Visual Behaviors and Mobile Information Acquisition

NUWAN JANAKA, National University of Singapore, Singapore
XINKE WU*, National University of Singapore, Singapore and Alibaba, China
SHAN ZHANG*, National University of Singapore, Singapore
SHENGDONG ZHAO, National University of Singapore, Singapore
PETR SLOVAK, King's College London, United Kingdom

It is common for people to engage in information acquisition tasks while on the move. To understand how users' visual behaviors influence microlearning, a form of mobile information acquisition, we conducted a shadowing study with 8 participants and identified three common visual behaviors: *glance*, *inspect*, and *drift*. We found that *drift* best supports mobile information acquisition. We also identified four user-related factors that can influence the utilization of mobile information acquisition opportunities: situational awareness, switching costs, ongoing cognitive processes, and awareness of opportunities. We further examined how these user-related factors interplay with device-related factors through a technology probe with 20 participants using mobile phones and optical head-mounted displays (OHMDs). Results indicate that different device platforms significantly influence how mobile information acquisition opportunities are used: OHMDs can better support mobile information acquisition when visual attention is fragmented. OHMDs facilitate shorter visual switch-times between the task and surroundings, which reduces the mental barrier of task transition. Mobile phones, on the other hand, provide a more focused experience in more stable surroundings. Based on these findings, we discuss trade-offs and design implications for supporting information acquisition tasks on the move.

CCS Concepts: • **Human-centered computing → Mobile devices**; **User studies**; **Empirical studies in ubiquitous and mobile computing**; • **Applied computing → E-learning**.

Additional Key Words and Phrases: Visual behaviors, Attention fragmentation, Mobile information acquisition, Microlearning on the move, Shadowing, HMD, Smart glasses, Mobile phones

1 INTRODUCTION

Think about the last time you commuted: how many times did you shift your attention away from your phone to navigate a busy street or look at a signboard?

---

*Authors contributed equally to this research.

Authors' addresses: Nuwan Janaka, nuwanj@comp.nus.edu.sg, National University of Singapore, NUS-HCI Lab, School of Computing, Singapore; Xinke Wu, xinke.wxk@antgroup.com, National University of Singapore, NUS-HCI Lab, School of Computing, Singapore, Alibaba, Z Space, No. 556 Xixi Road, Hangzhou, China, 310000; Shan Zhang, shan_zhang@u.nus.edu, National University of Singapore, NUS-HCI Lab, School of Computing, Singapore; Shengdong Zhao, zhaosd@comp.nus.edu.sg, National University of Singapore, NUS-HCI Lab, School of Computing, Singapore; Petr Slovak, petr.slovak@kcl.ac.uk, King's College London, Department of Informatics, Bush House, 30 Aldwych, London, United Kingdom, WC2B 4BG.

Visual attention is a critical resource for processing visual information on the move. This process includes filtering information that one receives, then selectively processing the content [18, 21, 22]. It is more difficult to perform information processing tasks with computing devices while on the move than in stationary settings as higher levels of attention are required for both the mobility task (e.g., walking) and mobile Human-Computer Interaction (HCI) task (e.g., reading an email) [68, 93]. Mobility tasks require users to focus on their surroundings in an effort to maintain situational safety as well as react to social or personal need-based cues, while mobile HCI tasks require users to maintain visual attention on their device [80, 82, 93]. Consequently, visual attention fragmentation occurs more frequently when users are on the move [68].

The topic of visual attention allocation (or, more broadly, visual behaviors [32, 40]) is an important area of research that has been extensively studied. Yet, previous investigations have mostly focused on visual behaviors associated with stationary settings [18, 55, 74]. Relatively few studies have concentrated on visual behaviors in mobile contexts [68, 83], and there has been a lack of categorization around visual behaviors in this context. In this study, we aim to deepen our understanding of visual behavior patterns on the move, as well as precise and in-situ mobile interaction designs to support mobile information acquisition tasks. While previous studies have investigated the fragmented nature of visual behavior in mobile settings [68], we aim to examine the different aspects of fragmentation, formally classify them, and investigate their influence on the effectiveness of information acquisition in mobile contexts.

We contextualized our investigation in a specific mobile scenario (commuting) and information acquisition task (microlearning vocabulary). Commuting was selected as it is a typical mobile situation in everyday life [52]. It is naturally accompanied by complex and dynamic external distractions (e.g., people moving, sudden noises), which influence on-the-move information processing [19, 79]. Since visual attention is often fragmented during on-the-move situations [68], we avoided using long/complex information acquisition tasks as they can hinder the resumption of ongoing (mobility) tasks [17, 64]. Instead, we focused on information acquisition tasks that consist of smaller tasks with fewer dependencies. Microlearning is well-suited for this purpose [27] as the microlearning technique divides complex learning tasks into small "bite-sized" sessions and integrates them into daily activities [34].

We identified three distinct visual behaviors resulting from attention fragmentation while commuting through a shadowing study: *Inspect*, *Drift* and *Glance*, based on the dimensions of *Purpose*, *Duration*, and *Perceived Visual Attention Intensity*. We studied how these patterns offer different opportunities for mobile information acquisition tasks on the move, specifically for microlearning during the commuting scenario. We found that *drift* presents the most suitable opportunity for mobile information acquisition. However, the utilization of visual behaviors largely depends on dynamic interruptions from the environment (*external interruptions*).

Mobile information acquisition and interactions are tied to specific devices. Mobile phones are the most common platform at present, but existing research shows that on-the-move information acquisition with mobile phones leads to fatigue and reduced learning gains [19, 50, 93, 95]. Optical see-through Head Mounted Displays (OST HMDs, OHMDs) or smart glasses are an emerging mobile interaction platform that have been shown to minimize the issue of split attention. This platform can provide peripheral information to users, reducing interferences between the surrounding environment and on-the-move mobile interactions [46, 59, 62, 81].

For more insight into visual interaction designs during on-the-move situations, we investigated how different devices (mobile phones and OHMDs) utilize *drift*s differently for second-language microlearning during commute through a technological probe. Results revealed that OHMDs enabled a better balance between information acquisition tasks and situational awareness and allowed commuters to utilize shorter *drift*s amid of frequent *glance*s for information acquisition.

On the other hand, mobile phones provided a more focused experience for mobile information acquisition when the surroundings were more stable and had fewer *external interruptions*.

Based on these findings, we discuss the trade-offs and design implications for supporting mobile information acquisition on the move, especially for microlearning on the commute, and propose a system that can utilize opportunistic visual behaviors in a more general setting.
Our contribution is twofold:

1) We establish commuters' visual behaviors with respect to three categories (*glance*, *inspect*, and *drift*) and identify their effects on mobile information acquisition (i.e., microlearning) opportunities. We discuss design implications for better supporting mobile HCI tasks.

2) We empirically evaluate the *receptivity* for mobile information acquisition on both mobile phones and OHMDs during a dynamic mobile context (e.g., commute) and identify the trade-offs of using both platforms. In doing so, we better understand how device platforms affect mobile information acquisition and missing interactions.

2 RELATED WORK

Our work relates to three broad areas.

2.1 Multitasking, attention fragmentation, and visual behaviors

There are multiple theories and frameworks regarding attention management and allocation across various tasks, such as Kahneman's resource theory [47], Wicken's multiple resource theory [91], and the Resource Completion Framework [68]. In each of these theories, attention is modeled as a finite (or elastic) resource in which multitasking can be cognitively, perceptually, physiologically, and socially costly [68, 93]. In multitasking scenarios, attention is simultaneously shared across different tasks [68, 91], which leads to attention fragmentation [68]. Despite the attention costs involved in multitasking, people continue to multitask with their mobile devices on the move [69, 89], suggesting that there is a strong demand for information acquisition on the move. We seek to investigate how visual attention is allocated as an essential step to supporting this demand.

To understand various aspects of attention in real-world mobile HCI tasks in an effort to design interfaces for limited attention spans, HCI researchers have investigated attention allocation on the device or environment based on task levels [15, 68, 83]. For example, establishing that users' attention span is 4 to 8 seconds on mobile devices [68] has helped designers to size information chunks accordingly such that information can be effectively consumed in a short duration or glance.

However, the current understanding of the topic (e.g., duration of attention fragments) has proved insufficient to guide detailed designs on when and how to present information to users in mobile scenarios [15, 83]. Our study investigates visual behavior patterns in terms of duration, purpose, and intensity when users are on the move. With that, we develop a better understanding of various visual behaviors and their opportunities for effective information acquisition.

2.2 Mobile HCI, information acquisition, and microlearning

It is common for people to engage in mobile HCI tasks while on the move and the majority of these tasks are for the purpose of information acquisition (e.g., reading articles, checking social network updates, watching videos) [35, 68, 77].

Microlearning is one such information acquisition technique that divides complex learning tasks into small and quick learning interactions distributed across time [10, 34]. Microlearning is commonly applied to language learning as it is relatively easy to break down language learning tasks such as vocabulary learning into smaller ones [20, 30, 34]. Various methods of improving mobile language microlearning such as adaptations to individual learners [29], spaced repetitions

[30, 90], multi-modal presentations of content [20], and contextual encoding [10, 26, 30, 86] have been explored in previous studies.

Investigators also have researched on *when* to present microlearning content, in which studies focused on identifying ubiquitous micro-moments during daily life when users are less engaged with their mobile devices [16, 17, 34, 76]. Some research has honed in on the internal factors (e.g., boredom [27]) that influence microlearning opportunities as a means of identifying suitable timing independent of tasks.

However, how visual attention influences microlearning opportunities remains an underexplored area. This work further undertakes this area of research by investigating the interplay between visual attention with opportunistic moments for microlearning, its implications so that this knowledge can be applied more broadly to general information acquisition tasks.

2.3 Influence of platforms on mobile HCI tasks

Mobile phones are the most commonly used platform for information acquisition tasks such as language microlearning [16, 20, 26, 30, 76, 90]. With the advancement in mobile technologies, users can access information anytime and anywhere without physical and social boundaries [65, 70]. However, mobility presents interactional challenges due to constant situational and contextual changes in the user's environment, and it is difficult for users to sustain such high levels of attention on HCI tasks [70, 78].

OHMDs, on the other hand, have emerged as a promising platform to support multitasking as users can maintain direct visual contact with their physical surroundings while performing mobile HCI tasks displayed on screen [59, 67, 73].

Both mobile phones and OHMDs can support mobile information acquisition tasks such as microlearning due to their portability, ease of access, and support of contextual encoding [30], though there are trade-offs to each platform. For instance, mobile phone users do not have as much access to their peripheral vision when interacting heads-down with their phones [62]; thus, any information displayed on the mobile screen is potential 'distraction' for their mobility tasks [31]. In contrast, OHMD users have better visual access to their environment, given that their visual headset is worn and their display remains perpetually in view [60]. Still, focusing on learning content on a see-through display can be more challenging than on a mobile screen, as the background can change.

This research intends to investigate the advantages and disadvantages of information acquisition opportunities offered by the two platforms.

3 OVERALL STUDY DESIGN

Multitasking on the daily commute is common and habitual for most [79]. As previously highlighted, a key issue with such behaviors is that it requires visual attention to split between navigational and mobile HCI tasks [68]. We unpack our understanding of this problem by first conducting an observational shadowing study (*study 1*), followed by a technology probe (*study 2*) on mobile phones and OHMDs, comparing on-the-move microlearning between the two platforms.

4 STUDY 1: UNDERSTANDING VISUAL BEHAVIORS DURING COMMUTING

In this first study, we focus on exploring the following research questions.

RQ1: What are the typical on-the-move visual behaviors when factoring for purpose, duration, and intensity? What are the observable characteristics of visual behaviors?

Previous work has shown that a person's attention shifts based on the purpose, duration, and perceived intensity of focus change [63]. We adopt these categories as dimensions in our investigation of visual behaviors.

Since purpose and perceived intensity are subjective in nature and cannot be directly measured, we focused on head movements (often used with gaze estimation) [88] as the observable trait associated with purpose and intensity.

RQ2: How can visual behaviors support information acquisition tasks such as microlearning?

There are many opportunities to divert attention to information acquisition tasks when a person is on the move. We explore the visual behaviors that support such information tasks and the effect of behavior switching.

4.1 Method

We first carried out a shadowing study (with video recording) to investigate real-life individual and social interactions in-situ and the effects of the surroundings [6, 72].

We also used contextual inquiries [56][Ch 8] to understand how participants can potentially use visual attention and how receptive they are towards diverting attention to microlearning. Following Isaacs et al. [45], we operationalized receptivity as the willingness to engage with microlearning. To identify the *instantaneous* and *in-situ receptivity* in dynamic commuting scenarios, we conducted contextual inquiries in a fashion similar to event-contingent experience sampling with verbal responses [23, 66]. To reduce interruptions to natural behaviors, each inquiry was at least 15-minutes apart and had a maximum duration of 3 minutes. Each participant was shadowed 2-3 times within the same day to increase the generalizability of the results.

