>
     according to the <UserInstruction>:${prompt}$. Then
     you should only edit the related content according
     to the <UserInstruction> and return the edited
     <Email>.
            The output should be a JSON format:
            {
                subject: "",
                content: "Each paragraph should be
     separated by \n\n",
                settings: "The latest <EmailSettings>",
            }.",
        }
    }]
}
```

A.3 Email Tasks Used in Study 1

Table 2: Email Tasks Used in Study 1

Social Ties	Scenarios 1	Scenarios 2	Scenarios 3
Supervisor	Kim is your supervisor. Your team is working on a project, and there's a milestone review meeting next week. You need Kim's feedback on the project before the meeting.	Kim is your supervisor. You recently attended a conference and learned about a new approach that could improve a current project. You want to share this information with Kim.	Kim is your supervisor. The team is facing challenges in meeting a project deadline. You want to suggest a meeting to discuss strategies to overcome the obstacles.
Sister	Lily is your sister. You both share a love for cooking. There's a cooking class next weekend that you think she might be interested in joining.	Lily is your sister. Your family is planning a surprise birthday party for your brother Tony. You want to coordinate with Lily to ensure a successful event.	Lily is your sister. You recently watched a movie that she might enjoy. You want to recommend the movie to her and suggest watching it together.
Friend	Lucy is your friend. You recently attended a concert and bought an extra ticket for the next concert by the same artist. You want to offer her the ticket.	Lucy is your friend. You are both interested in a new book that's launching. You want to suggest a book club meeting to discuss the book.	Lucy is your friend. You're organizing a surprise farewell party for a friend Sam who's moving away. You want to coordinate the party details with Lucy.