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

ANONYMOUS AUTHOR(S)

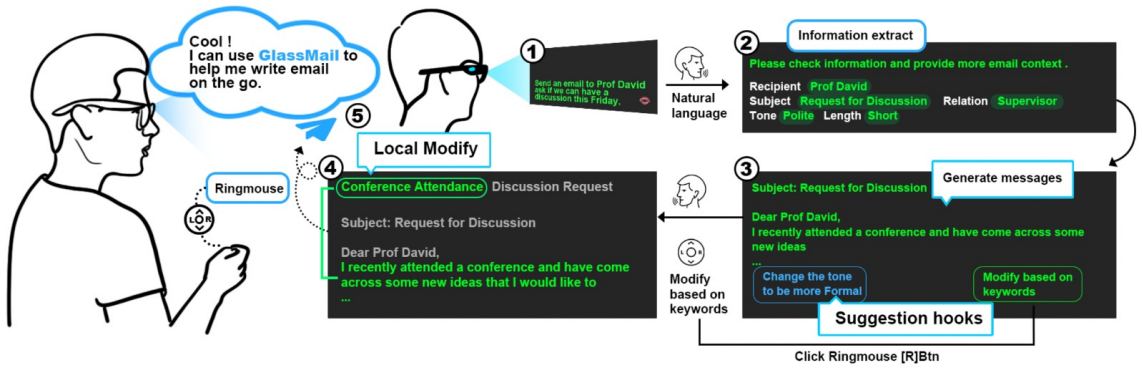


Fig. 1. GlassMail interactions: 1) GlassMail is able to understand the context and extract the user's requirements. 2) Then GlassMail will confirm the critical requirements with the user. 3) Once the user confirms, GlassMail will generate an email and suggest editing hooks. 4)In the local edit mode, the users can further customize the email from macro-level (e.g. structure) to micro-level (e.g. wording).

Managing personal information tasks such as composing professional emails while on the go can be quite challenging. These challenges primarily arise from the mobile devices' limited I/O capabilities and multitasking constraints. We introduce GlassMail, a wearable digital assistant leveraging Large Language Models (LLMs) on Optical See-through Head-Mounted Displays for mobile personal email creation. Our initial formative study explored the challenges faced by users when using smartphones and collaborating with LLM-based assistants for mobile email creation. Our findings highlight the delicate balance between personalization and workload management needed to effectively collaborate with an LLM-based assistant. Through a pilot study and two controlled experiments, we developed the GlassMail interface and evaluated the interface's usability in realistic mobile scenarios. Results showed that GlassMail supports seamless human-AI collaboration, easy information absorption, and personalization. We further discuss insights and implications for the future design of intelligent wearable Personal Information Management (PIM) systems.

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

Additional Key Words and Phrases: OHMD, Smart Glasses, Large Language Model, Personal Information Management, PIM, Voice Assistant, Mobile email creation

Manuscript submitted to ACM

ACM Reference Format:
Anonymous Author(s). 2018. GlassMail: Towards Personalised Wearable Assistant for On-the-go Email Creation on Smart Glasses. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 35 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In the world of fashion portrayed in the movie "The Devil Wears Prada"[1], a critical message needs to be conveyed from Miranda to her junior assistant, Emily. Through a brief dialogue, Emily grasps the important messages and aids Miranda in completing it. This cinematic portrayal underscores the potential of intelligent and well-trained personal assistants, capable of assisting in intricate tasks with just a few instructions. They are particularly useful in on-the-go situations, where performing such tasks independently would compromise the primary mission of their masters. While this may seem like a narrative confined to the silver screen, our formative studies reveal that such occurrences are commonplace in real life. However, this level of personal assistance remains a luxury not accessible to all.

For many, crafting a relatively complex email in on-the-go scenarios can be a significant challenge [39, 98]. Often, individuals choose to postpone such tasks, waiting for more favourable conditions, like a quiet environment, a desk, and the requisite equipment, such as a laptop. Alternatively, when desperate, they reluctantly transform into "smartphone zombies," [8] dedicating their attention to the small screen and the tiny keyboard to laboriously compose an email. Yet, this experience remains tedious, error-prone, and unsatisfying [60].

The emergence of Large Language Models (LLM) has revolutionized text creation, rendering it remarkably simpler. Users can now succinctly express their intentions in a text field within a chatbot, and in return, receive a complete email draft of high quality. Given this new "magical" capability offered by LLM, one could be inclined to believe that bringing an intelligent mobile, personal email assistant to life could be as straightforward as integrating LLM into a mobile or wearable device.

However, the reality is far from simple. Through a formative study ($n = 12$), we have discovered that merely introducing LLM to wearable devices (smart glasses or Optical Head-Mounted Display henceforth, just OHMD) does not automatically transform them into effective on-the-go personal email assistants. Numerous challenges, ranging from input/output interactions to personalisation, stand in the way, rendering the concept of a seamless, efficient on-the-go email assistant an ambitious aspiration.

In this paper, we explored how the capability of LLMs can be effectively integrated into OHMDs to enhance personal productivity while ensuring the generated content aligns with the user's individual style and minimizes the need for attention-intensive editing. We present GlassMail, a proof-of-concept prototype system designed to address the complexities identified in our formative studies. The development of GlassMail serves as a case study in our exploration of the intersection between LLM-based assistants and personal information management. To better illustrate the functionalities of GlassMail, we begin by presenting its usage scenario.

Abby is currently interning at a company. One morning, while wearing her smart glasses on her way to the subway station, she received an urgent email from her colleague, Tony, requesting additional details about the location of an upcoming sponsorship event venue. Tony was visibly anxious and needed an immediate response.

Abby promptly provided the email's context (i.e., informing Tony that the venue has been rented and Emily will meet with him this afternoon to finalize all details) to GlassMail, an application integrated into her smart glasses that had previously

[1] https://en.wikipedia.org/wiki/The_Devil_Wears_Prada_(film)

learned Abby's personal writing style. GlassMail efficiently identified the missing information (i.e., Tony's preferred meeting time) through a single supplementary question. It then generated an email draft based on this context. GlassMail's Fade Context visual output, accompanied by optional audio, allowed Abby to edit the email while remaining aware of her surroundings. Abby also reviewed the email while walking outdoors, with the assistance of audio.

The email already matched Abby's writing style, but she wanted to make slight adjustments in length and the order of content. Therefore, she utilized the 'suggestion hooks' feature (e.g., shortening the length) to request these modifications, aligning the email more closely with her and Tony's collaborative style (i.e., keeping it within 200 words). Abby further refined the email's structure using the 'keywords-based editing' feature, simplifying keywords to adjust their order, such as placing the Time inquiry at the end.

After making these refinements, Abby reviewed the email with satisfaction, confirming its readiness. She then asked GlassMail to send it to Tony. Although the entire process took only a few minutes, Abby recognized the crucial role this email would play in ensuring the success of the sponsorship event.

The design of GlassMail leverages insights on OHMDs' text display design guidelines from previous research [32, 39, 86, 116] and findings from our empirical studies to pave the way for a more efficient and user-friendly on-the-go personal email assistant. More specifically, we gathered insights from 1) a controlled study ($n = 12$) to identify the modality and visual output mode needed to ease information access from OHMDs, and 2) an observational study ($n = 12$) to elicit how users should be supported in their coarse and fine-grained editing needs with an LLM, to inform our design GlassMail.

We evaluated the usability of GlassMail in a real-world user study ($n = 8$) spanning a variety of locations and mobility tasks. The feedback we received from participants was highly positive, with most users finding it a considerable improvement over their traditional approach using phones, highlighting the success of GlassMail in bridging gaps between LLM-driven assistance and the challenges of ubiquitous text creation. While GlassMail may not yet match the intelligence of an expert personal assistant like Emily in "The Devil Wears Prada," it effectively fulfils its design objective of aiding users in composing emails of various complexities within the constraints of mobile scenarios with significantly reduced distraction to the primary tasks, marking a significant step toward that vision in the future. In this paper, we present the following contribution.

- An in-depth understanding of the gaps between AI (LLM) capabilities and users' needs in creating formal text in mobile scenarios.
- Empirically-informed design and development of GlassMail, an artifact that optimizes human-AI collaboration to facilitate email writing in mobile and multitasking contexts, which includes: 1) Support for users to retain their personalized writing styles. 2) A three-stage approach to gather information, compose a draft, and edit it to generate emails that are more accurate and robust. 3) Improved visual output mode for glanceability on OHMDs and choice to switch to audio in high cognitive load situations. 4) Interaction mechanisms tailored for coarse and fine-grained editing needs of users.
- Real-world usability study of GlassMail, demonstrating our system's capability as a personal writing assistant and design guidelines for future AI-based ubiquitous personal writing assistants.

2 RELATED WORK

2.1 Background

The overarching goal of our envisioned wearable intelligent assistants rivals the intelligence of an expert personal assistant, such as Emily in "The Devil Wears Prada." A key task that an intelligence assistant can support is Personal Information Management (PIM) on the go [117]. PIM, which are essential activities for knowledge workers, refers to an individual's daily practice of storing, organizing, retrieving, controlling and using information [14, 51]. In our fast-paced, information-driven society, the demand for PIM anytime and anywhere has never been greater. While existing mobile handheld devices, such as smartphones, allow us to access and manage our personal information on the go, they also present significant issues - situational impairment due to resource conflict and information fragmentation [117].

Intelligent Personal Assistants (IPAs) have emerged as a potential solution (e.g. Apple Siri[2] and Amazon Alexa [3]). IPAs leverage voice commands and contextual awareness, enabling hands-free and eyes-free interactions with mobile devices [93]. However, despite their availability, most people do not use IPAs regularly on the go [73]. Research reveals that infrequent IPA users often encounter a challenge known as "Hands-Not-Full-Free Interaction." [26]. While IPAs appear to offer hands-free usage, users often find themselves needing to touch the screen for certain actions using mobile phones, compromising the hands-free experience. Moreover, even frequent IPA users tend to treat interactions with these assistants as secondary tasks, hesitating to delegate complex actions, such as composing emails, due to concerns about accuracy and trustworthiness [71].

Optical See-Through Head-Mounted Displays (OHMDs) present a promising solution to the challenges at hand. These devices champion a heads-up, hands-free interaction model, potentially delivering in-context, real-time digital aid when and where it's most needed [113]. Nonetheless, fine-tuning OHMDs for efficient on-the-move information processing remains a formidable hurdle. Our driving question centres around designing OHMD-integrated mobile PIM systems that adeptly reconcile users' content generation needs with their efficacy in mobile multitasking environments. While addressing every task in PIM within a single paper is overly ambitious, we've prioritized exploring on-the-go Email assistants due to their centrality in PIM. An effective mobile email assistant should mimic the nuance and adaptability of human assistance, navigating fragmented attention, partial inputs, and occasionally vague or incomplete directions. These interactions ought to be crafted to minimize visual and cognitive loads, enabling users to remain engrossed in their primary activities. In the next section, we dive into the topic of using LLM to improve content creation, another crucial challenge to address.

2.2 Challenges of on-the-go Formal Content Creation

Email creation on mobile devices presents unique challenges compared to desktop environments. These challenges revolve around limited input and output capabilities and constrained multitasking demands of on-the-go scenarios.

2.2.1 Input Constraints. Typing on mobile devices is notably slower, averaging 36.2 words per minute (WPM) compared to over 120 WPM on desktop keyboards [27]. Mobile users often encounter increased error rates when typing, especially when engaged in activities like walking [10, 49, 67, 81]. Mobile typing also demands significant visual attention and imposes a high cognitive workload [39, 49]. While voice input offers a potential solution, its effectiveness for precise editing in formal content creation remains a concern [39]. Existing voice editing systems, such as EYEditor [39] and SmartEdit [29], are designed for shorter texts and lack proven efficacy in editing longer, formal paragraphs. Multiple

[2]https://www.apple.com/siri/
[3]https://developer.amazon.com/en-US/alexa

input methods have been developed for OHMDs. Among these, virtual keyboards represent a prominent text input approach [55]. Two prevalent implementations of virtual keyboards include tap-typing and word gesture typing [55]. However, tap-typing achieves a modest typing speed of 15.44 words per minute (WPM) with a 0.97% error rate [96]. Although research has indicated improved performance with word gesture typing compared to tap-typing [74, 110], users still attain an average typing speed of only 24.73 WPM after 60 minutes of training [110]. This speed remains notably slower than typical desktop typing rates. Additionally, these methods often require a substantial amount of user attention and effort, which contradicts the principle of receiving seamless assistance from intelligent personal email assistants [114].