4.1.1 Apparatus. As our focus is on task-level visual behaviors instead of micro-level eye movements, we used video recordings that were similar to those used by Oulasvirta et al. [68], but only with a head-mounted camera (weight ≈ 24g) and side camera, as shown in Fig 1. The head-mounted camera was for identifying the focus of users' attention, while the side view camera was for identifying attention switches (with spatial changes) during task engagement.
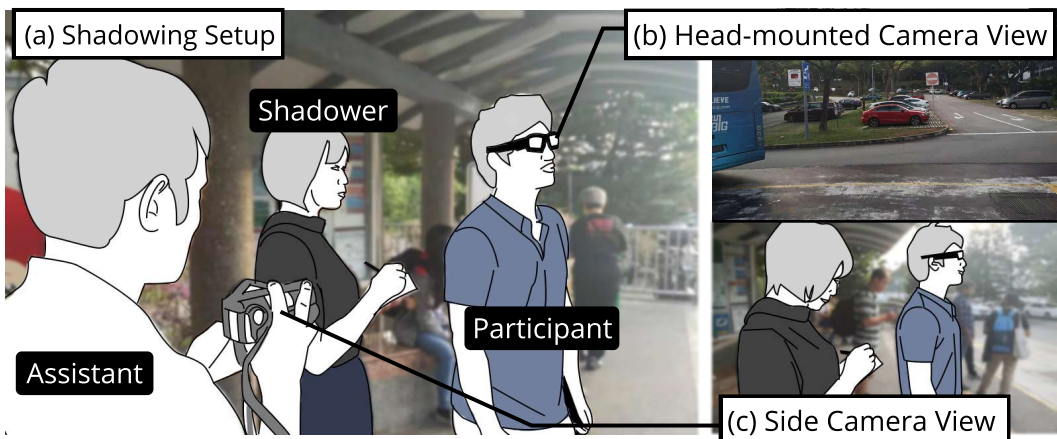

Fig. 1. Shadowing configuration (a) The shadower observes the participant who is wearing a head-mounted camera and takes notes, while the assistant records the side view of the participant (b) Head-mounted camera view from the head-mounted camera (c) Side camera view recorded by the assistant

Table 1. Participants' demographic information and shadowing settings in *study 1*

Participant	Gender, Age, Native language, Profession/Major	Language learning usage	Shadowing duration, Medium, Commuting purpose, Time of the day
P1	Male, 23, English, Undergraduate (Real Estate)	Use *Jisho* to learn Japanese	180 min, Metro/Bus/Walking, To & from School/Dining, Morning/Evening
P2	Female, 23, English, Undergraduate (Industrial and System Engineering)	Use *AnkiDroid* to learn Japanese. Do not use mobile on the bus/during walking due to motion sickness	220 min; Metro/Bus/Walking, To & from School/Dining/Exercise, Morning/Noon/Evening
P3	Male, 22, English, Undergraduate (Linguistics)	Used *Memrise* to learn Spanish	150 min, Bus/Walking, To & from School/Dining/Library, Noon/Evening
P4	Female, 22, Mandarin, Master Student (Finance)	Used *Shanbay* to practice English	140 min, Bus/Walking, To & from School/Dining, Morning/Evening
P5	Male, 28, Sinhale, Software Engineer (Electronics)	Used *Magoosh GRE* to practice English. Do not use mobile during the bus due to motion sickness	200 min, Bus/Metro/Walking, To & from Work/Dining/Shopping, Morning/Noon/Evening
P6	Female, 21, English, Undergraduate (Environmental Engineering)	Use *TenguGo Hangul* to learn Korean. Do not use mobile while walking	230 min, Metro/Bus/Walking, To & from School/Shopping, Noon/Evening
P7	Male, 28, Spanish, PhD Student (Design & Environment)	Use *Duolingo* to learn Mandarin	200 min, Metro/Bus/Walking, To & from School/Shopping, Morning/Afternoon
P8	Female, 21, English, Undergraduate (Economics)	Use *Duolingo* to learn French	240 min, Metro/Bus/Walking, To & from School/Dining/Shopping, Morning/Evening

4.1.2 Participants. Since we targeted populations with mobile information needs, we selected participants who were already personally motivated to perform information acquisition tasks on the move. Eight participants [P1-P8] (4 females, age $M = 23.5$, $SD = 2.8$) were selected based on their language learning experiences and commuting profiles (Table 1). All participants were formally educated in English, had been on two or more types of commute (metro/bus/walking), and had experience using mobile vocabulary learning apps, albeit not in all commuting scenarios.

We compensated each participant with $\approx$ USD 7.40/h for their time in **both** studies and received their informed consent before conducting the studies.

4.2 Procedure

Each participant was followed by a shadower and an assistant for 2-4 hours a day in different commuting scenarios, as shown in Fig 1 (a) and **Table 1 (column 4)**. After participants were briefed and asked for consent, they put on a head-mounted camera, which recorded a view of their vision field (Fig 1 (b)).

During the study, the shadower took observational notes, and the assistant recorded the side view of the participant, especially his/her head and hand movements (Fig 1 (a), (c)). When participants' behavior or context changed, the shadower noted down the time, changes in context, as well as their attention focus. Participants were asked a set of questions, including their receptivity (i.e., willingness) to microlearning (5-point Likert scale, 1 = Very Low, 5 = Very High), factors that affected their willingness to microlearning, and the perceived amount of visual attention they paid to the primary task (5-point Likert scale, 1 = Very Low, 5 = Very High). The monitored contexts and details of contextual inquiries are in Appendix A.5 and Appendix A.6.

At the end of the shadowing, the shadower carried out a 30-40 minute semi-structured interview and asked participants about the reasons behind the changes in their visual behaviors, how that or other factors influenced their receptivity to microlearning (refer to Appendix A.7 for interview topics). The entire interview was audio-recorded and later transcribed. Whenever required, such as when participants needed help to recall details, the assistant played them the relevant parts of the video recording.

4.3 Data Analysis

We conducted a mixed analysis (mainly qualitative) by triangulating four data sources: contextual inquiry notes from 112 inquiry sessions, interview transcriptions of 8 sessions, observation notes, and video recordings from approximately 26 hours of shadowing (see Appendix A.2 for details). Text data (i.e., contextual inquiry notes, observation notes, and interview transcriptions) contained the perceived behaviors, reasons, and contextual information. Video data (i.e., head-mounted camera view time-synced with a side view) consisted of head/eye movement data, attention focus, and timing information. Since the two forms of data showed different but complementary dimensions of visual behaviors, we used two different coding schemes initially and combined them with the themes later.

Using the *QDA Miner* software package, two researchers (co-authors and shadowers) independently performed open coding [24, 25] on **two** participants' text data and video recordings.

The two researchers then discussed and developed initial coding schemes, one for text data and another for video data. Then, they independently reanalyzed the same participants' data, resolved any disagreements, and refined the coding by discussing and re-watching video recordings. After analyzing another participant's data independently, the two coders reached 93% agreement on text data, but the video coding timing varied by 0-13s. This discrepancy happened because the purpose of visual attention was only captured during contextual inquiries but not in all video data.

The researchers used the resulting codes to independently analyze the remaining participants' data to develop themes and patterns by grouping codes. Two researchers watched the video recordings to reach an agreement whenever there was a discrepancy between the subsequent codes and themes. The data with discrepancies were reanalyzed using the codes that were agreed upon.

Videos were coded into datasheets by pausing and replaying them based on the agreed codes: time, posture, location, familiarity, crowd levels, and visual behavior (with an accuracy of one second - refer to Appendix A.1 for sample data). Observation and contextual inquiry notes were merged with the video coding based on the respective event time for (quantitative) descriptive analysis. Any time discrepancies were resolved using the video data.

5 FINDINGS: STUDY 1

We present our findings based on the research questions and themes that emerged from the data analysis.

5.1 What are the typical on-the-move visual behaviors?

5.1.1 Visual behavior patterns. To analyze the visual behavior patterns in the contexts of mobile human-computer interactions, we describe the visual behaviors at the task level based on their characteristics, i.e., purpose, duration, and intensity. We identified three distinct patterns from the mixed analysis: *glance*, *inspect*, and *drift*, as illustrated in **Fig 2**. For each pattern, we distinguish between two main reasons for engagement: decision-making and action-taking.
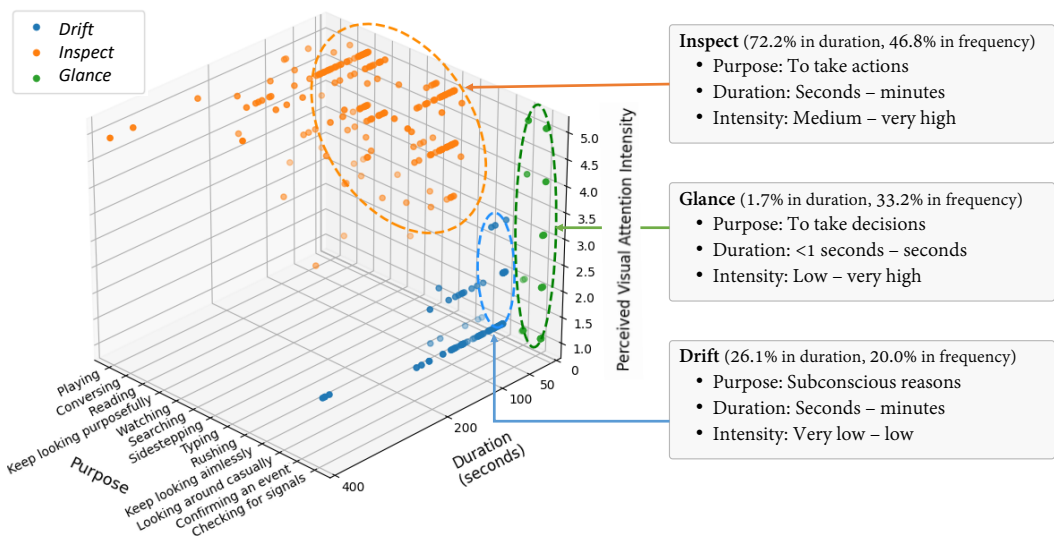

Fig. 2. 3D scatter plot of the three visual behavior patterns: *glance*, *inspect*, and *drift*, clustered by purpose, duration of the behavior, and perceived visual intensity. For example, checking the arrival of a bus, confirming bus numbers, or checking for notification type are *glance*s. Playing a mobile game, looking for someone, or watching the outside scenery intentionally are *inspect*s. Visually scanning the surroundings without intention or staring at something without purpose are examples of *drift*s. We further classified the purpose for engaging in each behavior pattern into different subcategories. Duration (seconds) is displayed on a log scale. Perceived visual attention intensity is represented as a 5-point Likert scale (1 = Very Low, 5 = Very High). See Appendix A.3 for duration distributions. NOTE: The visual patterns shown here are non-exhaustive as they only cover instances of inquiry.

Glance: Commuters utilize *glance*s for the purpose of decision-making, such as to check if action is required. *Glance*s are less than a second to several seconds in duration since decision-making does not require much time. Furthermore, the intensity of the visual attention required for *glance*s varies; relatively low-intensity attention focus is required when users can return to their primary task at hand without taking any post-action, such as when a commuter looks up to check the bus number, only to resume his previous task when he finds that it is not the correct bus. On the other hand, greater intensity in attention focus is required when a post-action is necessary, such as when this same commuter *glance*s up to check the bus number and prepares to board it when he finds that it is indeed the right bus that has arrived. During a *glance*, the commuter's gaze and head

orientation change quickly within a few seconds and in a highly alert state[1]. Checking the arrival of a bus or sudden notifications on the phone are thus examples of *glance*s.

Inspect: *Inspect* occurs when people need to keep their gaze on an object related to their task at hand or engage in action. It can last for several seconds to minutes, depending on task duration. Texting, reading, and chatting are ubiquitous activities that fall under the category of *inspect*ion. *Inspect*ion requires continuous monitoring and thus generally requires higher visual attention intensity. During *inspect*s, the commuter's gaze and head orientation change slowly or are kept static with high alertness. Intentionally looking for someone, reading notices, and watching the outside scenery are instances of *inspect*s.

Drift: *Drift* is a natural behavior to reduce physical fatigue, refresh the commuter's mind. Subconscious activities such as pondering, recalling past events, or daydreaming fall under this category. People usually *drift* when they are not visually engaged with any task, such as when a commuter unintentionally looks around or stare at something to rest his mind after having read an article for some time. *Drift* can last from seconds to minutes, depending on the duration of the interruption, or until *drift* is switched to *glance* or *inspect*. *Drift* requires low levels of visual attention since it is not a full engagement of a task. During *drift*, the commuter's gaze and head orientation change slowly or remain static with low levels of alertness.

5.1.2 Visual behavior transitions. As illustrated in Fig 3, participants fluidly switch between patterns while performing different tasks on the move. They may occasionally *glance* at their surroundings while *drift*ing and *inspect*ing (e.g., Fig 3 (1) 'check the path' (*glance*) or 'check for a bus' (*glance*) during 'texting/reading on the phone' (*inspect*)). These *glance*s can interrupt the ongoing tasks and divert their visual attention to new tasks/actions, such as when a commuter stops texting on his phone (*inspect*) when he notices (*glance*) that someone is walking towards him in order to avoid collision (*inspect*).
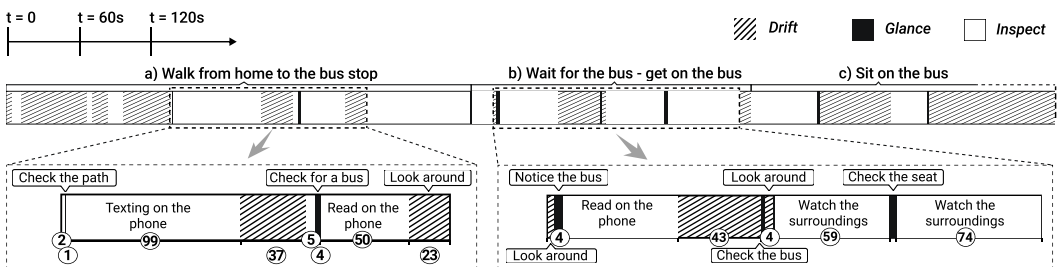

Fig. 3. A participant's (P1) visual behaviors during commuting from home to school illustrate how visual behaviors fluidly transition. The numbers inside circles represent the duration for each pattern in seconds. a) Walking from home to the bus stop (time taken: 8min 49s), b) Waiting to get on the bus (time taken: 4min 50s), and c) Commuting on the bus. Refer to Appendix A.1 for further details.

The duration and frequency of visual patterns and transitions were dependent on contextual factors [45] such as location, time, crowdedness, and personal habits like mobile phone usage. Two participants who showed the highest (P1, 48.2%) and lowest (P6, 14.8%) overall duration for *drift*s exemplify this dependency in Fig 4.

Furthermore, *glance*s and *inspect*s were affected by mobile HCI tasks. 23.8% in duration (17.5% in frequency) of *glance*s and 71.8% of duration (65.8% in frequency) of *inspect*s were attributed to mobile HCI tasks that occurred during commuting (e.g., *glance*s: checking smartwatch, *inspect*s:

---

[1]the state of being attentive and prepared to react [3]

reading/texting/watching on the phone or tapping travel card). The duration of visual patterns lasted between 1 second (*glance*: when checking the phone/smartwatch for new notifications/time while walking) and 695 seconds (*inspect*: when playing a mobile game while standing on the metro), depending on the task and context. As expected, greater engagement with mobile HCI tasks reduced the overall duration of *drift*s and increased the overall duration of *inspect*s.
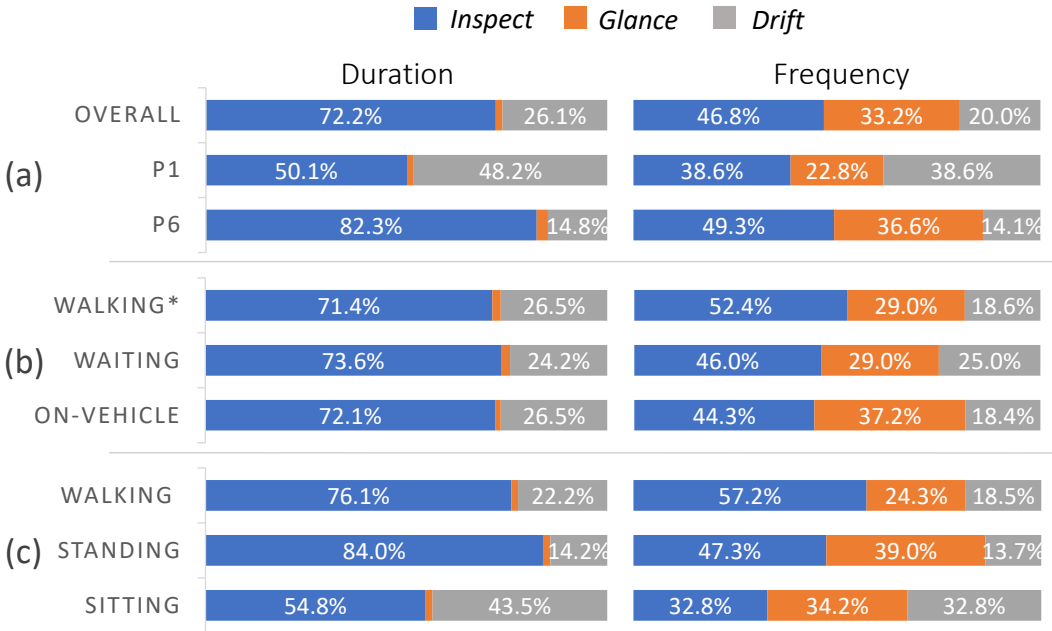

Fig. 4. Overall variation in visual behaviors across duration and frequency, based on (a) individual behaviors, (b) commuting stage (walking for commuting, waiting for commuting, and on-vehicle commuting), and (c) posture. * Includes scenarios such as traveling on an escalator/lift which does not involve walking and does not belong to other commuting stages. Refer to Appendix A.4 for further details.

5.2 How can visual behaviors support information processing tasks such as microlearning?

5.2.1 Visual behaviors and receptivity to microlearning. Our findings suggest a significant connection between visual behavior patterns and receptivity towards microlearning. Comparing the receptivity ratings between patterns, we found a significant effect of *pattern* (Kruskal-Wallis test, $H(2) = 198.4$, p < 0.001) and Dunn's post-hoc comparison showed significant difference (p_{bonf} <0.001) between each pair from high to low receptivity: *drift > inspect > glance*.

Drift. Generally speaking, *drift*s presented the most opportunistic moments for microlearning since they had the highest receptivity (willingness $M = 3.8$, $SD = 0.33$, $n = 132$, n represents the number of patterns related to contextual inquiries). We differentiate between two types of *drift*: *passive drift* and *active drift*. *Passive drift* indicates when the participant has nothing to do, such as boredom, e.g., *"[sitting on an empty bus,] I am doing nothing now [observed drift], so I think it's good to learn words (P7)"*. *Active drift* indicates relaxation and ponderation, e.g., *"I was thinking something important at that time [during observed drift]. I don't think it's a good time [for microlearning]. (P4)"*. Interestingly, the observed ratio between *passive drift* and *active drift* was 126:6 = 95%, indicating that users were willing to engage in more meaningful activities if given a choice for the majority of

*drift*s. Cognitive load monitoring [37] may help us distinguish between *passive drift* and *active drift* since the latter involves non-visual cognitive tasks such as pondering.

Our analysis also shows a positive, medium correlation between receptivity to engage in mobile information acquisition task and duration of *drift*s ($Spearman's\ \rho$ = 0.356, p = 0.052), i.e., user receptivity increases when available duration increases. Participants shared in their interview that microlearning was better supported by *drift*s that are longer than 3 minutes ($MIN = 3$, $MAX = 6$, $M = 4.3$, $SD = 1.18$): *"I think I want to learn [3-6 words] when the duration is about 5 minutes [on the metro] (P5)"*. As expected, receptivity towards microlearning was lower during *drift*s with shorter durations. This depends on whether the value of switching to microlearning outweighs the switching cost, i.e., the time to switch from one task to another, including both physical and mental preparation time [4, 48]. As described by some participants, *"[standing on the metro] I will get down in 2 minutes, it's not worthy to open the app (P6)"*, *"[waiting for the bus] I need to get on the bus in a short time, maybe after I start, I can learn only a few words before I have to pay attention to the bus. So why not start learning after sitting down on the bus? (P8)"*.

In addition, the purpose and method of microlearning also mattered; participants highlighted that their preferred learning time depends on their personal learning technique, such as word repetition, making sentences or puns, and associating with objects/sounds.

While the examples provided above are mainly based on microlearning, they can be extended to other information acquisition tasks. In general, *drift* moments are best utilized for presenting information that is unrelated to their current activities. In addition, longer *drift* moments are considered better for information acquisition than shorter ones.

Inspect. *Inspect* is generally unsuitable for microlearning (willingness $M = 2.4$, $SD = 1.10$, $n = 237$) and other information acquisition tasks. However, receptivity levels differed based on two different types of engagement with the ongoing tasks; *essential engagement* and *inessential engagement*. As participant P3 summarized during an *essential engagement*, *"[chatting on social media] I don't want to learn at all. I want to chat with my friends and see their updates."* Similarly, frequent *inspect*s, like flipping through social media, texting friends, and intentionally looking at the road in crowded places, led to low receptivity towards other information acquisition tasks. However, during *inspect*s related to *inessential engagements* which were carried out to *"kill"* time (e.g., playing games, surfing the internet), or for tasks with low priority, participants showed greater receptivity to other information acquisition tasks: *"Oh, maybe it [observed inspect] is a good time to learn. I am just reading some gossip. You know when you are reading you won't think about language learning... I prefer doing more valuable things like vocabulary learning to refresh my mind. (P1)"* The observed ratio between *essential* and *inessential engagement* was 224:13 = 95%, indicating that users were mostly unwilling to engage in additional information acquisition tasks during *inspect* instances, as it disrupts their engagement with ongoing tasks.

Glance. As expected, *glance*s due to their short duration and specific use are unsuitable for additional information acquisition tasks (willingness $M = 1.2$, $SD = 0.38$, $n = 58$). However, we noticed that *glance*s interact with other visual behaviors to determine information acquisition suitability. During the commute, *glance*s were needed in many navigational scenarios, and each *glance* required the user to navigate their attention away from their current task. When the frequency of *glance*s increased, users frequently experienced interruptions of their current task, which naturally leads to less efficient information acquisition. For example, when waiting for the bus to come, P2 was less willing to engage in an information acquisition task due to a lack of proper estimation for their bus arrival time, *"I think the bus is coming soon, but I am not sure. I have to check from time to time. [it] is too distracting"*. According to our observations, the frequency of *glance*s increased when there is unfamiliarity with place/road, uncertainty with the surroundings, vehicle motion (e,g., sudden jerks), and *external interruptions* (i.e., dynamic interruptions from the

environment). For example, P3 stated, *"if I see people are coming, I need to look at them more often to avoid bumping. Then I don't want to learn."*

External interruptions. Overall, there were two types of dynamic signals/cues from the surrounding environment that led to *glance*s and potentially interrupted information acquisition: *background signal*s and *trigger signal*s. *Background signal*s lead to the perceiving or cognitive processing of the cue without resulting in any associated action, while *trigger signal*s lead to user actions (see Table 2). For instance, the scene outside the bus is a *background signal*, while a change in the traffic light is a *trigger signal* when the participant is waiting to cross the road, as it calls for action.

The frequency of perceived *trigger signal*s and *background signal*s was highly dependent on the time of the day, route, posture, and medium of commute. For example, the *trigger signal* (flocking in/out of a crowd) frequency while commuting by bus or metro during weekdays happened more frequently in the morning ($\approx$ 1 signal per 5-15 min) than at noon ($\approx$ 1 signal per 30-60 min).

Table 2. *Signal*s from the surroundings

Background signals	Trigger signals
1 or 2 passengers getting on/off the vehicle	The flocking in of a crowd of passengers
The movement of a few people on the road	The movement of a crowd of people on the road
The fluctuation of the bus	The color change of the traffic light
The slight changes of people's postures who are around	The arrival of the metro or the bus
Announcements not related to commuting routes	Warning sounds on the metro-station

5.2.2 Factors affecting the utilization of visual behaviors for information acquisition tasks. In summary, we identified four *user-related* factors to consider when adapting opportunistic visual behaviors in dynamic environments for information acquisition, such as microlearning.

(1) *Balancing between situational awareness and information acquisition.* When *signal* frequency or intensity increases, users would be less likely to engage in additional information acquisition tasks. Instead, they would prefer to focus on the surroundings only. *"I have to be more concentrated on the surroundings now. If I continue looking at my phone, I am afraid of bumping into others"*, said P3 during a crowd influx. Since users have to act on *trigger signal*s, ongoing information tasks should be paused when such signals are detected to allow users to focus on *external interruptions*. Similarly, when any *signal* is present, delaying the sending of new information tasks until users can focus on information tasks reduces the need to switch between *signal*s and tasks, hence also reducing perceived information overload.