2.2.2 Output Limitations. Mobile screens are less suitable for reading lengthy content due to size constraints [11, 34], and reading on mobile devices can isolate users from their surroundings, potentially increasing safety risks [68]. The heads-down posture associated with mobile reading can also negatively impact user health [39]. OHMDs offer a promising alternative for on-the-go content consumption [39]. Research [70, 83] indicates that OHMDs, when paired with voice output, can enhance the user experience for short text processing. Vadas et al. [100] compared the performance with audio or visual output in text comprehension tasks while walking. While usually requiring more time for users to complete, audio output imposed a lower workload, allowing users to maintain a higher walking speed compared to visual output. Different kinds of visual representations of text have also been explored [89, 116]. Rzayev et al. explored [89] the effects of different text positions and methods and proved that scrolling methods are more suitable for comprehension while walking and centre and the bottom-centre are better positioned for text display. Zhou et al. [116] compare the influence of line spacing and word spacing on OHMD and found increasing interline spacing to be an effective way to improve mobile OHMD text reading. However, designing effective email displays for OHMDs remains challenging, especially regarding structured email content.

2.2.3 Multitasking and Situational Impairments. The nature of on-the-go formal content creation introduces multitasking challenges. This endeavour necessitates users to concurrently perform several cognitive and physical tasks, often in intricate and dynamic environments, increasing the risk of accidents [20, 44]. In addition to the inherent multitasking difficulties, the situational demands originating from the surrounding environment can substantially disrupt the content creation process[105]. These situational impairments might manifest as unexpected interruptions, ambient distractions, or unforeseen obstacles, all of which can hinder the user's progress and lead to a higher incidence of errors [57]. Moreover, multitasking can also expose users to higher accident risks while walking or driving [20, 44], individuals who attempt to multitask while walking or driving may not only be at a higher risk of accidents but also more prone to producing content of lower quality or accuracy [106]. Therefore, balancing the need for productivity and creativity with the potential consequences of multitasking and situational impairments is a critical aspect of harnessing this technology effectively and responsibly.

2.3 Challenges for on-the-go Voice-based Editing

In eyes-free scenarios, voice-based text input is a natural and efficient modality [107], allowing users to engage in tasks like messaging, note-taking, or drafting while occupied with activities such as walking or driving. However, editing voice input, a crucial step, presents challenges. It involves removing colloquial expressions, correcting recognition errors, and revising inappropriate words to ensure precise communication [21]. This editing process requires spatial referencing, making it challenging due to the linear [95] and temporal nature of audio [62], which can increase the mental load and error rates [38]. Recent studies have introduced two primary strategies for voice editing. The first

involves using descriptive commands, such as "Change bat to cat". EDITalk [38] have implemented this for core editing tasks like insertions, replacements, and deletions, but this approach can be cognitively taxing as it requires users to remember the original text and specific command syntax [40]. The second strategy is re-speaking a part of the text with corrections, like changing "the bat sat" to "the cat sat". EYEditor [39], a smartglass-based voice text editor, employs this method for mobile users, combining it with manual input for text navigation and selection. This method, however, demands sufficient context to avoid alignment errors, often leading to multiple attempts by users for successful edits [39, 102]. In mobile contexts, editing AI-generated content, such as that from LLMs, adds another layer of complexity. The AI's output might not always match the user's style, necessitating further edits that could interrupt primary tasks, raising questions about how LLM-based assistants can efficiently meet users' editing needs through voice input. Previous research also indicates that switching from voice-only to multi-modal inputs (e.g., voice + mouse) can enhance error correction efficiency [39, 97]. Thus we propose a hybrid interaction approach of voice and wearable ring-mouse input. Voice is used to interact with LLM-based assistants, while ring-mouse facilitates quick navigation and seamless confirmation tasks.

2.4 Large Language Model Support Complex Content Creation

With the great success of the Large Language Model (LLM) in the natural language processing area, human-level performance on text-based tasks by the LLM, as exemplified by GPT-4 [79], has inspired various research on human-AI interactions [13, 48, 59, 103], especially LLM support writings [36, 37, 42, 63, 101]. Leveraging the impressive writing capabilities of LLMs holds promise for on-the-go formal content creation.

2.4.1 AI-assisted writings. Various studies have explored the capabilities of AI, especially LLM in assisting various types of human writings, such as storytelling [13, 24, 37, 111], theatre scripts [75], messages [101], emails [42] and scientific materials [33, 36, 108]. Despite their prowess, LLM-powered writing assistants are not without limitations [32, 37, 42, 75, 84]. Fok and Weld [32] summarized the limitation of LLMs as writing assistants into five aspects: hallucination, inconsistent content and style, repetition, mediocrity and ethical concerns.

Style problems in AI-assisted writing have been identified by various studies [42, 84, 101]. For example, in a study comparing texts written by AI assistants against human authors, the stylistic variations were found to be less in AI-generated texts [84]. When developing an LLM support communication system, Valencia et al. [101] reported participants' need for AI-generated phrases to reflect their own communication styles and preferences. In the study of LaMPost, an email writing with LLM support, users often felt the need to rewrite paragraphs in specific styles [42].

2.4.2 LLM supports email writing. AI has been widely used in supporting human email writing. One common type is automated response suggestions such as Smart Reply in Gmail [53]. Additionally, continuous email generation with AI has also been explored and is proven to have the effect of reducing the feeling of authorship while making users edit more [64]. Goodman et al. [42] proposed LaMPost, an LLM-support desktop email-writing interface to enable email writing for people with dyslexia. However, the evaluation study indicated that LaMPost may not meet the needs of writers with dyslexia in terms of accuracy and quality due to hallucination and misunderstanding [42]. A naive approach with few shot learning prompts is not enough to eliminate the existing obstacles in utilizing LLM for email generation.

In addition, previous studies on AI suggestions or AI-generated content in emails have highlighted maintaining personalized style as a crucial criterion for user satisfaction [42, 88, 101]. While current LLMs, like GPT-4, can produce high-quality writings comparable to or even surpassing those of average human writers [79], using them directly for

email composition may not result in highly personalized emails without further specifications regarding users' unique styles [32]. Additional efforts may be necessary for an on-the-go email creation system to genuinely assist users in easily composing ready-to-send emails.

In summary, creating formal content on the go presents numerous complexities [98, 100]. Nevertheless, the growing importance of personal information management in the digital age and the impact of wearable devices on information management are evident. Specifically, crafting content on mobile platforms and OHMDs entails unique challenges [74, 96, 110]. However, it is imperative to seamlessly integrate these technologies into users' daily lives, setting the stage for a deeper exploration of the role LLMs can play in addressing these challenges. In particular, focus on the feasibility of the on-the-go approach to email, and address technical, stylistic, and personalization issues.

3 FORMATIVE STUDY

In the introduction, we highlighted the prevalent use of Large Language Models (LLMs) for text tasks in stationary desktop environments. But can they be directly used on mobile and wearable devices to achieve the desirable effects of wearable email assistants?

To better understand the practicality of using smartphones for email writing and the feasibility of leveraging LLMs for meaningful content generation in mobile settings and the challenges that users might face therein, we conducted a study involving 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, fluency in English, experience in writing emails using smartphones and experience using ChatGPT for emails on desktops. Half of them had previously used OHMDs.

3.1 Apparatus and Procedure

In the study, participants were tasked with mobile email creation through two approaches: initially using smartphones, and subsequently collaborating with LLM-based assistants. We then detail the apparatus and procedure for each approach.

To comprehensively assess the practicality of using smartphones for email composition in mobile contexts, we implemented real-world tasks as outlined in previous research [39]. These tasks included indoor walking, ascending and descending stairs indoors, outdoor walking, taking a bus, and indoor shopping. For the study, participants used an iPhone XR equipped with the Google Gmail[4] and the Gboard [5] applications, which provided them with the choice to either type or dictate emails on the mobile phone. Participants were required to continue writing emails while actively engaged in these real-world mobility tasks. Before commencing the task of email composition using the large language model (LLM)-based system, participants were given a 10-minute break. This intermission was provided to ensure they were well-rested and could approach the LLM-based email writing task with renewed focus.

We used Nreal Light glasses [2] due to its lightweight (106 grams) and clear visual clarity (stereoscopic display resolution is 1920x1080 pixels) which is necessary for dense text display. During the study, participants wore smart glasses [2] while walking indoors, guided by an experimenter carrying a connected MacBook Air (M1, 2020) [3]. They interacted with a ChatGPT web page mirrored on the glasses using voice control through a Chrome plugin [6] to compose emails with LLM assistance on the OHMD. Participants submitted a self-written email typical of those they frequently

[4]https://www.google.com/gmail/about/
[5]https://en.wikipedia.org/wiki/Gboard
[6]https://voicecontrol.chat

send in their daily lives before the study. They then recreated the pre-submitted self-written email while walking. Afterwards, semi-structured interviews gathered feedback on the issues they faced and their requirements. We also inquired about their experience of using their mobile phones to create emails. The entire study took approximately 90-120 minutes per participant.

3.2 Results and Discussion

Interviews were audio-recorded, transcribed, and finally coded using Thematic Analysis [16]. We used an inductive approach to craft themes related to the major challenges that users faced in using smartphones and LLMs for mobile email creation. Themes were derived from data based on both their frequency of occurrence and perceived substantive significance.

3.2.1 Smartphones' Practicality for Mobile Email Writing. The study revealed significant limitations in using smartphones for email writing in mobile scenarios. Participants invariably had to stop moving to concentrate on their emails, interrupting their main activities. Tasks like ascending/descending stairs or shopping proved impossible for email composition on smartphones. In scenarios of walking, both indoors and outdoors, only half of the participants could complete their emails, yet they were not confident about the quality and were hesitant to send them. On a bus, resembling a more stable desktop-like environment, the experience still wasn't favourable. Challenges included the bus's movement, ambient noise, difficulty focusing on a small screen, increased errors in typing or dictation, and discomfort in multitasking. However, it was observed that smartphones are suitable for composing simpler emails, similar to short message responses (e.g., one sentence or few words), where informal language and less stringent formatting are acceptable. In these less demanding contexts, using smartphones for email writing was practical.

3.2.2 Overall Experience for OHMD LLM-based Mobile Email Writing. Eight participants successfully initiated their first interaction with LLM to generate email content. However, the remaining four participants faced difficulties during their initial interaction with LLM, as it failed to provide the desired email content due to a lack of understanding of their needs, prompting them to seek further clarification. Additionally, once their needs were initially misunderstood, continuing the dialogue interaction did not resolve the issue in a timely manner. Two of these participants wanted to abandon the process and start over after 2-3 attempts, even expressing a preference for manually writing the email themselves (P2, P7).

While participants appreciated the speed at which LLM generated emails, they expressed dissatisfaction with the initially generated content. They found it to be overly polite and lengthy, often containing extraneous material that required deletion. Participants struggled, for the most part, to get LLM to spell various names correctly, including people's names and places, when using voice input. Notably, P5 made ten unsuccessful attempts to correct his name. These spelling errors and unusual content in emails underscored the potential negative impacts, leading most participants to hesitate when it came to sending such emails as they didn't want others to feel they were impolite.

While they were confident in using LLM for email composition when trying it on their own laptops before, they felt that it did not save them a significant amount of effort, and they found it more of a chore to edit the content to match their writing style while walking. As P6 mentioned, "I think it's too much to do at the moment. in the time it took to compose emails with LLM while walking, I could complete many of them more efficiently by hand." Additionally, all participants found the text displayed in front of them to be overwhelming while walking. P2 and P8 even complained about excessive content in their view, such as conversation notes, email content, and other irrelevant information. They expressed a preference for only seeing the email content.

We further summarized the main challenges of applying LLM on OHMD for mobile email creation (see Table 1).

3.2.3 Lack of Interaction Support. Utilizing LLMs for Personal Information Management (PIM) tasks, including email composition through OHMDs while mobile, presents unique challenges, primarily arising from the quality of voice-based input prompts. Whereas desktop users can fine-tune prompts to get the necessary output from the LLM through a feedback loop, achieving similar refinement with limited I/O options in mobile contexts is more challenging, particularly when using voice. Participants also expressed discomfort using voice-based interactions without the option for manual editing. Once their intentions are misinterpreted, they are also at a loss as to how to correct them through further interaction with LLM. A notable instance involved a participant (P9) whose mention of "Robert" was misinterpreted as "robot" by the automatic speech recognition (ASR), leading to the LLM generating content suggesting an AI assistant wished to join dinner. Despite P9's attempts to correct the name using different phrases, the LLM retained the incorrect context until P9 clarified that there was no AI assistant, only herself and Robert. This highlights the necessity for more reliable voice recognition and better mechanisms to interact with the LLM.

3.2.4 Lack of Support for Effective Information Output. Participants struggled to understand email content on OHMDs while walking due to the need to split their visual attention between the display and their surroundings. This multitasking led to fragmented attention, impacting comprehension. While prior research recommended displaying text sentence-by-sentence for mobile OHMD voice-based editing of a piece of text [39], 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 whereas, always presenting everything at once can overwhelm users, especially during editing [116]. Hence, it is crucial to find a method that highlights edits without forcing users to reread the entire content.

3.2.5 Difficulty Achieving Personalization. Personalization is essential in email communication [78], and participants expect their unique writing styles to be reflected in the emails. While LLM has demonstrated the potential of generating content that contains most of the necessary information, such content often lacks a personal touch. Moreover, it is unclear how they can interact with the LLM to personalize the email style.

In summary, while LLM has demonstrated potential for email writing on desktops, using LLMs for the same purpose on OHMDs in mobile scenarios presents several challenges. These include the inefficiency of editing the generated content with voice if it did not meet their expectations, difficulty understanding the edits made by the LLM based on their prompts, and the generic writing style of mails created by LLMs. Solving these challenges requires tailored input and output designs, which we explore through our design studies in the following section.

4 ITERATIVE DESIGN AND EVALUATIONS

After gaining a better understanding of the challenges, we initiated a series of investigations to design optimal solutions for these problems(see Table 1). We describe the structure of our investigations below.

- **Improve the interaction style with LLMs to enhance communication accuracy while minimizing attention and effort required from users.** To begin, we wanted to minimize the need for users to make edits to the generated content given the high cognitive demand it imposes on users. For this, we conducted pilots ($n = 6$) comparing different voice-based interaction styles with LLMs in terms of their ability to generate output of desired quality, while balancing user experience.

- **Create an OHMD output mode that clearly displays LLM edits in a user-friendly format, reducing the cognitive and visual load to improve user experience without hindering the primary task.** Once users generate content with LLMs in mobile scenarios, they need to view and comprehend it efficiently on OHMDs. We conducted a controlled study ($n = 12$) to systematically investigate the suitable modality and specific visual output modes that can help users utilise the generated content more efficiently in these situations.
- **Improve personalization of LLM-generated content by designing user-friendly editing schemes for the mobile OHMD context, aimed at enabling user-driven final customization.** We finally looked at how to better support the remaining editing needs of users with LLMs. For this, we designed an initial version of GlassMail (Section 4.3) based on findings until this point and literature that supports the composition and

Table 1. Design Goals, Challenges and GlassMail's Solutions for OHMD LLM-based mobile email writing

Design Goals	Challenges	GlassMail Solutions
Improve the interaction style with LLMs to enhance communication accuracy while minimizing the attention and effort required from users (D1).	In mobile multitasking contexts, the refinement of voice input prompts presents a dual challenge: • C1: The correction or addition of information via voice is hindered by the absence of detailed feedback mechanisms. • C2: Users often have insufficient attention to confirm whether their prompts were accurately captured leading to potential miscommunications with the language model.	Lazy one-shot approach to interacting with LLM.
Create an OHMD output mode that clearly displays LLM edits in a user-friendly format, reducing the cognitive and visual load to improve user experience without hindering the primary task (D2).	• C3: The difficulty understanding the edits made by the LLM based on their prompts. • C4: Due to multitasking constraints and situational impairments, effectively understanding or processing output with limited visual and cognitive loads without affecting the primary task is challenging.	Fade Context visual feedback for glanceability with optional audio.
Improve personalization of LLM-generated content by designing user-friendly editing schemes for the mobile OHMD context, aimed at enabling user-driven final customization (D3).	• C5: The generic writing style of emails created by LLMs lacks a personal touch. • C6: It is unclear whether and how LLM-based assistants can meet users' editing needs to achieve final personalisation in the OHMD mobile context.	Few-shot learning with users' email to generate personalized emails with LLMs. GlassMail provides two distinct editing schemes tailored to users' coarse and fine-grained editing needs: 'Suggestion Hooks' provide quick shortcuts for rapidly adjusting the email in a global manner (e.g., tone and length), while 'Keywords-based Editing' enables more detailed and specific modifications, focusing on localized scoping of the content.

viewing of OHMDs. We then conducted an observation study ($n = 12$) to understand the types of edits that users often need to perform and how we can better support these edits.

Combining insights from these studies, we developed the final version of GlassMail, which we evaluated in a real-world usability study to gather qualitative feedback on its usability for email writing in mobile scenarios. All studies (including the previously mentioned formative study) received consent from participants and approval from the university's institutional review board (IRB).

4.1 Pilot Study: Explore Ease of Interaction with LLM-based Wearable Assistant

In the mobile context, users no longer have the luxury of using a manual input (e.g., keyboard or mouse) to perform editing tasks. Also, they have limited patience and attention resources, which poses challenges for them in achieving their personal writing styles with LLM. From our initial formative pilot findings, a simple inquiry with LLM often fails to achieve such personalization although they could easily give their needs in one shot interaction. Prior research has demonstrated that improved responses from LLMs can be attained through dialogue-based interactions [19]. Yet this may prolong the email composing process, thus compromising efficiency and user impatience. These methods raise a trade-off between efficiency and usability: there can't be too many interactions with the LLM while ensuring the accuracy of key elements for email writing. Hence, we conducted a pilot study with 6 participants between 19-24 (2 females, $M = 20.7, SD = 1.86$) to compare these three interaction approaches (see Figure 2) with LLM-based assistants for Heads-up mobile usage: 1) One-shot approach; 2) Dialogue approach; 3) Lazy one-shot approach. This approach involves users expressing their needs in a one-shot interaction, and then being presented with critical information to fill in any missing or supplementary details for composing an email.

4.1.1 Apparatus. Participants wear the Nreal Glasses connected with MacBook Air (M1, 2020) [3] 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 [47, 86, 94] (Apparatus see 5(b)). 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.

4.1.2 Materials and Procedure. Responding to a single email or email threads that require users to contemplate and catch up on the context is not the scope of our study to avoid the confounding effects of other variables.

We chose email tasks from email analysis pilots according to their utility and usage in daily life and private concerns. They are event coordination email tasks, such as sending invitations, reminders, or updates about upcoming events. Participants were required to compose emails according to the given email tasks using three interaction styles to three different social ties (i.e., supervisor, sister, friend) [28, 31]. The order of materials and the order of interaction approaches are counterbalanced with the Latin Square design. On completion of the study, they need to fill in a post-questionnaire survey followed by a semi-structured interview to understand their issues and requirements.

4.1.3 Results and Discussion. Our findings indicate a preference for the *lazy one-shot* approach for OHMD mobile use with LLMs. Most participants (5 of 6) preferred this approach because it could "ask for clarification for important parts of email" (P2, P3) while not going too far in "asking information that is already available" (P4), thereby balancing content quality and workload effectively. This approach also helped users to visualize what they wanted to write before

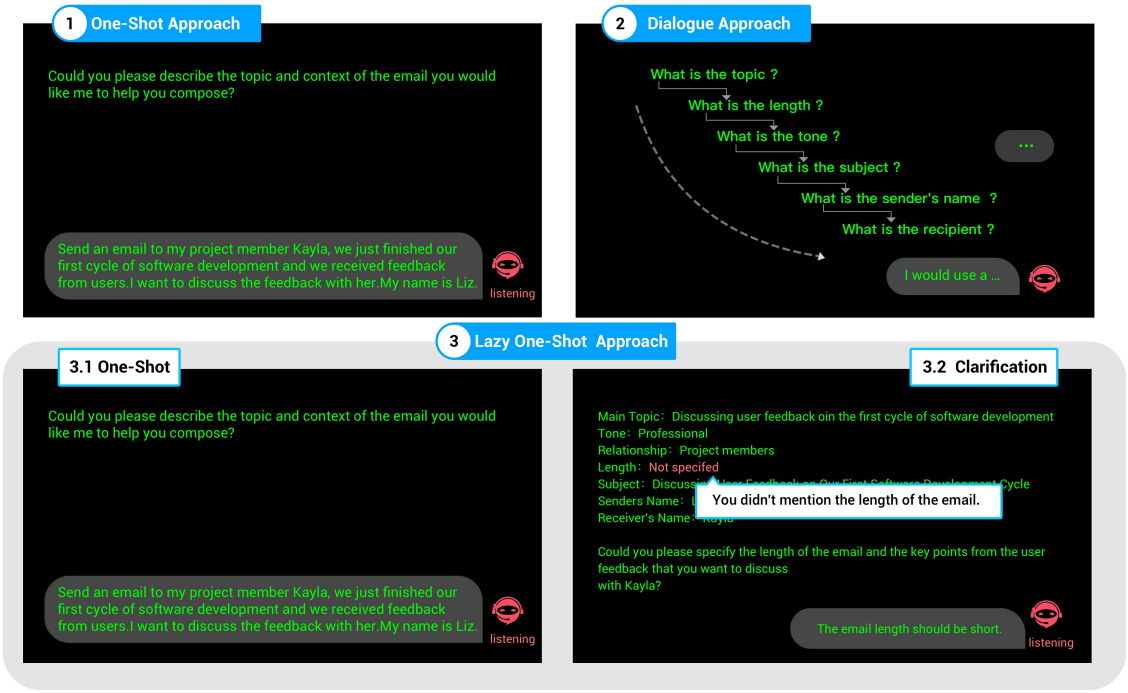


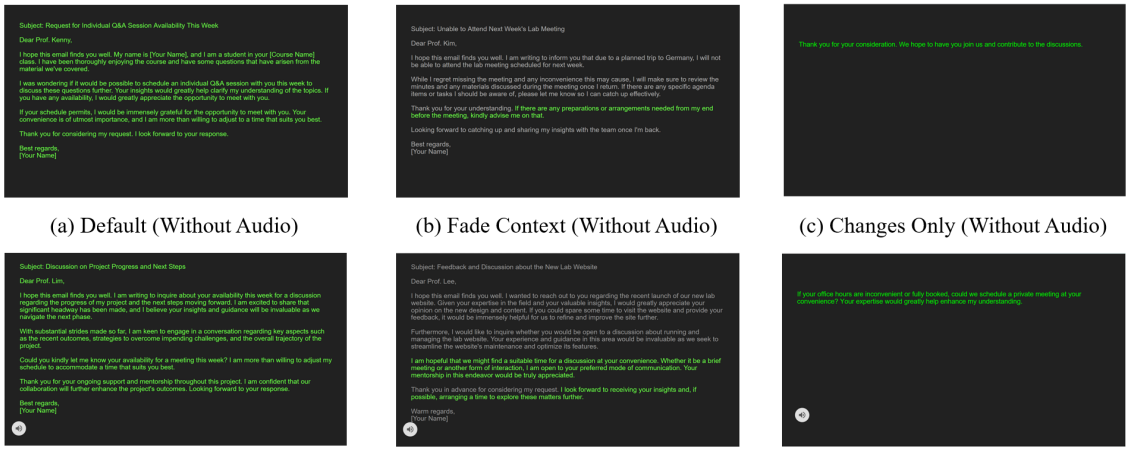
Fig. 2. Examples to show the three interaction approaches used in the Pilot Study: 1) One-shot, 2) Dialogue and 3) lazy one-shot approach.

the email was generated, offering users the possibility of rectifying it whenever they wanted (P5). One participant favoured the swifter one-shot approach despite its lower content quality (P4). None opted for the dialogue approach as it "still sometimes missed details" (P2) 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.