(2) *Managing switching costs between navigational tasks and information acquisition.* Participants expressed that they need a minimum available time before considering task switching. The amount of time required depends on the type of information acquisition task. In the case of microlearning vocabularies, participants expressed that they need at least 3 to 6 minutes to learn a few words. Hence, the perceived switching cost impedes the utilization of short *drift*s for information acquisition, given that actual acquisition can occur in a lesser duration (e.g., less than 20s is required to associate 1-word pair [27]). The higher perceived switching costs also depend on other factors, such as the time taken to start the mobile app after taking it out of the pocket and unlocking it. Thus, in reducing switching costs, we can enable the utilization of shorter durations for productive information tasks.

(3) *Balancing between the ongoing cognitive processes and information acquisition.* Even though *drift*s are the most opportunistic visual behavior for information acquisition, when users are

engaged in other cognitive processes (i.e., *active drift*s), they become unwilling to engage in microlearning, *"[sitting on the bus, during observed drift] I had a class. I need to relax, I don't want to learn now. (P7)"* The presentation of information tasks during *active drift*s should be avoided, so as to minimize any potential annoyance to the user.

(4) *Difficulties in self-identifying opportunistic moments for information acquisition.* Since their own behavior patterns were not observable to participants, they did not engage with *"valuable"* activities even if they were more receptive. This indicates a need for developing methods of identifying and providing feedback about potential opportunities (e.g., *passive drift*s).

5.2.3   Device-task interplay. From the inquiries, we identified two mobile phone limitations for microlearning.

(1) The reduction of peripheral vision field when visually focused on phone screens. This visual field constriction made it difficult for participants (*7 out of 8*) to maintain situational awareness while learning with their phones during commuting. As P3 mentioned, *"Sometimes I am focused on finishing a [mobile learning] session, and I can not see when somebody needs me to move out of their way."*

(2) Ergonomics. *4 participants* complained about the heads-down posture and associated fatigue with mobile phones, *"Focusing on the phone is really tiring. I need to nod down to look at the screen, and it hurts my neck. (P4)"*

Additionally, three participants did not prefer using mobile phones for learning vocabulary while commuting due to their propensity to motion sickness, *"I can not focus [a] long time on [the] phone especially on [the] bus as I feel carsick. (P6)"*

Existing literature supports the device limitations that we identified. Visual field constriction [62] is a key factor in the 'smart-phone zombies' phenomena [5], which describes mobile phone users who are obsessively engaged with their phones and compromise on situational awareness [9, 58]. Moreover, detrimental long-term effects of the head-down posture include health problems such as the 'Text Neck' [36, 71], a form of chronic musculoskeletal disorder.

5.3   Discussion

Different visual behavior patterns cater to the specific needs of different tasks. Patterns transition when users switch tasks, self-interrupt (e.g., choosing to relax after interactions), or react to *external interruptions* (e.g., *signal*s).

User receptivity to information acquisition depends on the behavior pattern and perceived attention needs. The Resource Completion Framework (RCF) [68] describes the cognitive resource allocation during mobile HCI tasks, which we can use to explain user receptivity levels during each visual behavior pattern. According to the RCF, when an information task dominates one's attention, the working memory processes retained information, increasing the cognitive load. The cognitive load affects the perceived demand of tasks, the effort required to start or continue a task, and users' receptivity towards the information task. We found that *Drift* requires the least visual attention; thus, users have a greater capacity to divert their attention to other information acquisition tasks. In contrast, *inspect*s were poorly received since it is attention-demanding. Therefore, we can predict opportune moments for information acquisition tasks and forecast user receptivity by observing their visual behavior patterns.

It is essential to consider the threshold or lower bound time limit for effective information acquisition. The effective association of a word pair takes at least several seconds[16], suggesting that visual behaviors shorter than $\approx 8s$ (minimum time required for users to associate a word pair [27]) are unsuitable for microlearning. Some *glance*s and very short *drift*s are too brief for most information acquisition tasks. Visual behaviors of shorter durations (e.g., *glance*s) impact

users' receptivity, as the frequency of such behaviors also affects the frequency of transitions. This is because *glance*s involve checking and remembering problem-states, i.e., working information relevant to the ongoing task [12]. Hence, when the prediction of the problem-state is inaccurate, more frequent *glance*s are required, requiring greater effort from the user to store and restore multiple problem-states. This cognitive demand may lower user receptivity.

According to our study, *drift*s of longer durations are common and useful. In the case of microlearning, as much as 65.4% of observed *drift*s are longer than 8s. These account for 92.3% of the total *drift*ing time. Even with a higher duration threshold of 33s (median duration), there will still be 50.9% of the *drift*ing time (83.7% of the total duration). This provides considerable opportunity for information acquisition tasks since *drift*s account for more than one-fifth of commuting time (21.8% of total commute time). Furthermore, more than 60.8% of *inspect*s durations are carried out for mobile-HCI tasks (each duration lasting longer than 33s), where some of them are carried out for nonproductive tasks. While our observations indicate many opportunities for information acquisition on the go, it is vital to avoid overusing these opportunities to the point of information overload and mental fatigue.

6 STUDY 2: OPPORTUNITIES WITH VISUAL BEHAVIORS FOR MICROLEARNING WITH MOBILE PHONES AND OHMDS

Study 1 established the visual behaviors that commuters engage in and the desirability of utilizing them for information acquisition tasks. However, since we did not test the actual usage in any mobile platform, the results were not validated ecologically.

We proceed to investigate visual behaviors and information acquisition opportunities on different platforms, particularly by comparing the de-facto mobile phone platform with the emergent OHMD platform. Both platforms are designed for mobile usage but have distinct characteristics that can provide different information acquisition opportunities.

Our research questions for *Study 2* includes:

RQ1: How are visual behaviors for microlearning utilized differently on the mobile phone and OHMD platforms? How do the platforms affect user receptivity to additional information acquisition tasks (i.e., microlearning)?

RQ1.1: What are the limitations of each platform in utilizing visual behaviors for information acquisition (e.g., microlearning)?

To answer these research questions, we probed [42] microlearning on both mobile phones and OHMDs separately, focusing on *drift*s that users were most receptive to. Probing allowed us to focus on identifying the relationship between device platforms and user-related factors identified in *study 1* (sec 5.2.2) without being subject to the technical constraints of an actual implementation. Therefore, we used the *push* strategy [45] to **remind** users of potential microlearning opportunities and verify whether the platforms (mobile phone and OHMD) enable users to utilize them and identify the associated tradeoffs.

6.1 Probe study design

To make the information acquisition task more realistic, we created a vocabulary list for microlearning, similar to what is presented in current vocabulary learning apps such as Duolingo[2], which contains both visual (spelling) and auditory (pronunciation) cues. To ensure that no participants have prior experience with the selected vocabulary list, we used Vimmi corpus [61], an artificial

---

[2]https://www.duolingo.com/

corpus, as the second language. We created 90 Vimmi-English word pairs (e.g., "toze" in Vimmi refers to "flower" in English) as part of the microlearning material.

We designed two Android mobile apps, one for participant microlearning (Fig 5 (b1), (b2)) and the other for the experimenter to trigger microlearning (Fig 5 (a)) in the participant app. Whenever the experimenter triggered a microlearning session, Vimmi-English word pairs were automatically displayed and audio pronunciation sounded on the participant's device (see Fig 5 (b1), (b2)).

Each microlearning session included 6-word pairs, a design choice based on previous study results [27]. To determine the display duration of each word pair, we conducted a pilot on three volunteers, where they had to learn vocabulary on both platforms on the move. All participants preferred the duration of 10 seconds.

While learning involves different stages (i.e., acquisition, retention, and transference of knowledge [7]), we focused on the **acquisition stage** as it is the first step and fundamental to the other stages of retention and transfer. Therefore, we did not adopt any vocabulary learning techniques (e.g., spaced repetition).
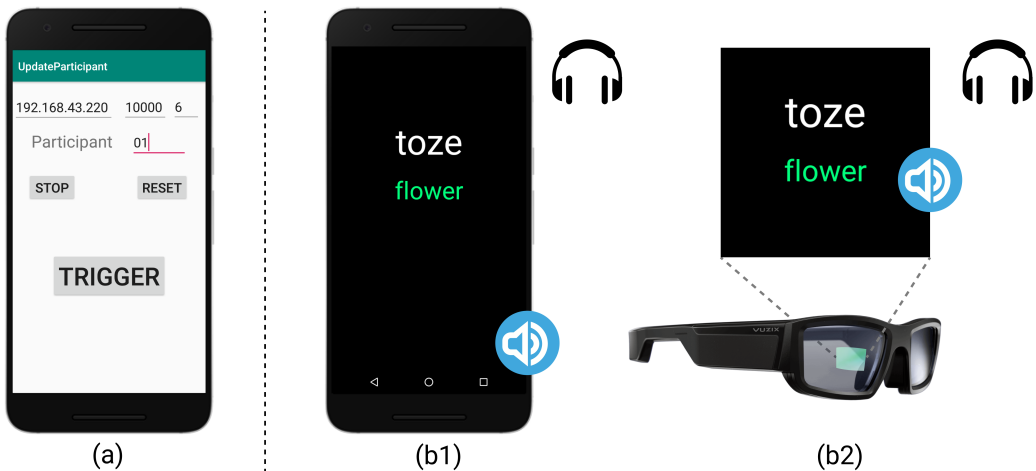

Fig. 5. Apparatus used in the probe-based study design: (a) Mobile application interface to trigger the microlearning session. The participant's device (mobile phone or OHMD) IP address, microlearning session parameters (e.g., 10000ms time gap between words, 6 words per session), and participant id were configured during the training. The microlearning session was triggered by clicking the 'TRIGGER' button. (b1) Mobile microlearning application interface (b2) OHMD (Vuzix Blade) microlearning application interface. The trigger displayed the Vimmi-English word pairs and played the audio pronunciation in the device. NOTE: In OHMDs, the black color background represents the transparent background.

In our informal pilot study, participants detected words shown on their OHMD without the need for any additional notification though this was not the case for mobile phones. Therefore, mobile phone words were triggered with vibration and a 1s audio beep to notify participants (who might have had their attention off-screen) that a word had been triggered. For both platforms, words were shown on screen for a fixed duration before disappearing. Thus, users who chose not to pay attention missed out on the displayed content.

6.2 Method

6.2.1 Apparatus. For the *phone probe*, participants installed the microlearning mobile app (Fig 5 (b1)) on their own phones. For the *OHMD probe*, participants wore a pair of Vuzix Blade[3] smart glasses (480x480 px display, centered on the right glass as shown in Fig 5 (b2)), installed with a customized version of the mobile app (see sec 6.1). In both probes, participants wore earphones to listen to the audio. The experimenter used a Google Pixel 4 phone to trigger the microlearning sessions.

6.2.2 Participants. A total of 20 volunteers were recruited for the two probes (12 females, age $M = 23.6$, $SD = 3.1$, $MIN = 19$, $MAX = 30$). 16 of them were students, 3 were IT professionals, and 1 was a business professional. All participants had received formal education in English and had experience using mobile learning apps during commuting. However, none of them had prior experience using OHMDs. In addition, all participants were regular commuters, spending an average of 131 minutes ($MIN = 80$, $MAX = 300$) per day commuting.

We recruited 16 participants with a between-subject design to obtain diverse feedback and minimize possible interference between the two conditions. Eight (P1-P8) participated in the *phone probe*, and eight (G1-G8) participated in the *OHMD probe*. We balanced participants between the two probes based on their commuting time and medium. While between-subject design eliminates the possible interference between conditions, it is less sensitive to detect subtle differences between conditions due to subjective differences. To compensate for this, we also tested 4 more participants (**GP1-GP4**, 1 IT-professional, 3-students) with a within-subject design in which all 4 participants underwent both probes.

6.3 Procedure

Fig 6 illustrates the process of the probe study (for details, refer to Appendix B.1). We did not test the number of words participants remembered (i.e., information *retention*) as the objective was to identify how the two platforms support information **acquisition** of new concepts and **receptivity** of platforms.
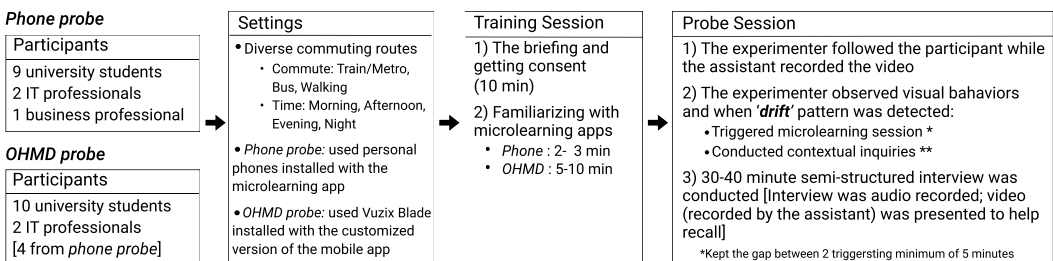
Phone probe		Settings	Training Session	Probe Session
Participants		• Diverse commuting routes	1) The briefing and getting consent (10 min)	1) The experimenter followed the participant while the assistant recorded the video
9 university students		• Commute: Train/Metro, Bus, Walking		2) The experimenter observed visual bahaviors and when '**drift**' pattern was detected:
2 IT professionals		• Time: Morning, Afternoon, Evening, Night	2) Familiarizing with microlearning apps	• Triggered microlearning session *
1 business professional		• *Phone probe:* used personal phones installed with the microlearning app	• *Phone* : 2- 3 min	• Conducted contextual inquiries **
OHMD probe			• *OHMD* : 5-10 min	3) 30-40 minute semi-structured interview was conducted [Interview was audio recorded; video (recorded by the assistant) was presented to help recall]
Participants		• *OHMD probe:* used Vuzix Blade installed with the customized version of the mobile app		
10 university students				
2 IT professionals				*Kept the gap between 2 triggersting minimum of 5 minutes
[4 from *phone probe*]				

Fig. 6. Study 2 procedure. See Appendix B.1 for further details.