4.2 Controlled Study: Explore Ease Information Absorption for OHMD Mobile Usage

We move on to address the next essential question: How can we design the output of LLM for heads-up mobile usage on OHMDs? Information overloading can be a serious issue in mobile multitasking scenarios, where users' attention is fragmented between the display and the environment. We aim to enable the user to swiftly and effortlessly absorb LLM's edits to the email without dedicated focus, necessitating an optimal balance between these two requirements.

We focus our investigation on two pivotal elements: output modality (visual, or a visual and audio hybrid) and the visual output mode of text content on OHMDs [22, 89]. While the output modality can be audio, visual, or a hybrid, audio only might not be suitable for email content that requires specific structure and is carefully wordy. Previous work also revealed that editing text on OHMD in an audio-only mode was cognitively very challenging if a continuous stream of audio was presented [39]. This study seeks a harmonious balance in presenting edited changes to facilitate quick comprehension while minimizing cognitive load.



(a) Default (Without Audio) (b) Fade Context (Without Audio) (c) Changes Only (Without Audio)

(d) Default (With Audio) (e) Fade Context (With Audio) (f) Changes Only (With Audio)

Fig. 3. All conditions used in the Controlled Study: a) Default, b) Fade Context, c) Changes Only, d) Default (with Audio), e) Fade Context (with Audio) and f) Changes Only (with Audio). The key difference lies in the presentation of the email context (e.g., unchanged content of the email): Default displays the entire email content, including both edits and unchanged parts; Fade Context uses a less noticeable colour (e.g., gray) to fade the unchanged context, and Changes Only omits the context entirely, displaying only the edits.

4.2.1 Modes of Visual Output. We designed 3 visual output modes, each exploring a combination of audio and/or visual modality. The visual output can be rendered in three different presentations: Default, Fade Context, and Changes Only (Figure 3 shows all output modes). The difference is how to show the email context except for the edited changes: display as default settings (Default), using unnoticeable colour to fade the context (Fade Context) or not display the context (Changes Only). The default display settings were chosen according to previously researched best practices for mobile OHMD text display. The text was green against a black background that appears transparent on OHMD [61, 80, 90].

4.2.2 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$7.5 per hour.

4.2.3 Apparatus and Materials. We conducted a Wizard-of-Oz (WOZ) study [91] using Nreal Light glasses. The smart glasses mirrored the display of an iPad Mini tablet [1], showcasing various displays (the visual output modes) crafted using PowerPoint slides[72], hosted on a MacBook Air (M1, 2020)[3] and shared to the iPad via Zoom[9].

We replicated the email tasks from Pilot 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 [99] at a default pace and incorporated into the appropriate slides.

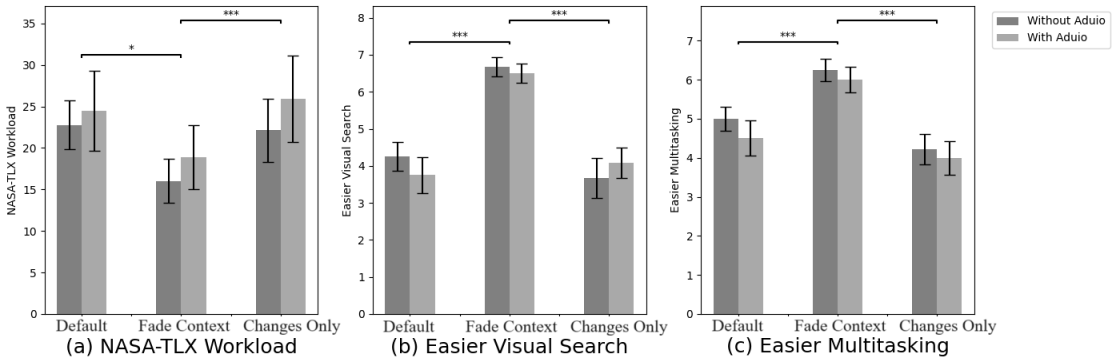


Fig. 4. Controlled Study Results: Means and standard deviations (SD) of significant measures: (a) NASA-TLX workload, (b) Easier Visual Search, (c) Easier Multitasking for both visual output with and without audio mode.

4.2.4 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).

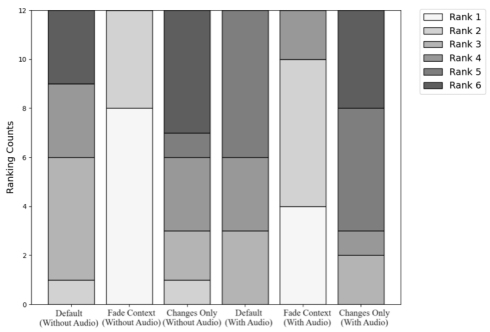
A fully crossed design resulted in 6 combinations per participant. The study starts with participants signing a consent form and completing a questionnaire about their demographic information. We then provided an example task to familiarise participants with the experiment and devices as part of the training session. Afterwards, they began the actual experimental task. 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.

4.2.5 Results and Discussion. We measured each output mode's task workload (i.e., NASA-TLX [43]), information absorption in easier visual search for edited changes, and support for multitasking using 7-points Likert Scales, as well as user overall preference for all output modes. This scale is adapted from prior works [18, 63] and we further make it more suitable for our settings.

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 main effects of 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 4). 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. There was no significant difference between Default and Changes Only on all measurements. 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.

Overall Preference and Feedback. Eight Participants preferred the Fade Context while four participants preferred the Fade Context with Audio (see Figure 5(a)). 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



(a) User Pereference Rank

(b) Apparatus

Fig. 5. (a) Users' Preference Ranking for Six Conditions: Visual Output Mode (Default, Fade Context, Changes Only) x Audio (With, Without) in the Controlled Study. (b) The apparatus shows all devices including the NReal OHMD and Ring Mouse.

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). For those who preferred visuals only because they felt adding audio was "confusing" (P3), "distracting" (P5, P10, P11), and even "makes the task even harder" (P5).

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 with a simple walking task that required a little visual load. In this task, participants needed to shift their visual attention between the OHMD display and environmental signs while on the move similar to the setting in Zhou et al., [116]. 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 [6, 15]. Given this variability in preference and the fact that audio may still be useful during situations that are more cognitively demanding [82, 104], we propose adopting the Fade Context mode with optional audio as the optimal output design for our system.

4.3 First Version: GlassMail

The findings from our design exploration studies inform our initial interface design. GlassMail incorporates a hybrid interaction approach of voice and wearable ring-mouse input. Voice is used to interact with LLM-based assistants, while

ring-mouse facilitates quick navigation and seamless confirmation tasks. The full view of the GlassMail interface is illustrated in Figure 6.

Input Interaction style with LLMs is lazy one-shot interaction. Our initial pilot study suggests that this 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 (C1) and users' limited attention (C2), leading to misinterpretations. Rectifying these misinterpretations needs further effort from users thereby defeating the simplicity of the lazy one-shot approach. To address this, we propose a design with real-time voice interactions (C1) and "Fragmented attention-friendly chunking" to break information into manageable parts for divided attention scenarios (C2). 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.

4.3.1 Real-time voice transcription. Real-time voice transcription 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 [76] findings, users often opt to quit or restart upon encountering misinterpreted intent. To enhance GlassMail's usability, we implemented real-time voice transcription visible to users as they speak (see Figure 6 (1)), coupled with a straightforward method for making corrections by repeating words or phrases.

4.3.2 Fragmented Attention-Friendly Chunking. It is designed for efficient error detection and correction in voice-based inputs, minimizing cognitive strain. 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.

Fade Context visual output mode with optional audio mode. GlassMail features the 'Fade Context' coupled with an optional audio output, specifically designed for the OHMD mobile environment (see Figure 6 (4)). This feature utilizes an unnoticeable colour, such as gray, to diminish the visibility of unchanged text, thus highlighting edits. This approach effectively addresses the issue in C3 by making it easier for users to understand the modifications made by the LLM in response to their prompts. Furthermore, it tackles the C4 challenge by enabling 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.

4.3.3 Implementation. The first version of GlassMail employed the GPT-3.5-Turbo-16K model as its primary Large Language Model (LLM) to process voice 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 email. Several prompt engineering techniques were utilized to enhance the output (further details in Appendix A). 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[7] 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 [5] from web browsers to enable TTS.
- **ASR engine**: First, the RecordRTC [35] library is used to capture and record the user's voice input. Speaking detection is implemented through the hark library [77], monitoring when the user starts and stops speaking to initiate and terminate recording. Audio data is encoded into MP3 format using the lamejs library [30] to reduce file size. Recorded audio data can be sent in real-time to the OpenAI Whisper API [85], which returns translated text. Robust error handling is implemented to manage exceptions, including issues with acquiring media streams, recording failures, or translation errors.
- **CTRL module**: This module integrates ASR and TTS engine into a closed-loop design and manages control operations such as start, stop, and reset.

4.3.4 GlassMail Interaction Process. As shown in Figure 6, users begin their lazy one-shot 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 upper key on the ring mouse again. Then they press the lower button on the ring mouse activates 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.

4.4 Observation Study: Understanding Editing Needs

In the two previous design exploration studies, users did not really engage in email editing. It is unclear whether GlassMail, developed based on previous findings, can fulfil users' editing requirements for achieving final personalization in the OHMD mobile context. We conducted an observational study to understand the challenges users faced and determine the level of editorial precision required to achieve users' desired personal email content when collaborating with LLM-based AI assistants (i.e., GlassMail) in the OHMD mobile context.

4.4.1 Participants. We recruited 12 participants (5 females) between 19 and 23 ($M = 21.3, SD = 1.5$) from the university community. All participants had 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$7.5 per hour.

4.4.2 Apparatus and Materials. We used the first version of GlassMail as described in Section 4.3. For a realistic composition and editing experience, participants were asked to provide three email scenarios involving supervisors,

[7] https://github.com/kpdecker/jsdiff

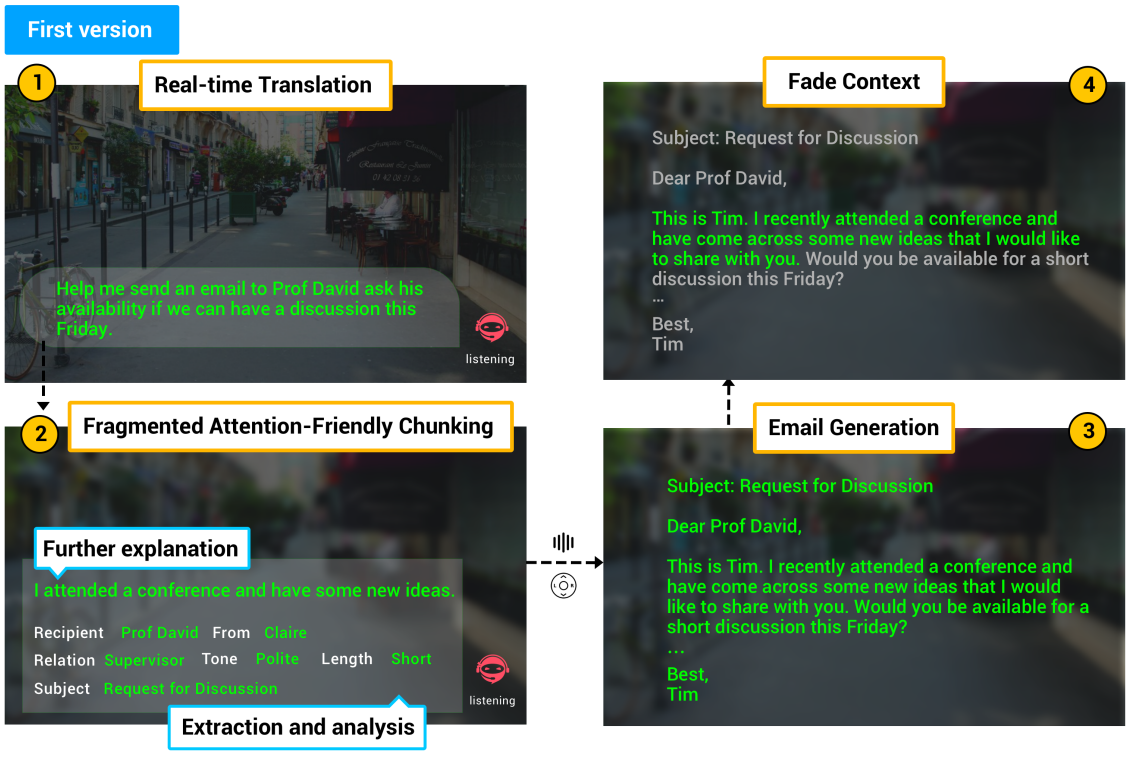


Fig. 6. The initial interface of GlassMail features the lazy one-shot approach to interact with the LLM-based assistant for composing emails. The lazy one-shot approach includes (1) Real-time translation and (2) Fragmented Attention-Friendly Chunking display. It also utilizes the (4) Fade Context mode with optional audio as the output mode for displaying editing changes.

friends, and family members they often mailed in reality. These scenarios were used as the email composition and editing tasks in the study.

4.4.3 Procedure. The study began with participants signing consent forms and providing demographic information. Following this, a system demonstration was conducted, accompanied by a training session. The observation study comprised three sessions, each involving a real scenario provided by the participant. Participants were tasked with composing and editing emails using our system until they felt that the content was similar to their writing style. After completing all three sessions, semi-structured interviews were conducted to gather insights about their GlassMail interface experience.

4.4.4 Results and Discussion.

Editing Process and Frequency. Overall, all participants began with high-level editing adjustments, with a particular focus on email length and tone. Users 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

	Global edit: tone	Global edit: length	Fine edit: paragraph	Fine edit: sentence	Fine edit: word
Supervisor	2	3	9	68	20
Friend	2	2	17	57	24
Brother/Sister	5	2	9	59	32
Total	9	7	35	184	76

Table 2. Frequency of different levels of editing under different scenarios

incorporated brief sentences between paragraphs to align with his personal writing style. We summarize the frequency of different levels of editing in the study (See Table 2).

Insights from the interview. We outlined the main challenges of editing with LLM-based AI assistance in the OHMD mobile context:

Global Editing Challenges and Needs: Global editing involves adjusting the tone, writing structure, and length of emails. However, users often encounter difficulties due to discrepancies between their intended tone and length and the AI's understanding. One issue is that altering the global features, be it language alone or dialogue, can be challenging due to imprecision in natural human speech and the mismatch in intention captured by LLM. We need to strike a balance in addressing this. Another issue is that despite users instructing the AI to make the content casual or friendly, generated content often remains overly formal and wordy. Furthermore, when users aim for concise emails, the AI tends to produce longer responses than desired. This mismatch requires users to invest extra effort in editing, either by deleting or adding content. Therefore, users need easy-to-use global adjustment interactions.

Local Editing Challenges and Needs: 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). AI assistants struggle to accurately identify the email's structure, often leading to the misplacement of new content. Adjusting sentence-level order, especially within paragraphs, becomes a time-consuming process. Users express a need for enhanced control over word choice and arrangement, as well as improved contextual comprehension to accurately incorporate names and context. Challenges also arise when merging separate sentences into coherent paragraphs, which frequently results in unnecessary line breaks. Users envision a system that can deduce their editing requirements based on descriptions, allowing for a more efficient editing process with reduced workload.

4.5 Final Version of GlassMail

Combining insights from the observation study with our previous findings, we designed the final version of GlassMail (See Figure 7). The final version of GlassMail builds upon the first version and includes improved editing schemes for achieving final personalization with LLM (D3). To enable ease and accuracy in email writing to achieve personalization with a mobile OHMD LLM-based assistant and minimize multitasking workload we adopted the following design choice (see Table 1):

4.5.1 ***Achieving Personalization through Few-shot Learning.*** Our findings from the observational study indicate that accommodating users' personal writing styles is essential for an LLM-assisted email system. We found that users often want to modify email tone, length and email structure features such as greetings, openings, content order, closings or signatures, as they believe these aspects represent their personalization. This is also indicated by previous work[42, 88, 101]. According to Robertson et al.,[88], it is not only the email content but also the social context that

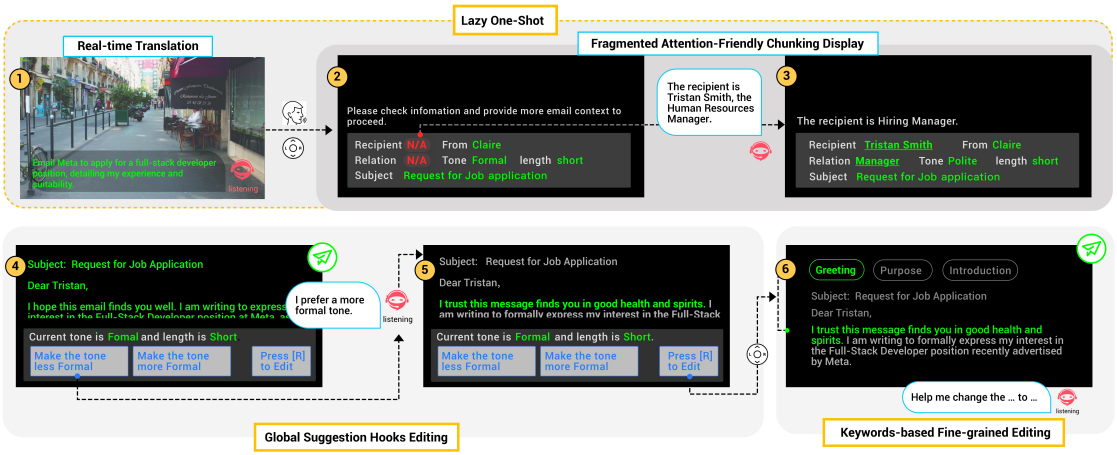


Fig. 7. The final interface of GlassMail also features the lazy one-shot approach to interact with the LLM-based assistant for composing emails and utilizes the Fade Context mode with optional audio as the output mode for displaying editing changes. It adds two distinct editing schemes: (4-5) suggestion hooks for facilitating global adjustments and (6) keywords-based editing for local edits.

should be considered in AI-assisted email systems. The main themes for email content personalization include structural features, personal authenticity, and semantic and tonal coherence.

GlassMail is powered by GPT-4 and it employs few-shot prompting methods [69, 115] for adaptation. We switched from GPT-3.5 to GPT-4 because it can handle analysing more complex user writing styles, including writing structure. When the user provides the initial email context, the system constructs a custom few-shot learning prompt and sends it to the language model to learn the user's personal writing styles. Writing effective prompts requires rigorous testing and iteration to achieve reliable and accurate responses from the model. The performance of the prompt is highly sensitive to word choice, formatting, and the content of the exemplars [115].

The GPT model tended to 'hallucinate' [112] additional information that did not exist in the original email context. Additionally, due to the model's limited context window [17, 23], which accepts prompts with a finite length, we could include just one exemplar to demonstrate the email task, ensuring consistent results. To overcome these challenges and also balance the tradeoff between mobile multitasking workload and generating the desired personal email content, we break up the email creation task via a prompt-chaining process [42, 58]: with the output of one step being the input to the next. We proposed a three-stage approach for email content creation in the mobile context to minimize the mobile multitasking workload and improve each task's accuracy: pre-, during, and post-creation (See details prompts in Appendix A). Here is an example that illustrates how GlassMail adapts to user instructions to create a personalized email in the desired tone and content structure, compared to the default LLM response (see Figure 8).

Based on the insights obtained from the study, we developed two distinct editing schemes in GlassMail, one for facilitating global adjustments while crafting emails with LLM-based assistance and the other for local edits. These two editing schemes are achieved by prompt engineering (see Appendix A for detailed prompts). We describe each editing scheme in detail.

subsubsection**Global Editing Schemes** Global editing is required to correct mismatches in desired tone, structure, and verbalisms, exacerbated by human speech imprecision and the AI's inclination towards formal or verbose email content. While personalisation reduces, to address these issues, GlassMail provides suggestion hooks that users can

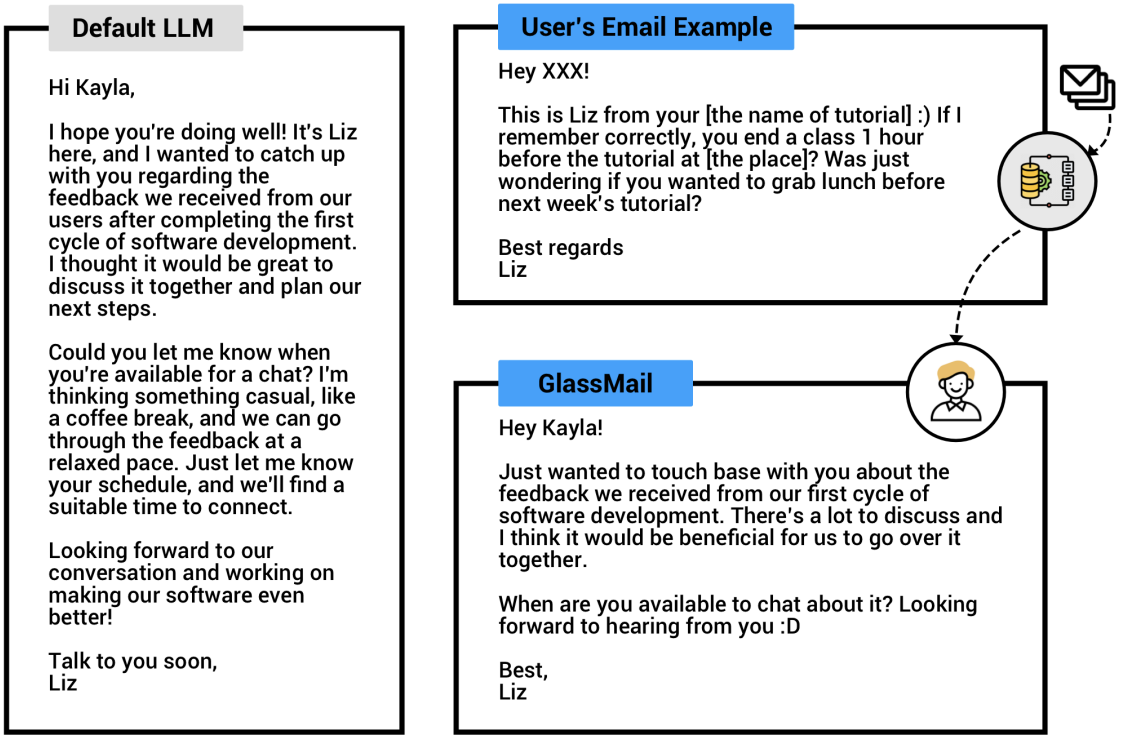


Fig. 8. An example showing how the responses of GlassMail and the default LLM differ when given the same user instruction. The user instruction is 'Hi please send an email to my project member Kayla. We just finished our first cycle of software development and we received feedback from the users. I want to discuss the feedback with her. Please make the tone casual and ask Kayla when she is available.'

simply speak out to facilitate easier and quicker global adjustments to bridge the gap between the AI's understanding and the user's intention. **Suggestion hooks** (see Figure 7) are user-friendly, quick commands that are generated based on the current email context and user sample emails. These hooks simplify the process of indicating user preferences to the assistant and help users quickly make global adjustments within the current system. For example, if a user wants to make the email tone more casual, they can simply use a suggestion hook like "change the tone to be more casual" to convey this preference to the system, without the need to specify the exact changes in the email contents as the system already learn their writing styles and preferences. This concept draws inspiration from "Contextualized help" [25, 54]. This approach ensures that help content revolves around actions that are relevant and available in the current context, rather than offering general information (e.g., suggestion hooks, as seen in Figure 7, for adjusting email tone or length based on the current context and email history). Once an email is generated, GlassMail shows the LLM-based assistant's current perception of the email, such as its current tone or length. This helps users to better direct their subsequent prompts for adjustments.

4.5.2 ***Local Editing Schemes.*** Local adjustments in mobile content, where traditional tools are absent, present challenges in content placement, sentence modification, and contextual understanding. AI's frequent misinterpretation

of email structure further complicates this, leading to a demand for improved content control and contextual accuracy. GlassMail's solution is **"Keyword-based Editing"** (see Figure 7). GlassMail enhances local editing efficiency using keywords extracted from 1-3 relevant email sentences. These keywords act as location markers. When selected, associated sentences are highlighted in the 'Fade context' mode. This eliminates the need for manual word or sentence selection, streamlining dialogue interactions and content order modifications. Specifically, it could address the following challenges.