6.3.1 Measures. Similar to *study 1*, the primary measure was participants' feedback. A 5-point Likert scale (1 = Very Low, 5 = Very High) was used to gather participants' receptivity and concentration on microlearning. The contextual inquiry topics and interview topics can be found in Appendix B.2 and Appendix B.3.

---

[3]https://www.vuzix.com/products

6.4 Data Analysis

We conducted a qualitative analysis with 247 sessions of contextual inquiry notes, observation notes, and 24 interview transcripts. Two researchers (co-authors, one experimenter) used the *QDA Miner* software package and analyzed notes using the open coding technique [24, 25]. An initial set of common codes were generated by analyzing 4 participants' data (2 per platform). This initial set of codes were further refined after watching video recordings and further discussion. The refined coding scheme was then iteratively tested and revised to reach 94% agreement among the coders before coding the rest of the participants.

7 FINDINGS: STUDY 2

All participants engaged with microlearning during the observed 247 *drift*s, except for 6 technical fault cases where the microlearning app failed to start upon triggering. We compared the overall probe duration and session count statistics to verify if participants had similar experiences with the two platforms. On average, *phone probe* participants spent 70 minutes (SD = 25), while *OHMD probe* participants spent 73 minutes (SD = 32) on commute during the study, excluding briefing, training, and post-interview time. On average, *phone probe* participants received 9.2 sessions/hour (SD = 1.6) while *OHMD probe* participants received 9.4 sessions/hour (SD = 1.5). Additionally, we observed that between-subjects participants had similar feedback on their experiences as within-subject participants.

We organized the results based on the research questions and themes that emerged from our data analysis.

7.1 RQ1: How are visual behaviors for information acquisition utilized differently on the mobile phone and OHMD platforms?

Overall, we found that the receptivity to information acquisition (microlearning) primarily depended on *external interruptions* (i.e., *signals*, sec 5.2.1). However, the platform also had a strong influence on how these opportunities were utilized.

7.1.1 External interruptions, microlearning, and situational awareness. In general, participants' receptivity to engage in microlearning, as well as their concentration levels reduced as the *signal* (sec 5.2.1) intensity increased (Fig 7, compare *no signals* or *background signal*s with *trigger signals*). However, the two platforms differed in this aspect. As shown in Fig 7a, the *overall* receptivity did not vary much across platforms. This is because users' receptivity depended on their information needs and motivations.
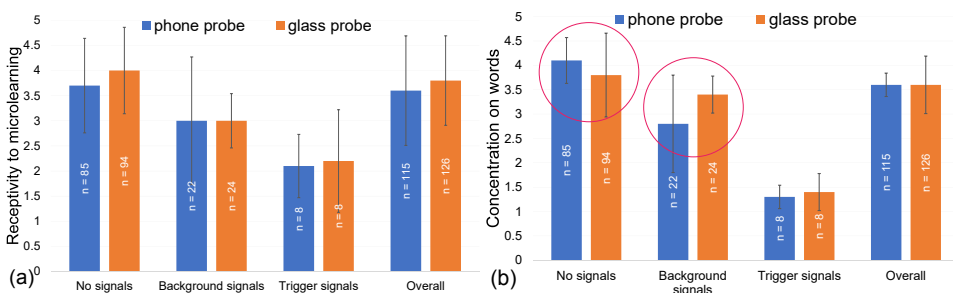

Fig. 7. *Signal*s' effect on (*a*) receptivity and (*b*) concentration on information acquisition (microlearning). The Y-axes of both graphs represent 5-point Likert scales, and error bars represent the standard deviation. NOTE: *Signal*s occurred after the information acquisition task was triggered during *drift*s.

As shown in Fig 7b, when there were few *signal*s, mobile phones enabled better concentration on information acquisition due to their static (opaque) backgrounds and high-resolution screens. On the other hand, when there were more *signal*s, information acquisition on mobile phones was more frequently interrupted, significantly affecting participants' concentration. In contrast, OHMDs supported better concentration as users tended to be more situationally aware of *signal*s in most commuting scenarios (refer to Fig 8 (a2), (b2)).

All OHMD users checked their surroundings through the see-through displays and used their peripheral vision to monitor *signal*s and decide on the type of signal: *background signal* or *trigger signal*. Their microlearning was only interrupted during *trigger signal*s: *"[waiting for the bus, background signal] I don't have to change my gaze to check the bus deliberately. I know clearly whether the bus is coming even when my eyes [are] still glued to the words. (G1)"*, and *"[walking in the metro station, suddenly a child runs in front, i.e., trigger signal] I managed not to bump into the kid even though I was looking at words, but I lost my focus (G3)"*. The see-through nature of OHMDs helped users to avoid situational hazards, allowing them to pay attention to information acquisition and dynamic *signal*s.

Fig. 8. Situational awareness on mobile phones and OHMDs: (a1) The mobile phone user needs to lift the phone to pay peripheral attention to the crowd; (a2) The OHMD user looks ahead and focuses on the words while paying peripheral attention to the crowd; (b1) While walking, the person needs to look up to check the surroundings and look down to focus on the phone; (b2) While walking, the OHMD user can check the road while paying minor attention to words

In contrast to OHMDs, mobile phones were not designed for users to effectively acquire information on screen while monitoring the environment.

There were also disadvantages to using peripheral vision for OHMD content and focal vision for the surroundings. When *3 participants* had to pay more attention to the *signal*s, words that appeared in the peripheral vision were distracting: *"[waiting for the bus] Actually, I was quite okay with learning when I was checking the bus. But now [walking towards the bus], I find the words really distracting. I can't help looking at the words from time to time, but I become quite worried about bumping into others when I do it. (G2)"* This indicates that information acquisition tasks on OHMD are best paused when *signal*s require the users' focused attention.

7.1.2 Learning strategies and behavior changes. There were differences between the two platforms regarding how users engaged in information acquisition and how it affected their visual behavior.

OHMD users could adopt a "covert learning" strategy (*8 participants*) when they had to pay more attention to the surrounding information (e.g., *signal*s), which enabled a more continuous learning experience compared to mobile phones. A "covert learning" strategy refers to the use of peripheral vision to check on learning content while focal vision is occupied by *signal*s. For example, according to G2, *"When I was checking the number of the bus, I can still see the words in the range of vision.".* G1, on the other hand, described it as a "half-look": *"even when I am looking*

elsewhere, I can 'half-look' at the words at the same time." Moreover, this "covert learning" strategy can minimize interruption to microlearning with frequent *glance*s in between *drift*s, and is suitable for use during commute when attention is often fragmented. Given that *external interruptions* are unavoidable on the move, "covert learning" can reduce the cost of context switching between information acquisition and attending *signal*s and improve the efficiency of information acquisition. We suggest future investigations for validating this hypothesis.

Mobile phone users could hardly leverage "covert learning" since keeping their phone screen within the same range of vision as when their gaze is directed towards their immediate environment was challenging. Instead, all participants used these moments away from the screen to digest and memorize words: *"When I look down at the phone, I am intaking the word, just to leave a rough impression. When I look up and check the road, I can digest the word and try to memorize it. After checking the road, I refocus on the word for [a] deeper impression. (GP3)"* For this case, we recommend further studies to identify the most effective strategy for information *retention* between two platforms on the move.

Information acquisition tasks clearly influenced visual behaviors. In the case of microlearning, participants switched from *drift* to *glance* (check the word) or *inspect* (recognizing the word) and then back to *drift* (trying to memorize words). In the *phone probe*, all participants used to look *down (word) and up (environment)*; while in the *OHMD probe*, all participants used to *focus (word) and defocus (word)*, to check and remember words: *"I focus on the words [be]cause I saw it, but then when it [word] changes then I'll refocus on the change. (G8)"* While mobile phone users explicitly used both head movements and gaze to focus and defocus on words (e.g., focus on environment), OHMD users only used gaze to do so.

7.1.3 Switching cost. Overall, OHMDs can lower switching costs as less physical effort and gaze time are required of users. OHMDs require shorter transition times *"fraction of second"* ($\approx$ 150 ms according to [75]) than phones (we observed a 2-6 seconds transition time). The ease in transition considerably lowers the psychological barrier towards information acquisition tasks, as G4 commented: *"the smart glasses quickly bring me into a learning state without any delay. But usually, when I receive notifications from the phone, I need more time to switch into a good learning condition. Maybe I will wait for a while before picking it up. Or maybe I don't bother."*

Due to low switching costs, OHMDs could utilize shorter chunks of time for information acquisition and thus offer more opportunities for information acquisition tasks than mobile phones. For example, in certain places such as the escalator where participants had 40-60 second *drift*s, most participants (*7 out of 9*) showed low receptivity to the *phone probe*. P4 highlighted this, *"the time is very short, so after I picked up the phone and started learning, maybe I have to put it down and prepare for getting off before I can remember a single word. I'd rather do nothing but wait."* On the contrary, participants were more receptive during short *drift*s in the *OHMD probe* (*7 out of 8*): *"I think even 10 seconds is possible to be utilized for smart glasses [to microlearn]. I just need to quickly glance at the words and then quickly defocus to prepare for the next task. (GP2)"*

However, for users who are not ready to accept information, the low switching costs of OHMDs could cause unwanted interruptions to ongoing mental processes. For example, two participants complained that their thought processes were interrupted by OHMD content that appeared, *"I was thinking something important just now. But the words suddenly appeared, and I forgot what I was thinking (GP4)".* On the other hand, one participant from the *phone probe* missed two sessions as he did not notice the vibration or the audio beep while he was pondering. Therefore, OHMDs are more likely to result in accidental triggers that potentially annoy users.

7.1.4 Ergonomics, postures, and motion effects. OHMDs could mitigate the ergonomics issues identified in *study 1* (sec 5.2.3) by supporting the natural heads-up posture. It could also reduce the

perceived effects of vehicle motion, reduce disturbances to information acquisition, and improve receptivity. Three participants in the *phone probe* expressed that looking up and down was *"tiring"* or *"annoying"* when they were microlearning. In contrast, five participants in the *OHMD probe* expressed that OHMDs were *"much more comfortable"* than mobile phones (based on their past experience) and thus, increased their willingness to learn.

Moreover, we observed the influence of posture in the *phone probe*: *5 participants* changed their posture from standing to sitting with no *signal*s, and as a result became more receptive (willingness $M = 3.5$, $SD = 1.02$, $n = 42$ **to** willingness $M = 4.4$, $SD = 0.41$, $n = 32$). P1 mentioned in the *phone probe*, *"when I am sitting down, I feel more comfortable. I don't move my head so often even [...] my willingness is higher [compared to standing]"*. Yet, three participants mentioned that they preferred standing over sitting due to the potential *signal*s. For example, P5 mentioned that *"I wouldn't have to keep looking up to check for people who need [my seat]"*. In contrast, OHMD users were more "relaxed" as the OHMD supported personal preferences for posture as they engaged with microlearning. This made them more receptive to learning on OHMDs than on mobile phones during those particular scenarios.

The see-through and view-stabilized nature of OHMD content reduced the adverse effects of vehicle motion (e.g., jerking) and sudden speed changes (*5 out of 7*). Mobile phone participants required more effort to keep their eyes on the screen in moving vehicles, which sometimes led to motion sickness. GP3 highlighted that *"[with the mobile phone, sitting on the bus] If it is bumping up and down, I need to look at the screen for a longer time to intake the information. And I will look up less frequently because it makes me sick.", "[with OHMDs] It's easy to focus on words [...] even with shaking"*. As suggested, OHMDs facilitate better concentration in more dynamic contexts (see Fig 7b) and lower the barrier to information acquisition task engagement. In the next section, we expand on the limitations of each platform's use.

7.2 RQ1.1: What are the limitations of each platform in utilizing visual behaviors for information acquisition?

7.2.1 Limitations of mobile phones. All limitations identified in sec 5.2.3 were also present in the *phone probe* when participants physically experienced experimenter-triggered microlearning sessions during *drift*s.

7.2.2 Limitations of OHMDs. **Social limitations**. Social context plays an essential role in participants' ability to acquire information. For instance, participants disliked having digital content appear on their OHMDs when they were also looking at the faces of others. In our study, all participants except one reported that they would first check that they were not interacting (directly/indirectly) with others before beginning to read their OHMD content.

Focusing on OHMD content could lower participants' awareness of their social environment. Two participants encountered the issue of unintentionally staring at others when they were, in fact, focusing on displayed content: *"I suddenly realized that I was looking at someone's face when I looked away from the words, so I quickly changed the direction I was staring at. I am afraid that when I focus on the words, I will again unconsciously look at someone, so I became more careful. (G6)"* In crowded environments, some would look at the ground or the ceiling to avoid such unintentional gazes at others; however, this *"unnatural"* posture made learning *"uneasy"*, an issue which OHMDs usually helps overcome. According to Akechi et al. [2], some cultural contexts consider gazes to be inappropriate. In cultures where this poses a sensitive social issue, OHMD users may be more inclined to ensure that they maintain a base level of situational awareness to avoid landing unintended gazes upon others.

Technical Limitations. We identified three areas of technical constraints for OHMDs that limit information acquisition on the move. The first constraint relates to the high external brightness and contrast, which is common in most optical see-through displays [8, 49, 59]. Eight participants mentioned that when the environment was bright, or the surrounding colors matched the text colors on display, they had to look for *"contrasting"* surfaces to see OHMD words clearly. Three participants complained that reflections in the outdoor environments distracted them. The second limitation arises from the hardware properties of OHMDs. Four participants mentioned that OHMDs were heavy, inconvenient for use, and difficult to customize. The third limitation relates to the issue of eye strain [38, 41]. Only one participant encountered this due to unfamiliarity with the OHMD prototype and its monocular nature: *"[I] feel the right eye is used more, even when the screen is centered (G2)"*.

Nevertheless, we anticipate that future advancements in OHMD technology will seek to resolve these technical limitations. For example, better projection technology will resolve issues of visual contrast and reflections [1]. We saw significant improvements in some of the recently released OHMDs, such as the Nreal[4] smart glasses, which is comfortable to wear and delivers a better viewing experience.

7.3 Summary of the findings

Table 3 summarizes the key findings of this study.

8 DISCUSSION AND DESIGN IMPLICATIONS

We aimed to understand the on-the-move visual behaviors and utilize them to support mobile information needs on different mobile platforms. Using a shadowing study in the wild, we identified 3 visual behaviors patterns: *glance*, *inspect*, and *drift*. Subsequently, we probed on OHMDs and mobile phones in the wild to determine the influence of device platforms on microlearning opportunities created by *drift* behavior. We found that the OHMD platform provides more opportunities for mobile information acquisition, while the mobile phone platform facilitates a limited yet more focused information acquisition experience. In the next section, we discuss methods for supporting information acquisition on the move, then, more specifically, the use of OHMDs for this purpose.

8.1 Design for information acquisition on the move: Information acquisition vs. Navigation

8.1.1 Detection of 'opportunistic' drifts. We gained two key insights on opportunistic behaviors: (1) Out of the three visual behavior patterns, *drift*s provide the most opportunistic moments for mobile information acquisition, (2) Not all *drift*s present opportune moments. Since *drift*s are mostly of sufficient duration for information acquisition (sec 5.1.1: 83.6% in duration above the median, or 22% of commuting duration), we propose a way to detect and utilize opportunistic *drift*s for information acquisition tasks. The first step involves detecting *drift*s. This can be achieved through mobile gaze tracking such as via object/location of eye focus [94], head orientation [84], and alertness detection [85]. After the *drift* visual behavior is identified, the second step involves filtering out moments when users are engaged in nonvisual cognitive tasks (i.e., *active drift*). We can achieve this through cognitive load monitoring with physiological sensing of pupil size, ECG, or EEG [11, 14, 37, 44].

8.1.2 Review vs. New. Since different opportunistic *drift*s are of different durations, information acquisition tasks should be assigned based on the available duration so as to avoid overloading users with information. For example, tasks involving word-learning are not suitably achieved with short *drift*s as they require time and concentration. All participants in *study 2* preferred reviewing

---

[4]Nreal light: https://www.nreal.ai/light

Table 3. Summary of the key findings. (+) indicates positive qualities, (−) indicates negative qualities, and (●) indicates neutral qualities.

Mobile phones	OHMDs / Smart glasses
(+) Allow users to better concentrate on the information acquisition task when no/fewer *signal*s exist	(+) Allow users to better concentrate on the information acquisition task even when *background signal*s are present
(−) Limit the user's ability to be situationally aware and engage in the information acquisition task due to visual field constriction	(+) Allow users to have situational awareness and engage in the information acquisition task
(−) Higher switching cost between the environment and task due to the greater physical effort required	(+) Lower switching cost due to direct gaze interaction
	(−) Can interrupt cognitive processes when sudden visual stimuli (e.g., digital content) appear in front of eyes
(●) Require users to look up and down to memorize words	(●) Require users to focus and defocus to memorize words
	(+) Can use "covert learning" to continue information acquisition
(−) Can cause neck fatigue due to the head-down posture during interaction	(+) Do not cause neck fatigue as the head-up posture is more relaxed
(−) Can be challenging to concentrate on words with vehicle vibration and motion	(+) Easy to concentrate on words with vehicle vibration and motion
	(−) Can be socially awkward when words appear in front of faces
	(−) Can cause eye strain when users are not accustomed to OHMDs. More training is required to increase familiarity levels.

materials they had previously seen instead of learning new ones during short *drift*s. As expressed by P2: *"[before boarding the bus] The interval is too short for me to intake new information. I think I'd rather review old words now since I need less time for each word."*

However, it is difficult to predict the duration of *drift*s due to the dynamic nature of *signal*s. It is thus helpful to assign a shorter information acquisition task (e.g., review session) once a *drift* has been identified, regardless of its predicted duration. As elucidated in sec 7.1.2, when users tried to memorize words, they switched from *drift*s to a series of *glance*s and *inspect*s. In this case, the system should initiate a longer information task (e.g., a learning session with new words) upon detection of user behavior from a previous information acquisition task. If there are unexpected *signal*s from the environment, users may switch away from word-memorizing behaviors, and such deviations can be used to stop/pause information tasks.

8.1.3 Automatic vs. Manual control. Automatic pushing has many advantages: participants do not have to retrieve information manually, and the pushed content reminded and encouraged

them to take on the information acquisition opportunities available from their devices. Despite these benefits, some participants considered automatic pushing to be inappropriate and annoying, primarily when they were already engaged in something else (e.g., pondering, relaxation, sec 7.1.3). Whenever their environmental conditions were changing and uncertainty high, users frequently switched their attention between information acquisition and commute-related tasks, preferring to eliminate acquiring digital information. Thus, we suggest automatic pushing for kick-starting information acquisition (e.g., microlearning), then allowing users to manually postpone or cancel the session through peripheral interactions if preferred. We present these suggestions in a flowchart, Fig 9 (see Appendix B.4 for the proposed solution for microlearning).
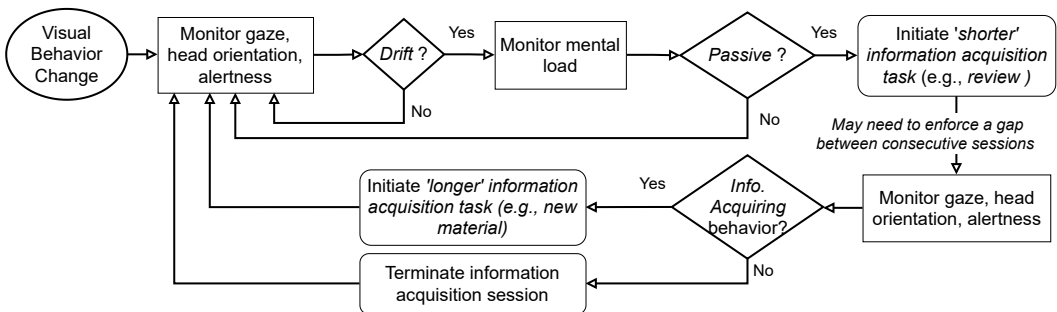

Fig. 9. A flowchart for automatic detection and utilization of opportunistic *drift*s for information acquisition. See Appendix B.4 for proposed solution for microlearning which also considers consecutive session limits to minimize fatigue. In practical usage, designers also need to consider user safety, which is not captured in this diagram.

8.2 Design for information acquisition on OHMDs

8.2.1 Divided vs. Focused Attention. Due to the dynamic change of *signal*s on the move, we observed that users could engage in two information acquisition modes: *divided-attention mode* and *focused-attention mode*. Users enter the focused-attention mode when the frequency and intensity of the *signal*s are low and the divided-attention mode when the frequency or intensity of *signal*s is high. From our observations, OHMDs are suitable for the divided-attention mode as the see-through nature of the display screen allows users to use their peripheral vision to monitor *signal*s and divide their attention between multiple tasks. On the other hand, mobile phones suit the focused-attention mode as users can focus on the screen without as many distractions from their surroundings. To strike a balance between the two modes, we propose two design solutions. The first design changes the OHMD foreground from transparent to opaque (i.e., words appear on an opaque overlay or "light mode" [33]) when the focused-attention mode is triggered. This can improve users' concentration as it can block out the physical background and *signal*s. The second design uses a combination system as shown in Fig 10 to allow for flexible transitions between mobile phones and OHMD. However, we did not test the combination system and recommend further exploration of it. A balance needs to be struck between platform limitations and allowing users to maintain situational awareness as well as focus on display. Adjusting the opacity of OHMD content may help users focus on what is displayed but affect their level of situational awareness. Future investigations should undertake these design challenges.

8.2.2 Peripheral notifications to minimize interruptions. To render users more control of their OHMDs, information triggering applications can subtly notify users of any upcoming information
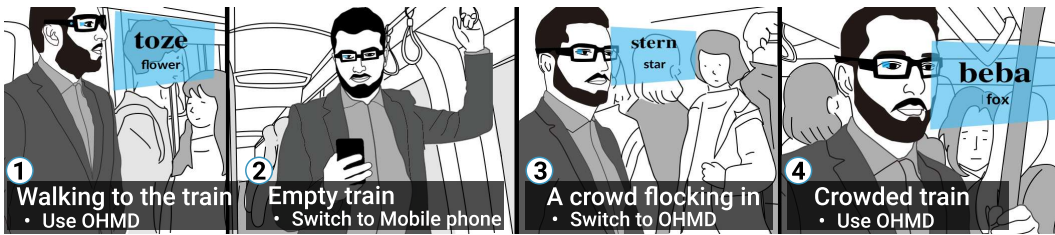
Fig. 10. A potential scenario combining the use of OHMD and mobile phone to support commuters with microlearning

acquisition sessions and allow them to ignore them if they wish. Moreover, to minimize the interruptions of such notifications, applications can either use nonvisual notifications or minimal UIs that reduce distractions from visual notifications. For example, NotifEye [59] used a minimalistic playful user interface to present notifications on the user's focal vision while Luyten et al. [60] explored a visual language on peripheral vision.

8.2.3 Peripheral interactions for postponing or canceling. Participants preferred using different interaction techniques to control microlearning sessions (e.g., continue, postpone, cancel) and learn content (e.g., word duration, appearing/disappearing) on OHMDs. The majority of participants did not prefer voice or touch input on the OHMD frame, all of which are standard OHMD interaction techniques, as they could be socially disturbing [43] on public transport. Some participants preferred hands-free interactions, such as gaze and head movements, while a few suggested using the mobile phone as a controller. This indicates that there is a need for a set of socially acceptable and user-definable [87] OHMD interactions to cater to different information acquisition needs on the move. Furthermore, the interactions should not require any extra effort on users' part to initiate or stop since they should be able to divert their attention to their surrounding *signal*s when necessary. Given that undesirable triggers can annoy users, we suggest using gaze behavior (e.g., blinking pattern) to cancel sessions since users can interact with the system without diverting their attention.

9   LIMITATIONS

We showed that our results apply to tech-savvy participants who are potential early adopters of OHMDs. Since device usage and motivation for information acquisition depend on users' technology acceptance and personal needs [28, 92], we should generalize these results to other populations with care.

We also note that novelty effects might have affected the *OHMD probe*. Participants had only tested the OHMD during our training sessions before the actual experiment compared to a more extended history of mobile phone usage.

In addition, our observations were limited to a relatively small number of (homogeneous) participants who had motivation to acquire information during commuting, which may be insufficient for understanding the larger range of user behaviors. However, we believe that we have covered all prominent behaviors with our participant count, which achieved data saturation during the analysis.