(1) **Contextual Understanding**: Voice-based email editing with AI can lead to inaccuracies. Users often resort to repeating entire sentences or specifying locations, like "the fourth sentence of the second paragraph". This is not only time-consuming but also cognitively taxing. GlassMail's keyword-based shortcuts reduce this burden. By invoking a keyword using the ring mouse, the system grasps the modification context, aiming for more accurate and context-aware edits.
(2) **Content Order Modification**: Adjusting content order, particularly within paragraphs, is traditionally time-consuming using voice-based interaction. GlassMail simplifies this with keyword-based editing. By voicing keywords like "Time inquiry" or "Greeting", users can swiftly select and rearrange sentences. For instance, users can command the agent to place the "Time inquiry" after the "Greeting". This method offers a streamlined and intuitive approach to content reordering during voice interactions.
(3) **Navigational Control**: Identifying an email's structure from voice input poses challenges for LLM, often resulting in content misplacement or disrupted structure. Users often need to specify locations, like "the second and fourth sentence of the last paragraph". GlassMail counters this with a ring mouse and keyword editing, granting users simplified navigational control. By invoking a keyword using ring mouse, it simplifies navigational structure from 2D to 1D and reduces the space of continuous navigation with many possibilities to a few commonly used possibilities. Users can precisely pinpoint the sections they want to edit and make adjustments with precision, reducing the likelihood of content misplacement.

5 REAL-WORLD USABILITY STUDY

We conducted a usability study to evaluate the effectiveness and user satisfaction of using GlassMail to create and send emails in real-world mobile scenarios. We chose this evaluation method over a comparative study with smartphones as our preliminary formative study indicated that smartphones offer poor user experiences in real-world mobile scenarios, making them an inadequate baseline for meaningful comparison with GlassMail's capabilities.

5.1 Participants

We recruited 8 participants (5 females, 3 males) between 20-24 years old ($M = 21.3, SD = 1.28$) from the university community. All participants had a normal or corrected-to-normal vision with no colour deficiency and were fluent in English at the university level. One of them had prior experience using OHMDs. Five of them reported that they write an average of three emails per week on their mobile phones. Participants were compensated at the standard rate of US$7.5 per hour.

5.2 Materials

During the observational study, participants indicated that they seldom or never send emails to close friends or family members on the go, preferring instead to message them directly. Although Table 1 shows a higher number of global

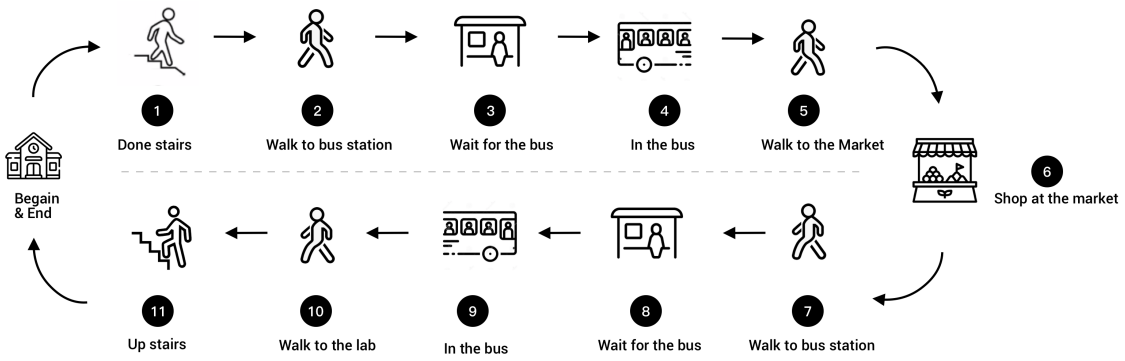


Fig. 9. Details of realistic mobile tasks: Participants start by taking a bus to a supermarket to purchase items and then go back. It includes three different mobile tasks: (a) There are three scenarios included on the way to the bus station (1)going downstairs, (2) walking indoors, and (3) walking outdoors. b) On the bus, the participants will 4) either sit or stand. After alighting, they will 5) walk to the supermarket. c) Shopping in the supermarket requires searching for the desired items and checking out at the cashier. These are all accomplished in the 6) indoor scenario by themselves while walking indoors (7-11) then they will take a bus and go back.

edits in personal emails to friends and family, this is likely due to our observation study's requirements. Participants needed to continually edit the generated emails, which were not initially satisfactory, until the content closely matched their own writing style, thereby explaining the increased need for edits. We believe that business-related emails are the preferred type of email for people to focus on while mobile, as the content is often more sensitive, timeliness, and they are less familiar with drafting such emails. In Contrast, while changing the tone in emails to family is common, family members are more tolerant of mistakes. We chose to focus on the more challenging case for our study.

We thus centered our study on three realistic email scenarios: communications with supervisors, and colleagues, and business-related interactions (e.g., order inquiries, and job applications). Before the study, participants provided three sample emails from these scenarios to help our system learn their writing style. They then identified three typical tasks per scenario and used GlassMail to compose these nine tasks in real-world mobile settings.

5.3 Real-world mobile tasks

We designed a route (See Figure 9) consisting of three realistic mobile scenarios where people are more likely to send emails based on our formative study interviews: On the way to or from work/lecture while using public transportation, and shopping at the supermarket. The route included mobility tasks such as sitting, standing, walking, and climbing up/down stairs in indoor, outdoor, and transport locations. On average the route took one hour to complete.

5.4 Apparatus and Procedure

The study used the final version of GlassMail. Participants began by signing a consent form and providing demographic information. They then received hands-on training with the system, including two example tasks. Next, participants wrote and edited emails using GlassMail while completing the route. They were asked to continue editing each email until they deemed them suitable for sending. After the study ended, they reviewed all sent emails to assess whether the quality matched their expectations. Finally, participants completed a post-questionnaire and answered open questions about their experience with GlassMail and mobile phones.

5.5 Results and Discussion

We measured the number of email tasks completed and a SUS survey [50] to evaluate the interface's usability. We also measured the overall experience of collaborating with the current system using the 7-point Likert scale in terms of satisfaction, confidence, ownership and quality of content [52].

Usability of GlassMail. Overall, GlassMail facilitated the successful creation and personalization of email content for users in real-world mobile settings. Participants managed to complete an average of 8.5 email tasks ($SD = 0.75$) during the route and provided an average SUS score of 75.31 for the GlassMail system. In terms of their experience collaborating with GlassMail, users reported feelings of ownership ($M = 5.75, SD = 0.89$), email writing task efficiency ($M = 6.13, SD = 0.84$), trustworthiness ($M = 6.13, SD = 0.99$), satisfaction ($M = 6.13, SD = 0.64$), and confidence ($M = 5.75, SD = 0.71$). Notably, five participants favoured GlassMail over mobile phones for drafting personal emails in mobile scenarios, appreciating the newfound flexibility it offers. The remaining three participants found both platforms to have unique merits.

GlassMail vs. phones for email writing. Participants unanimously recognized the primary advantage of GlassMail over mobile email writing as its speed in composing messages while ensuring grammatical correctness, and adapting to users' writing styles. As P8 mentioned, "[about GlassMail] I think it's much better than phones as I don't have to write the mail from scratch, thinking about the sentences and how to structure it. Also, the writing style [of GlassMail] is very close to my writing style". This allowed users to focus on making minor edits to the provided content.

Participants also found the system to be satisfactory in making necessary edits using keyword-based editing, showing that the feature supports scope adjustment for editing. While GlassMail was able to handle a variety of phrase or sentence-level edits, users occasionally faced difficulties either because the system failed to understand users' intentions (P3, P4, P8) or due to incorrect speech recognition in a noisy environment For instance, P3 asked the system to "change team activity to meetings" but the system might not make the change, due to LLM's instability, it may encounter technical issues or malfunctions, resulting in its inability to execute the user's request or inability to accurately determine the user's true intent, increasing the time needed to send out the email. Most participants also found the ability to switch to audio modality to be particularly useful in attention-heavy situations such as while climbing down stairs or when the OHMD had very low visibility due to high external lighting levels.

In contrast, participants expressed that mobile phones were challenging to use for similar tasks, requiring high amounts of attention, and making it difficult to remain aware of their surroundings. Creating email content from scratch, especially determining suitable subjects, proved cumbersome. However, the mobile phone system offered advantages over the smart glass system in some aspects such as better screen brightness (P1) and good visibility in various lighting conditions. Fine-tuning and making small edits were also easier on mobile phones (P3, P5, P7).

GlassMail's Boundary in Email Complexity. GlassMail is designed to handle a range of email complexities, though its effectiveness varies depending on the length and complexity of the email content. At the simplest level, emails consisting of just a single sentence or a brief reply may not necessitate the use of GlassMail. As P3 and P6 noted, such emails can be efficiently handled by more conventional methods (e.g., direct speech or manual typing on smartphones). GlassMail strength lies in composing emails of medium complexity that span a few paragraphs such as detailed replies, setting up meetings, or summarizing information. In these cases, GlassMail's features like lazy one-shot interaction, suggestion hooks, and keyword-based editing come into play, offering significant advantages in terms of efficiency and user experience. However, when it comes to highly complex emails, particularly those that

are lengthy and structurally intricate, GlassMail's effectiveness diminishes. For example, P2 struggled with sending detailed interview notes of three candidates to HR. P2 found it particularly challenging to align details with the desired structure and perform tasks like adding break lines or group content in a lengthy email. These complex tasks often demand extensive editing, formatting and nuanced language, which can surpass GlassMail's capabilities in a mobile context. GlassMail is thus ideal for emails that are neither overly simple nor excessively complex, aligning with the majority of typical email communications [45, 56, 65]. For emails at the extreme ends of complexity, alternative or traditional methods (e.g., desktop typing) might be more suitable.

6 OVERALL DISCUSSION

We began our exploration with the vision of developing an intelligent digital personal assistant, with which we can seamlessly collaborate to craft more complex textual content while multitasking or on the move. Our studies show that, by leveraging wearables OHMDs and Large Language Models (LLMs), such a vision is attainable but user interaction needs to be carefully designed.

6.1 Design Guidelines

AI has made considerable progress in creating personalized content, especially with few-shot learning techniques. However, it still struggles to perfectly capture personal styles and detailed editing nuances. Our studies reveal that after learning from a single user example, AI aligns with about 80-85% of a user's style in generated content. This significantly lessens the initial detailed editing required compared to traditional methods. AI's inability to fully grasp individual styles necessitates additional user-driven further editing. The problem is that in on-the-go scenarios editing itself demands substantial attention and may also conflict with the hands-free, head-up principles. The key takeaway from the whole lesson is the use of AI to quickly and effectively the big pictures (e.g., user's needs and personal style) in one shot as much as possible, aiming to minimize the need for subsequent editing. Also, designing convenient interaction schemes for editing is crucial. These should cater to both global and local editing requirements, thus streamlining the process in various contexts. Based on the results from our studies, we summarize the following design guidelines for developing human-AI collaborative interfaces on wearable displays:

- **Being transparent about how an AI system understands users' needs is crucial.** GlassMail employs a 'lazy one-shot' interaction approach for its agent. This method acknowledges that users naturally tend to provide as much information as they initially can, yet it also recognizes that they might inadvertently omit details. To address this, the approach allows the user to provide as much information as they can. Then it uses 'Fragmented Attention-Friendly Chunking' to reveal the AI system's understanding to users and request users to clarify or correct errors and gather additional information in a lazy manner. Our findings indicate that this strategy is effective in quickly identifying and correcting errors and in reducing the likelihood of the AI system misunderstanding user intentions.
- **Personalization through Common Ground.** In the case of applications such as email writing where a user's personality is expressed, it is important for the generated output to be personalized to the user.GlassMail uses both the context and the background knowledge of the user, as well as the task domain, to create personalized emails for each user. Our approach noted the importance of building common ground throughout the interaction. This common ground then serves as a context to guide interactions with the user and to limit the content generated by LLMs. GlassMail establishes this common ground by taking into account users' needs gathered

through conversation and inferences drawn from a single email example. It then requests users to clarify information of the common ground before generating content.