To mitigate the aforementioned limitations, large-scale longitudinal studies with different user groups (e.g., heterogeneous sample) and different OHMD prototypes are needed to identify the long-term effects and increase the generalizability of the results. Although our focus was not

on micro-level eye movements (e.g., saccades, fixation [54]), using advanced apparatus with eye-tracking (e.g., [39]) can help to uncover the links between identified visual behaviors and micro-level eye movements and serve to build upon our proposed system (Fig 9).

10 CONCLUSIONS AND FUTURE WORK

Supporting ubiquitous information acquisition during on-the-move scenarios is challenging as users' surroundings dynamically change and can easily distract them. This research investigates common visual behaviors when users commute and differentiates between three types of visual behaviors: *glance*, *inspect*, and *drift*. We identified how they are affected by *background signals* and *trigger signals* in the environment and the resulting impact on users' receptivity to microlearning. The study also investigates the existing challenges of using two types of mobile platforms in such situations; the emerging OHMDs that provide enhanced information acquisition support on the move and existing mobile phones that offer more focused information acquisition with fewer *external interruptions*. Based on the limitations that exist in these platforms, we highlight the opportunities for technological advancement and better designs to support on-the-move information acquisition tasks. This research also suggests the potential coexistence of OHMDs with mobile phones to offer benefits beyond what either platform can individually provide. While microlearning is the domain of our investigation, we expect that our results apply to other forms of information acquisition aimed at improving users' productivity on the move. Although this study mainly examines the visual behaviors of participants, we note that there are other essential factors such as social [13], affective [53], and environmental factors [51] that can influence users' receptivity to *information acquisition* and learning. These factors need to be considered as a whole to enhance information acquisition experiences effectively and help users become more receptive to such tasks. Therefore, we encourage longitudinal field studies for insight into how the aforementioned factors contribute to information *acquisition* and *retention* across platforms.

11 DATA COLLECTION

The data was collected on public transport just before the COVID-19 pandemic situation arose in Singapore (i.e., Sep. 2019 - Mar. 2020).

REFERENCES

[1] Evan Ackerman. 2021. Bosch Gets Smartglasses Right With Tiny Eyeball Lasers. https://spectrum.ieee.org/tech-talk/consumer-electronics/gadgets/bosch-ar-smartglasses-tiny-eyeball-lasers Retrieved February 06, 2021.

[2] Hironori Akechi, Atsushi Senju, Helen Uibo, Yukiko Kikuchi, Toshikazu Hasegawa, and Jari K. Hietanen. 2013. Attention to Eye Contact in the West and East: Autonomic Responses and Evaluative Ratings. 8, 3 (2013), e59312. https://doi.org/10.1371/journal.pone.0059312

[3] APA. 2021. alertness - APA Dictionary of Psychology. https://dictionary.apa.org/alertness Retrieved May 5, 2021.

[4] APA. 2021. Multitasking: Switching costs. https://www.apa.org/research/action/multitask Retrieved May 5, 2021.

[5] 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 (June 2019), 130–133. https://doi.org/10.1016/j.jenvp.2019.04.003

[6] Onur Asan and Enid Montague. 2014. Using video-based observation research methods in primary care health encounters to evaluate complex interactions. *Journal of Innovation in Health Informatics* 21, 4 (Aug. 2014), 161–170. https://doi.org/10.14236/jhi.v21i4.72

[7]   David P. Ausubel. 2000. *The Acquisition and Retention of Knowledge: A Cognitive View.* Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-015-9454-7

[8]   Ronald T Azuma. 1997. A Survey of Augmented Reality. 6, 4 (1997), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

[9]   Corey H. Basch, Danna Ethan, Patricia Zybert, and Charles E. Basch. 2015. Pedestrian Behavior at Five Dangerous and Busy Manhattan Intersections. *Journal of Community Health* 40, 4 (Aug. 2015), 789–792. https://doi.org/10.1007/s10900-015-0001-9

[10] Jennifer S. Beaudin, Stephen S. Intille, Emmanuel Munguia Tapia, Randy Rockinson, and Margaret E. Morris. 2007. Context-Sensitive Microlearning of Foreign Language Vocabulary on a Mobile Device. In *Ambient Intelligence.* Vol. 4794. Springer Berlin Heidelberg, Berlin, Heidelberg, 55–72. https://doi.org/10.1007/978-3-540-76652-0_4

[11] Chris Berka, Daniel J. Levendowski, Michelle N. Lumicao, Alan Yau, Gene Davis, Vladimir T. Zivkovic, Richard E. Olmstead, Patrice D. Tremoulet, and Patrick L. Craven. 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine* 78, 5 Suppl (May 2007), 231–244.

[12] Jelmer P. Borst, Niels A. Taatgen, and Hedderik van Rijn. 2015. What Makes Interruptions Disruptive?: A Process-Model Account of the Effects of the Problem State Bottleneck on Task Interruption and Resumption. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15).* ACM, New York, NY, USA, 2971–2980. https://doi.org/10.1145/2702123.2702156

[13] Michael P Breen. 1985. The social context for language learning—a neglected situation? *Studies in second language acquisition* 7, 2 (1985), 135–158.

[14] Anne-Marie Brouwer, Maarten A Hogervorst, Jan B F van Erp, Tobias Heffelaar, Patrick H Zimmerman, and Robert Oostenveld. 2012. Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering* 9, 4 (jul 2012), 045008. https://doi.org/10.1088/1741-2560/9/4/045008

[15] Mihai Bâce, Sander Staal, and Andreas Bulling. 2020. Quantification of Users' Visual Attention During Everyday Mobile Device Interactions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376449

[16] Carrie J. Cai, Philip J. Guo, James R. Glass, and Robert C. Miller. 2015. Wait-Learning: Leveraging Wait Time for Second Language Education. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15.* ACM Press, Seoul, Republic of Korea, 3701–3710. https://doi.org/10.1145/2702123.2702267

[17] Carrie J. Cai, Anji Ren, and Robert C. Miller. 2017. WaitSuite: Productive Use of Diverse Waiting Moments. *ACM Transactions on Computer-Human Interaction* 24, 1 (March 2017), 7:1–7:41. https://doi.org/10.1145/3044534

[18] Marisa Carrasco. 2011. Visual attention: The past 25 years. *Vision Research* 51, 13 (July 2011), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

[19] Soledad Castellano and Inmaculada Arnedillo-Sánchez. 2016. Sensorimotor Distractions When Learning With Mobile Phones On-the-move. (2016), 5. https://eric.ed.gov/?id=ED571434

[20] Sonya Cates, Daniel Barron, and Patrick Ruddiman. 2017. MobiLearn go: mobile microlearning as an active, location-aware game. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17* (Vienna, Austria, 2017). ACM Press, 1–7. https://doi.org/10.1145/3098279.3122146

[21] Marvin M. Chun and Jeremy M. Wolfe. 2005. *Visual Attention.* John Wiley & Sons, Ltd, Chapter 9, 272–310. https://doi.org/10.1002/9780470753477.ch9 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470753477.ch9

[22] Charles E Connor, Howard E Egeth, and Steven Yantis. 2004. Visual attention: bottom-up versus top-down. *Current biology* 14, 19 (2004), R850–R852.

[23] S. Consolvo and M. Walker. 2003. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing* 2, 2 (April 2003), 24–31. https://doi.org/10.1109/MPRV.2003.1203750

[24] Juliet Corbin and Anselm Strauss. 2008. *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd edition ed.). Sage publications.

[25] Juliet M. Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology* 13, 1 (March 1990), 3–21. https://doi.org/10.1007/BF00988593

[26] David Dearman and Khai Truong. 2012. Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12.* ACM Press, Austin, Texas, USA, 1391–1400. https://doi.org/10.1145/2207676.2208598

[27] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. 2017. Language learning on-the-go: opportune moments and design of mobile microlearning sessions. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '17* (Vienna, Austria, 2017). ACM Press, 1–12. https://doi.org/10.1145/3098279.3098565

[28] Zoltán Dörnyei and István Ottó. 1998. Motivation in action: A process model of L2 motivation. 4 (1998), 43–69.

[29] Darren Edge, Stephen Fitchett, Michael Whitney, and James Landay. 2012. MemReflex: adaptive flashcards for mobile microlearning. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12).* Association for Computing Machinery, New York, NY, USA, 431–440. https:

//doi.org/10.1145/2371574.2371641

[30] Darren Edge, Elly Searle, Kevin Chiu, Jing Zhao, and James A. Landay. 2011. MicroMandarin: mobile language learning in context. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 3169–3178. https://doi.org/10.1145/1978942.1979413

[31] Stuart Elder and Alex Vakaloudis. 2015. A technical evaluation of devices for smart glasses applications. In *2015 Internet Technologies and Applications (ITA)* (Wrexham, United Kingdom, 2015-09). IEEE, 98–103. https://doi.org/10.1109/ITechA.2015.7317377

[32] Phoebe C. Ellsworth and Linda M. Ludwig. 1972. Visual Behavior in Social Interaction. *Journal of Communication* 22, 4 (Dec. 1972), 375–403. https://doi.org/10.1111/j.1460-2466.1972.tb00164.x

[33] Austin Erickson, Kangsoo Kim, Alexis Lambert, Gerd Bruder, Michael P. Browne, and Gregory F. Welch. 2021. An Extended Analysis on the Benefits of Dark Mode User Interfaces in Optical See-Through Head-Mounted Displays. *ACM Transactions on Applied Perception* 18, 3 (May 2021), 12:1–12:22. https://doi.org/10.1145/3456874

[34] Gerhard Gassler, Theo Hug, and Christian Glahn. 2004. Integrated Micro Learning - An outline of the basic method and first results. *International Conference on Interactive Computer Aided Learning* (2004), 7.

[35] Zhan Guo, Alexandra Derian, and Jinhua Zhao. 2015. Smart Devices and Travel Time Use by Bus Passengers in Vancouver, Canada. 9, 5 (2015), 335–347. https://doi.org/10.1080/15568318.2013.784933

[36] Ewa Gustafsson, Sara Thomée, Anna Grimby-Ekman, and Mats Hagberg. 2017. Texting on mobile phones and musculoskeletal disorders in young adults: A five-year cohort study. *Applied Ergonomics* 58 (2017), 208–214. https://doi.org/10.1016/j.apergo.2016.06.012

[37] Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. 2010. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10* (Copenhagen, Denmark, 2010). ACM Press, 301. https://doi.org/10.1145/1864349.1864395

[38] Jungmin Han, Seon Hee Bae, and Hyeon-Jeong Suk. 2017. Comparison of Visual Discomfort and Visual Fatigue between Head-Mounted Display and Smartphone. *Electronic Imaging* 2017, 14 (Jan. 2017), 212–217. https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-146

[39] Dan Witzner Hansen and Arthur E.C. Pece. 2005. Eye tracking in the wild. *Computer Vision and Image Understanding* 98, 1 (April 2005), 155–181. https://doi.org/10.1016/j.cviu.2004.07.013

[40] John M. Henderson. 2003. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences* 7, 11 (Nov. 2003), 498–504. https://doi.org/10.1016/j.tics.2003.09.006

[41] Natasa Vujica Herzog and Amer Beharic. 2019. Effects of the Use of Smart Glasses on Eyesight. In *Human Systems Engineering and Design II*, Tareq Ahram, Waldemar Karwowski, Stefan Pickl, and Redha Taiar (Eds.). Springer International Publishing, Cham, 808–812. https://doi.org/10.1007/978-3-030-27928-8_123

[42] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology Probes: Inspiring Design for and with Families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 17-24. https://doi.org/10.1145/642611.642616

[43] Perficient Inc. 2021. Mobile Voice Usage Trends in 2020 | Perficient Digital | Perficient Digital Agency. https://www.perficientdigital.com/insights/our-research/voice-usage-trends Retrieved May 5, 2021.

[44] Shamsi T. Iqbal, Piotr D. Adamczyk, Xianjun Sam Zheng, and Brian P. Bailey. 2005. Towards an Index of Opportunity: Understanding Changes in Mental Workload during Task Execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. Association for Computing Machinery, New York, NY, USA, 311–320. https://doi.org/10.1145/1054972.1055016

[45] Ellen Isaacs, Nicholas Yee, Diane J Schiano, Nathan Good, Nicolas Ducheneaut, and Victoria Bellotti. 2009. Mobile Microwaiting Moments: The Role of Context in Receptivity to Content While on the Go. *PARC white paper* (2009), 10.

[46] Yoshio Ishiguro and Jun Rekimoto. 2011. Peripheral vision annotation: noninterference information presentation method for mobile augmented reality. In *Proceedings of the 2nd Augmented Human International Conference on - AH '11*. ACM Press, Tokyo, Japan, 1–5. https://doi.org/10.1145/1959826.1959834

[47] Daniel Kahneman. 1973. *Attention and effort*. Vol. 1063. Citeseer.