- **Fragmented-attention friendly output methods.** We are facing the general problem of enabling users to seamlessly process digital information without affecting their primary task in the OHMD mobile context. The design we are coming up with is that we allow users to modify without sacrificing their attention and primary task performance too much. GlassMail has developed fragmented-attention-friendly output methods by chunking schemes for getting missing information and Fade context mode for easy understanding of the edits made by the LLM. When users have very little attention, provide support for other modalities such as audio.
- **Minimizing Cognitive Load by Task Segmentation.** In mobile contexts, user-driven editing demands significant attention and cognitive resources. GlassMail addresses this challenge by dividing the editing task into two distinct interaction mechanisms effectively reducing cognitive strain. These mechanisms are Suggestion Hooks which provide users with contextually relevant suggestions, enabling them to make quick and effective global edits without extensive effort. Keyword-Based Editing: this allows users to make local edits by simply using keywords, streamlining both global and local editing needs and eliminating the need for extensive navigation or mental activities like typing. The key insights are to break up tasks into more manageable sub-tasks, to require less cognitive load, and to allow users to recall information and functionality more easily.

6.2 Human-AI collaboration to facilitate input in fragmented-attention scenarios

An essential aspect that we need to account for to support mobile formal content creation with LLMs is the fragmented attention span users have in mobile environments, which makes it difficult to rely on vision-heavy interaction mechanisms such as virtual keyboards [46, 66]. In this regard, voice presents a viable alternative and voice-based strategies such as commanding or redictating have been extensively explored for on-the-go text editing in the past [39, 41]. Although such voice-based strategies represent an improvement over using mobile phones for the same purpose, such a first-person style interaction paradigm wherein users actively execute tasks themselves [12], can still impose a considerable cognitive burden for complex text creation tasks in mobile environments.

Our findings show that employing an intelligent agent such as LLMs to interpret high-level voice commands facilitates a shift to a second-person style interaction paradigm [12] that can support more complex PIM tasks. This paradigm allows users to describe desired outcomes more broadly and delegate them to the agent, relieving them from concentrating on minute details and thereby easing the task of content creation while on the move. Field tests confirm that with this approach, users can indeed compose intricate, formal emails while on the go without affecting the primary task much, and not just short, informal texts meant for social media platforms such as Twitter.

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. Moreover, 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.

6.3 Wearable Displays to facilitate output in fragmented-attention scenarios

It is also important to make it easy for users to retrieve the created information in these visual attention-deficit mobile situations. We leveraged prior work on eyes-free interaction to relay the created content through the audio modality using TTS [7, 92]. The TTS feature, however, was found to be beneficial only in certain situations such as the case when the OHMD display had very low visibility in bright outdoor settings or in high visual demand scenarios such as walking on stairs. In other cases visual cues were still preferred, highlighting the need to incorporate the Fade Context visual output mode we identified from our studies together with other attention-efficient display principles proposed for OHMDs[87, 109].

6.4 Other Challenges with Using Current AI Technology for Wearable Personal Assistants

Our research also recognizes that current AI technology cannot fully meet wearable text creation needs due to a number of issues.

"That doesn't sound like me". It was apparent from our initial testing that users were concerned about losing their unique writing styles when creating emails through LLMs, fearing it would result in negative reception from recipients. To address this, we employed prompt engineering, utilizing a single sample of a user's email to generate more individualized content. The approach centred on two fundamental elements identified in our research: 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, we established the context for drafting new emails, effectively retaining users' individual writing preferences. This attempt was largely successful, with all eight participants in the field study finding that the emails they sent out using GlassMail retained their writing, achieving both satisfaction with the final content ($M = 6.13, SD = 0.64$) and confidence in sending them ($M = 5.75, SD = 0.71$). We anticipate that the incorporation of continual learning approaches could further enhance the preservation of personal writing styles.

The need for robust Speech-recognition. The speech recognition module in GlassMail is an important bridge in the second-person style voice interaction paradigm. Errors in speech recognition get propagated into the content generated by LLMs which then becomes difficult to edit out. During our field study, we noticed that these occurrences occur primarily in crowded noisy environments like inside a bus and due to variations in users' accents. We expect these challenges to phase out soon with the rapid advancements in automatic speech recognition models.

In summary, while AI forms a fundamental pillar in our vision towards a digital personal assistant, it alone cannot bridge the gaps. Harmonizing AI with the right human-computer interaction design principles in both input and output domains is vital in bringing the personal assistant vision to fruition. We hope that the insights derived could potentially extrapolate to other spheres of PIM such as calendar scheduling, and crafting memos and pave the way towards a more advanced personal assistant ecosystem that can cater to an array of PIM tasks ubiquitously.

7 LIMITATION AND FUTURE WORK

As a proof-of-concept prototype, GlassMail was designed for personal email creation in the OHMD mobile scenario. However, GlassMail does not provide all the functionalities necessary to address the identified challenges of mobile email creation with an LLM-based assistant. Currently, the system lacks the ability to create 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.

As for the real-world challenges, one is our current speech recognition features are unable to accurately extract a user's voice in real time. For instance, a user's voice recognition may be interfered with by noisy background sounds, such as someone talking nearby or a bus announcing a stop. Additionally, when a user speaks, we send a translation request for real-time translation once per second, which may experience delays due to network conditions. These limitations can lead to increased error rates in user interactions with the system and may impact the user's experience and trust in the system, especially for systems that rely exclusively on voice input.

Another real-world challenge is how to address environmental noise interference and potential issues related to the privacy [26] of GlassMail in public settings. For example, a user's speech recognition may be interfered with by noisy background sounds, such as someone talking nearby or a bus announcing a stop. 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. Additionally, future work could utilize participant feedback to identify needed new features. For example, participants (n=3) indicated that the microphone could be integrated into the ring mouse or smart glasses to allow a better experience without wearing other extra devices.

8 CONCLUSION

GlassMail is designed to bridge the gap between the user's workload for mobile email creation and the limitations of using mobile devices to achieve it. Results show that GlassMail enables users to create emails with high-quality personalized content while minimizing their workload in realistic mobile contexts. The prototype garnered positive feedback from all of our participants and demonstrated promising potential to leverage LLM capabilities as a wearable digital assistant for mobile personal information management. We believe that GlassMail has the real potential to stimulate future work in the field of designing intelligent wearable personal information management systems. This would also be another significant step toward the ultimate goal of creating an intelligent wearable personal assistant.

REFERENCES

[1] 2022. Apple iPad mini. https://www.apple.com/sg/ipad-mini/. Accessed: 2022-09-15.

[2] 2022. Nreal Light. https://www.nreal.ai/light/. Accessed: 2022-09-15.

[3] 2023. MacBook Air (M1, 2020) - Technical Specifications. https://support.apple.com/kb/SP825?locale=en_US. Accessed: 2023-9-15.

[4] 2023. Ring mouse Bluetooth connection 5 buttons USB rechargeable 400-MABT156BK available at Sanwa Direct. https://direct.sanwa.co.jp/ItemPage/400-MABT156BK. Accessed: 2023-9-15.

[5] Julius Adorf. 2013. Web speech API. *KTH Royal Institute of Technology* 1 (2013).

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

[7] Xavier Anguera, Nestor Perez, Andreu Urruela, and Nuria Oliver. 2011. Automatic synchronization of electronic and audio books via TTS alignment and silence filtering. In *2011 IEEE International Conference on Multimedia and Expo*. IEEE, 1–6.

[8] 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.

[9] 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.

[10] Nikola Banovic, Varun Rao, Abinaya Saravanan, Anind K Dey, and Jennifer Mankoff. 2017. Quantifying aversion to costly typing errors in expert mobile text entry. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4229–4241.

[11] Stuart J Barnes and Sid L Huff. 2003. Rising sun: iMode and the wireless Internet. *Commun. ACM* 46, 11 (2003), 78–84.

[12] Michel Beaudouin-Lafon. 2004. Designing interaction, not interfaces. In *Proceedings of the working conference on Advanced visual interfaces*. 15–22.

[13] Eden Bensaid, Mauro Martino, Benjamin Hoover, and Hendrik Strobelt. 2021. Fairytailor: A multimodal generative framework for storytelling. *arXiv preprint arXiv:2108.04324* (2021).

[14] Ofer Bergman, Richard Boardman, Jacek Gwizdka, and William Jones. 2004. Personal information management. In *Chi'04 extended abstracts on human factors in computing systems*. Citeseer, 1598–1599.

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

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

[17] Miriam Brinberg, Nilam Ram, Jinping Wang, S Shyam Sundar, James J Cummings, Leo Yeykelis, and Byron Reeves. 2023. Screenertia: Understanding "stickiness" of media through temporal changes in screen use. *Communication Research* 50, 5 (2023), 535–560.

[18] 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.

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

[20] 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.

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

[22] Dimitris Chatzopoulos, Carlos Bermejo, Zhanpeng Huang, Arailym Butabayeva, Rui Zheng, Morteza Golkarifard, and Pan Hui. 2017. Hyperion: A wearable augmented reality system for text extraction and manipulation in the air. In *Proceedings of the 8th ACM on Multimedia Systems Conference*. 284–295.

[23] Guanling Chen and David Kotz. 2000. A survey of context-aware mobile computing research. (2000).

[24] John Joon Young Chung, Wooseok Kim, Kang Min Yoo, Hwaran Lee, Eytan Adar, and Minsuk Chang. 2022. TaleBrush: visual sketching of story generation with pretrained language models. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–4.

[25] 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.

[26] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. " What can i help you with?" infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th international conference on human-computer interaction with mobile devices and services*. 1–12.

[27] Vivek Dhakal, Anna Maria Feit, Per Ola Kristensson, and Antti Oulasvirta. 2018. Observations on typing from 136 million keystrokes. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.

[28] Guillaume Erétéo, Michel Buffa, Fabien Gandon, and Olivier Corby. 2009. Analysis of a real online social network using semantic web frameworks. In *The Semantic Web-ISWC 2009: 8th International Semantic Web Conference, ISWC 2009, Chantilly, VA, USA, October 25-29, 2009. Proceedings 8*. Springer, 180–195.

[29] 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.

[30] Julia Fiksinski. 2021. *Practica: A Music Education Application for Learning Jazz Improvisation*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[31] Karen L Fingerman, Elizabeth L Hay, and Kira S Birditt. 2004. The best of ties, the worst of ties: Close, problematic, and ambivalent social relationships. *Journal of Marriage and Family* 66, 3 (2004), 792–808.

[32] 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*.

[33] Tsu-Jui Fu, William Yang Wang, Daniel McDuff, and Yale Song. 2022. DOC2PPT: automatic presentation slides generation from scientific documents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 634–642.

[34] Ruti Gafni and Nitza Geri. 2015. Evolving consumption patterns of various information media via handheld mobile devices. *Issues in Informing Science and Information Technology* 12 (2015), 083–093.

[35] Boni García, Francisco Gortázar, Micael Gallego, and Andrew Hines. 2020. Assessment of qoe for video and audio in webrtc applications using full-reference models. *Electronics* 9, 3 (2020), 462.

[36] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for science writing using language models. In *Designing interactive systems conference*. 1002–1019.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] Kunihiko Higa, Olivia R Liu Sheng, Bongsik Shin, and Aurelio Jose Figueredo. 2000. Understanding relationships among teleworkers'e-mail usage, e-mail richness perceptions, and e-mail productivity perceptions under a software engineering environment. *IEEE Transactions on Engineering Management* 47, 2 (2000), 163–173.

[46] Sébastien Hillaire, Anatole Lécuyer, Gaspard Breton, and Tony Regia Corte. 2009. Gaze behavior and visual attention model when turning in virtual environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. 43–50.

[47] 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).

[48] Ellen Jiang, Kristen Olson, Edwin Toh, Alejandra Molina, Aaron Donsbach, Michael Terry, and Carrie J Cai. 2022. Promptmaker: Prompt-based prototyping with large language models. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–8.

[49] Xinhui Jiang, Yang Li, Jussi PP Jokinen, Viet Ba Hirvola, Antti Oulasvirta, and Xiangshi Ren. 2020. How we type: Eye and finger movement strategies in mobile typing. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.

[50] Brooke John. 1996. SUS: a" quick and dirty" usability scale. *Usability evaluation in industry* (1996), 189–194.

[51] William P Jones and Jaime Teevan. 2007. *Personal information management*. University of Washington Press.

[52] 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.

[53] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, et al. 2016. Smart reply: Automated response suggestion for email. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 955–964.

[54] Adi Katz. 2012. Enhancing computer mediated communication by applying contextualization to email design: a case study. *Management* 64 (2012).

[55] Florian Kern, Florian Niebling, and Marc Erich Latoschik. 2023. Text Input for Non-Stationary XR Workspaces: Investigating Tap and Word-Gesture Keyboards in Virtual and Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2658–2669.

[56] Elaine B Kerr and Starr Roxanne Hiltz. 2013. *Computer-mediated communication systems: Status and evaluation*. Academic Press.

[57] 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.

[58] 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.

[59] Tae Soo Kim, DaEun Choi, Yoonseo Choi, and Juho Kim. 2022. Stylette: Styling the web with natural language. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

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

[61] Pin-Sung Ku, Yu-Chih Lin, Yi-Hao Peng, and Mike Y. Chen. 2019. PeriText: Utilizing Peripheral Vision for Reading Text on Augmented Reality Smart Glasses. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 630–635. https://doi.org/10.1109/VR.2019.8798065

[62] Jennifer Lai and Nicole Yankelovich. 2006. Speech interface design. (2006).

[63] 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.

[64] Florian Lehmann, Niklas Markert, Hai Dang, and Daniel Buschek. 2022. Suggestion lists vs. continuous generation: Interaction design for writing with generative models on mobile devices affect text length, wording and perceived authorship. In *Proceedings of Mensch und Computer 2022*. 192–208.

[65] Yili Li. 2000. Linguistic characteristics of ESL writing in task-based e-mail activities. *System* 28, 2 (2000), 229–245.

[66] Yang Li, Sayan Sarcar, Yilin Zheng, and Xiangshi Ren. 2021. Exploring text revision with backspace and caret in virtual reality. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–12.

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

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

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

[70] Andrés Lucero, Danielle Wilde, Simon Robinson, Joel E Fischer, James Clawson, and Oscar Tomico. 2015. Mobile collocated interactions with wearables. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. 1138–1141.

[71] Ewa Luger and Abigail Sellen. 2016. " Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5286–5297.

[72] Microsoft. 2023. *Microsoft PowerPoint Slide Presentation Software | Microsoft 365*. https://www.microsoft.com/en-sg/microsoft-365/powerpoint Accessed: 2023-09-01.

[73] Carolina Milanesi. 2016. Voice Assistant Anyone? Yes please, but not in public. *Creative Strategies* (2016).

[74] Paul Milgram, Herman Colquhoun, et al. 1999. A taxonomy of real and virtual world display integration. *Mixed reality: Merging real and virtual worlds* 1, 1999 (1999), 1–26.

[75] 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.

[76] 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.

[77] Kazuhiro Nakadai, Toru Takahashi, Hiroshi G Okuno, Hirofumi Nakajima, Yuji Hasegawa, and Hiroshi Tsujino. 2010. Design and Implementation of Robot Audition System'HARK'—Open Source Software for Listening to Three Simultaneous Speakers. *Advanced Robotics* 24, 5-6 (2010), 739–761.

[78] Waqas Nawaz, Kifayat-Ullah Khan, and Young-Koo Lee. 2016. A multi-user perspective for personalized email communities. *Expert Systems with Applications* 54 (2016), 265–283.

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

[80] 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

[81] Kseniia Palin, Anna Maria Feit, Sunjun Kim, Per Ola Kristensson, and Antti Oulasvirta. 2019. How do people type on mobile devices? Observations from a study with 37,000 volunteers. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.

[82] 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.

[83] Alisha Pradhan, Leah Findlater, and Amanda Lazar. 2019. " Phantom Friend" or" Just a Box with Information" Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–21.

[84] 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).

[85] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*. PMLR, 28492–28518.

[86] 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.

[87] Wyko Rijnsburger and Sven Kratz. 2017. Personalized presentation annotations using optical HMDs. *Multimedia Tools and Applications* 76 (2017), 5607–5629.

[88] 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.

[89] 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.

[90] Rufat Rzayev, Paweł W. Woźniak, Tilman Dingler, and Niels Henze. 2018. Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3173574.3173619

[91] 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.

[92] Davide Salvi, Brian Hosler, Paolo Bestagini, Matthew C Stamm, and Stefano Tubaro. 2023. TIMIT-TTS: a Text-to-Speech Dataset for Multimodal Synthetic Media Detection. *IEEE Access* (2023).

[93] Joao Santos, Joel JPC Rodrigues, João Casal, Kashif Saleem, and Victor Denisov. 2016. Intelligent personal assistants based on internet of things approaches. *IEEE Systems Journal* 12, 2 (2016), 1793–1802.

[94] 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.

[95] Johan Schalkwyk, Doug Beeferman, Françoise Beaufays, Bill Byrne, Ciprian Chelba, Mike Cohen, Maryam Kamvar, and Brian Strope. 2010. "Your word is my command": Google search by voice: A case study. *Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics* (2010), 61–90.

[96] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based text entry in virtual reality. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

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

[98] 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.

[99] TextMagic. 2023. *Free Text to Speech Online - TextMagic*. https://freetools.textmagic.com/text-to-speech Accessed: 2023-09-01.

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

[101] Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shaun K Kane. 2023. "The less I type, the better": How AI Language Models can Enhance or Impede Communication for AAC Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[102] 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.

[103] Bryan Wang, Gang Li, and Yang Li. 2023. Enabling conversational interaction with mobile ui using large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–17.

[104] 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.

[105] 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.

[106] 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.

[107] James R Williams. 1998. Guidelines for the use of multimedia in instruction. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 42. SAGE Publications Sage CA: Los Angeles, CA, 1447–1451.

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

[109] Qingguo Xu, Sen-ching Samson Cheung, and Neelkamal Soares. 2015. LittleHelper: An augmented reality glass application to assist individuals with autism in job interview. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, 1276–1279.

[110] Chun Yu, Yizheng Gu, Zhican Yang, Xin Yi, Hengliang Luo, and Yuanchun Shi. 2017. Tap, dwell or gesture? exploring head-based text entry techniques for hmds. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4479–4488.

[111] 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.

[112] Zhuosheng Zhang, Aston Zhang, Mu Li, Hai Zhao, George Karypis, and Alex Smola. 2023. Multimodal chain-of-thought reasoning in language models. *arXiv preprint arXiv:2302.00923* (2023).

[113] Shengdong Zhao, Felicia Tan, and Katherine Fennedy. 2023. Heads-Up Computing: Moving Beyond the Device-Centered Paradigm. *arXiv preprint arXiv:2305.05292* (2023).

[114] 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

[115] 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.

[116] 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.

[117] Lina Zhou, Ammar S Mohammed, and Dongsong Zhang. 2012. Mobile personal information management agent: Supporting natural language interface and application integration. *Information Processing & Management* 48, 1 (2012), 23–31.

A APPENDIX

We divided our prompts pipelines into three stages: pre (see Figure 10), during (see Figure 11, 12) and post (see Figure 13, 14).

```
export const exractEmailSettings = async (prompt, emailExample, handleResponse) => {
    const gptMessages = [{
            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: `Give the <UserRequirements>: ${prompt}>, you should extract the following information:\
            'Recipient, Subject, Relation, FromName, Tone, and Length'.\
            For any aspect you cannot extract from <UserRequirements>,\
            please infer or predict based on the provided <EmailExample>: ${JSON.stringify(emailExample)}.\
            If you are unable to predict or infer it from <UserRequirements> and <HistoryEmail>, please return 'N.A'.\
            The output should be a JSON format:
            {
                settings:{
                    Recipient: "Recipient's name",
                    Subject: "Suggest one if not provided in <UserRequirements>, within 5 words.",
                    Relation: "Keep the tone as a single word. Suggest the most suitable one based on the <UserRequirements>.",
                    Tone: "Keep the tone as a single word. Suggest the most suitable one for the Relation.",
                    Length: "Short(within 200 words), Medium(200 - 350 words), or Detailed(350 - 500 words)",
                    FromName: "Sender name placeholder (extract from <EmailExample>)",
                }
            }`
        }
    ]
    ...
}
```

Fig. 10. Pre Stage: extract Email settings based on user's initial requirement.

```
export const composeEmail = async (preOutput, prompt, emailExample, handleResponse) => {
    const gptMessages = [
        {
            role: 'user',
            content: `Given the <EmailRequirement>:${prompt} and <EmailSettings>${JSON.stringify(preOutput)}, update the <EmailSettings>.\
            Then you should strictly analysis and extract <WritingStyles>: greetings, openings, closings, signature,\
            and writing structure from the <EmailExample>: ${emailExample}.\
            For the writing structure, understand how the user writes the main topic of body in <EmailExample>, excluding other parts.\
            Compose an email strictly following the <WritingStyles> according to the latest <EmailSettings> and <EmailRequirement>\
            using the <Tone> in the latest <EmailSettings>.\
            You should also suggest two similar tones based on <EmailRequirement>, <Tone>, and <EmailExample>、
            and return them in the <EmailSettings>.\
            The output in the following JSON format and do not include any additional content:
            {
                subject: "Suggest one if not provided in the latest <EmailSettings>, within 5 words.",
                content: "Each paragraph should be separated by \n\n.",
                settings: "The latest <EmailSettings>. Keep the tone as a single word.",
                styles: "The <WritingStyles>",
            }
            `
        }
    ]
    ...
}
```

Fig. 11. During Stage: Compose the initial email.

```
export const regenerateEmail = async ({settings, previousEmail, writingStyles} = prevOutput, prompt, handleResponse) => {
    const gptMessages = [
        {
            role: 'user',
            content: `Given the <EmailRequirement>:${prompt}, update <EmailSettings>: ${JSON.stringify(settings)}>.\
            Compose a new email with <WritingStyles>:${JSON.stringify(writingStyles)}, tone and word count similar to the <PreviousEmail>:\
            ${previousEmail} according to the latest <EmailSettings>, <EmailRequirement>.\
            You should not return the same content as <PreviousEmail> to me.\
            The output in the following JSON format and do not include any additional content:
            {
                subject: "Suggest one if not provided in the latest <EmailSettings>, within 5 words.",
                content: "Each paragraph should be separated by \n\n.",
                settings: "The latest <EmailSettings>. Keep the tone as a single word.",
            }
            `
        }
    ]
}
```

Fig. 12. During Stage: Regenerate Email.

```
export const generateKeywords = async (currentEmail, handleResponse) => {
    const gptMessages = [
        {
            role: 'user',
            content: `Given the email: ${currentEmail}, you should categorize all sentences in the email into three to five major<Keywords>.\
            You should ensure each keyword in <keywords> is linked to 1-3 different sentences in the email.\
            Each sentence should be only associated with one keyword. You should not return me the keyword without associated any sentences.\
            The output should be a JSON file:
            {
                keywords: [
                    {
                        id: 0,
                        keyword: 'Each keyword should be summarized in 2 words.'
                        sentences: ['', '', '']
                    }
                ]
            }
            `
        }
    ]
    ...
}
```

Fig. 13. Post Stage: Generate keywords for current email.

```
export const editEmail = async ({currentEmail, currentKeyword, currentKeywords} = preOutput, prompt, handleResponse) => {
    const editScope = currentKeyword.sentences.join('')
    const gptMessages = [
        {
            role: 'user',
            content: `Given the <Email>:${currentEmail}, only edit the content in the <EditScope>${editScope} according to User's request:\
            ${prompt} and return the edited <Email>\
            If user want to add or modify content in <EditScope>, replace the new edited content in the index${currentKeyword.id}\
            of the <Keywords> ${JSON.stringify(currentKeywords)}\
            If user want to delete/remove the whole content in <EditScope>, delete the items: ${currentKeyword.keyword} from <Keywords>.\
            Then you should reset <Keywords> id make it start from 0.\
            The output should be a JSON format:
            {
                subject: '',
                content: 'Each paragraph should be separated by \n\n.',
                currentIndex: 'if the <EditScope> has been delete, return 0 else return the content's currentKeyword Id',
                keywords: [
                    {
                        id: 'must always start from 0',
                        keyword: 'Each keyword should be summarized within 2 words.'
                        sentences: ['', '', '']
                    }
                ]
            }
            `
        }
    ]
    ...
}
```

Fig. 14. Post Stage: Keyword-based editing.