[48] Ioanna Katidioti, Jelmer P. Borst, and Niels A. Taatgen. 2014. What happens when we switch tasks: Pupil dilation in multitasking. *Journal of Experimental Psychology: Applied* 20, 4 (Dec. 2014), 380–396. https://doi.org/10.1037/xap0000031

[49] Steven J. Kerr, Mark D. Rice, Yinquan Teo, Marcus Wan, Yian Ling Cheong, Jamie Ng, Lillian Ng-Thamrin, Thant Thura-Myo, and Dominic Wren. 2011. Wearable Mobile Augmented Reality: Evaluating Outdoor User Experience. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry* (Hong Kong, China) *(VRCAI '11)*. Association for Computing Machinery, New York, NY, USA, 209-216. https://doi.org/10.1145/2087756.2087786

[50] Taslim Arefin Khan, Dongwook Yoon, and Joanna McGrenere. 2020. Designing an Eyes-Reduced Document Skimming App for Situational Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. https://doi.org/10.1145/3313831.3376641

[51] Maria Klatte, Kirstin Bergstroem, and Thomas Lachmann. 2013. Does noise affect learning? A short review on noise effects on cognitive performance in children. *Frontiers in Psychology* 4 (2013), 578. https://doi.org/10.3389/fpsyg.2013.00578

[52] Stefan M Knupfer, Vadim Pokotilo, and Jonathan Woetzel. 2021. Elements of success: Urban transportation systems of 24 global cities. https://www.mckinsey.com/business-functions/sustainability/our-insights/elements-of-success-urban-transportation-systems-of-24-global-cities Retrieved May 06, 2021.

[53] L. Knörzer, R. Brünken, and B. Park. 2016. Emotions and multimedia learning: the moderating role of learner characteristics: Emotions in multimedia learning. 32, 6 (2016), 618–631. https://doi.org/10.1111/jcal.12158

[54] Eileen Kowler. 2011. Eye movements: The past 25years. *Vision Research* 51, 13 (July 2011), 1457–1483. https://doi.org/10.1016/j.visres.2010.12.014

[55] Michael F. Land. 2006. Eye movements and the control of actions in everyday life. 25, 3 (2006), 296–324. https://doi.org/10.1016/j.preteyeres.2006.01.002

[56] Jonathan Lazar. 2017. *Research methods in human computer interaction* (2nd edition ed.). Elsevier, Cambridge, MA.

[57] Eckhard Limpert, Werner A. Stahel, and Markus Abbt. 2001. Log-normal Distributions across the Sciences: Keys and Clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. *BioScience* 51, 5 (05 2001), 341–352. https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2 arXiv:https://academic.oup.com/bioscience/article-pdf/51/5/341/26891292/51-5-341.pdf

[58] 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. https://doi.org/10.1016/j.aap.2017.02.005

[59] Andrés Lucero and Akos Vetek. 2014. NotifEye: using interactive glasses to deal with notifications while walking in public. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology* (Funchal, Portugal, 2014) *(ACE '14)*. ACM Press, 1–10. https://doi.org/10.1145/2663806.2663824

[60] Kris Luyten, Donald Degraen, Gustavo Rovelo Ruiz, Sven Coppers, and Davy Vanacken. 2016. Hidden in Plain Sight: an Exploration of a Visual Language for Near-Eye Out-of-Focus Displays in the Peripheral View. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, Santa Clara, California, USA, 487–497. https://doi.org/10.1145/2858036.2858339

[61] Manuela Macedonia, Karsten Müller, and Angela D. Friederici. 2010. Neural Correlates of High Performance in Foreign Language Vocabulary Learning. *Mind, Brain, and Education* 4, 3 (2010), 125–134. https://doi.org/10.1111/j.1751-228X.2010.01091.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-228X.2010.01091.x

[62] W.C. Maples, Wes DeRosier, Richard Hoenes, Rodney Bendure, and Sherl Moore. 2008. The effects of cell phone use on peripheral vision. *Optometry - Journal of the American Optometric Association* 79, 1 (Jan. 2008), 36–42. https://doi.org/10.1016/j.optm.2007.04.102

[63] W. Cheyne McCallum. 2021. Aspects of attention. https://www.britannica.com/science/attention/The-intensity-of-attention Retrieved May 27, 2020 from https://www.britannica.com/science/attention/The-intensity-of-attention.

[64] Christopher A. Monk, J. Gregory Trafton, and Deborah A. Boehm-Davis. 2008. The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology. Applied* 14, 4 (Dec. 2008), 299–313. https://doi.org/10.1037/a0014402

[65] Laura Naismith, Peter Lonsdale, Giasemi Vavoula, and Mike Sharples. 2004. Literature Review in Mobile Technologies and Learning. (2004), 48.

[66] Christie Napa Scollon, Chu-Kim Prieto, and Ed Diener. [n.d.]. Experience Sampling: Promises and Pitfalls, Strength and Weaknesses. In *Assessing Well-Being: The Collected Works of Ed Diener*, Ed Diener (Ed.). Springer Netherlands, 157–180. https://doi.org/10.1007/978-90-481-2354-4_8

[67] Jason Orlosky, Kiyoshi Kiyokawa, and Haruo Takemura. 2014. Managing mobile text in head mounted displays: studies on visual preference and text placement. *ACM SIGMOBILE Mobile Computing and Communications Review* 18, 2 (June 2014), 20–31. https://doi.org/10.1145/2636242.2636246

[68] Antti Oulasvirta, Sakari Tamminen, Virpi Roto, and Jaana Kuorelahti. 2005. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05*. ACM Press, Portland, Oregon, USA, 919. https://doi.org/10.1145/1054972.1055101

[69] Hiltraut M. Paridon and Marlen Kaufmann. 2010. Multitasking in work-related situations and its relevance for occupational health and safety: Effects on performance, subjective strain and physiological parameters. *Europe's Journal of Psychology* 6, 4 (Nov. 2010), 110–124. https://doi.org/10.5964/ejop.v6i4.226

[70] Mark Perry, Kenton O'hara, Abigail Sellen, Barry Brown, and Richard Harper. 2001. Dealing with mobility: understanding access anytime, anywhere. *ACM Transactions on Computer-Human Interaction* 8, 4 (Dec. 2001), 323–347.

https://doi.org/10.1145/504704.504707

[71] Physiopedia. 2021. Text Neck. https://www.physio-pedia.com/Text_Neck Retrieved May 5, 2021.

[72] Elizabeth Quinlan. 2008. Conspicuous Invisibility: Shadowing as a Data Collection Strategy. 14, 8 (2008), 1480–1499. https://doi.org/10.1177/1077800408318318

[73] Philipp A Rauschnabel, Alexander Brem, and Young K Ro. 2015. Augmented Reality Smart Glasses: Definition, Conceptual Insights, and Managerial Importance. (2015), 22.

[74] K. Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124, 3 (Nov. 1998), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

[75] Keith Rayner, Tim J. Smith, George L. Malcolm, and John M. Henderson. 2009. Eye Movements and Visual Encoding During Scene Perception. 20, 1 (2009), 6–10. https://doi.org/10.1111/j.1467-9280.2008.02243.x

[76] Anji Ren. 2015. Pull-To-Refresh and Learn: Leveraging Mobile Email Load Time for Education. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '15* (Seoul, Republic of Korea, 2015). ACM Press, 145–150. https://doi.org/10.1145/2702613.2726968

[77] Marie Russell, Rachel Price, Louise Signal, James Stanley, Zachery Gerring, and Jacqueline Cumming. 2011. What Do Passengers Do During Travel Time? Structured Observations on Buses and Trains. *Journal of Public Transportation* 14, 3 (Sept. 2011), 123–146. https://doi.org/10.5038/2375-0901.14.3.7

[78] Mike Sharples, Inmaculada Arnedillo-Sánchez, Marcelo Milrad, and Giasemi Vavoula. 2009. Mobile Learning. In *Technology-Enhanced Learning: Principles and Products*, Nicolas Balacheff, Sten Ludvigsen, Ton de Jong, Ard Lazonder, and Sally Barnes (Eds.). Springer Netherlands, Dordrecht, 233–249. https://doi.org/10.1007/978-1-4020-9827-7_14

[79] F. Atiyya Shaw, Aliaksandr Malokin, Patricia L. Mokhtarian, and Giovanni Circella. 2019. It's not all fun and games: An investigation of the reported benefits and disadvantages of conducting activities while commuting. 17 (2019), 8–25. https://doi.org/10.1016/j.tbs.2019.05.008

[80] Amanda Spink, Charles Cole, and Mary Waller. 2008. Multitasking behavior. *Annual Review of Information Science and Technology* 42, 1 (2008), 93–118. https://doi.org/10.1002/aris.2008.1440420110 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/aris.2008.1440420110

[81] Michael Spitzer, Ibrahim Nanic, and Martin Ebner. 2018. Distance Learning and Assistance Using Smart Glasses. 8, 1 (2018), 21. https://doi.org/10.3390/educsci8010021

[82] Despina Stavrinos, Katherine W. Byington, and David C. Schwebel. 2011. Distracted walking: Cell phones increase injury risk for college pedestrians. *Journal of Safety Research* 42, 2 (April 2011), 101–107. https://doi.org/10.1016/j.jsr.2011.01.004

[83] Julian Steil, Philipp Müller, Yusuke Sugano, and Andreas Bulling. 2018. Forecasting user attention during everyday mobile interactions using device-integrated and wearable sensors. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '18*. ACM Press, Barcelona, Spain, 1–13. https://doi.org/10.1145/3229434.3229439

[84] Rainer Stiefelhagen and Jie Zhu. 2002. Head Orientation and Gaze Direction in Meetings. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI EA '02)*. Association for Computing Machinery, New York, NY, USA, 858-859. https://doi.org/10.1145/506443.506634

[85] Benjamin Tag, Andrew W. Vargo, Aman Gupta, George Chernyshov, Kai Kunze, and Tilman Dingler. 2019. Continuous Alertness Assessments: Using EOG Glasses to Unobtrusively Monitor Fatigue Levels In-The-Wild. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–12. https://doi.org/10.1145/3290605.3300694

[86] Andrew Trusty and Khai N. Truong. 2011. Augmenting the web for second language vocabulary learning. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 3179–3188. https://doi.org/10.1145/1978942.1979414

[87] Ying-Chao Tung, Chun-Yen Hsu, Han-Yu Wang, Silvia Chyou, Jhe-Wei Lin, Pei-Jung Wu, Andries Valstar, and Mike Y. Chen. 2015. User-Defined Game Input for Smart Glasses in Public Space. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 3327–3336. https://doi.org/10.1145/2702123.2702214

[88] Roberto Valenti, Nicu Sebe, and Theo Gevers. 2012. Combining Head Pose and Eye Location Information for Gaze Estimation. *IEEE Transactions on Image Processing* 21, 2 (Feb. 2012), 802–815. https://doi.org/10.1109/TIP.2011.2162740

[89] Zheng Wang and John M Tchernev. 2012. The "myth" of media multitasking: Reciprocal dynamics of media multitasking, personal needs, and gratifications. *Journal of Communication* 62, 3 (2012), 493–513.

[90] Stuart Webb. 2007. The Effects of Repetition on Vocabulary Knowledge. *Applied Linguistics* 28, 1 (03 2007), 46–65. https://doi.org/10.1093/applin/aml048

[91] Christopher D. Wickens. 1991. Processing resources and attention. In *Multiple-task performance*. CRC Press.

[92] Barbara H. Wixom and Peter A. Todd. 2005. A Theoretical Integration of User Satisfaction and Technology Acceptance. *Information Systems Research* 16, 1 (2005), 85–102. http://www.jstor.org/stable/23015766

[93] Jacob O. Wobbrock. 2019. Situationally-Induced Impairments and Disabilities. In *Web Accessibility*, Yeliz Yesilada and Simon Harper (Eds.). Springer London, London, 59–92.   https://doi.org/10.1007/978-1-4471-7440-0_5

[94] Xucong Zhang, Yusuke Sugano, and Andreas Bulling. 2017. Everyday Eye Contact Detection Using Unsupervised Gaze Target Discovery. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 193-203.   https://doi.org/10.1145/3126594.3126614

[95] Yue Zhao, Tarmo Robal, Christoph Lofi, and Claudia Hauff. 2018. Stationary vs. Non-stationary Mobile Learning in MOOCs. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization - UMAP '18* (Singapore, Singapore, 2018). ACM Press, 299–303.   https://doi.org/10.1145/3213586.3225241

A   STUDY 1: SHADOWING

A.1   Video Coding

Please refer to Table 4 for details.

A.2   Data triangulation

Fig 11 shows the triangulation of four data sources; contextual inquiry notes, interview transcriptions, observation notes, and video recordings.
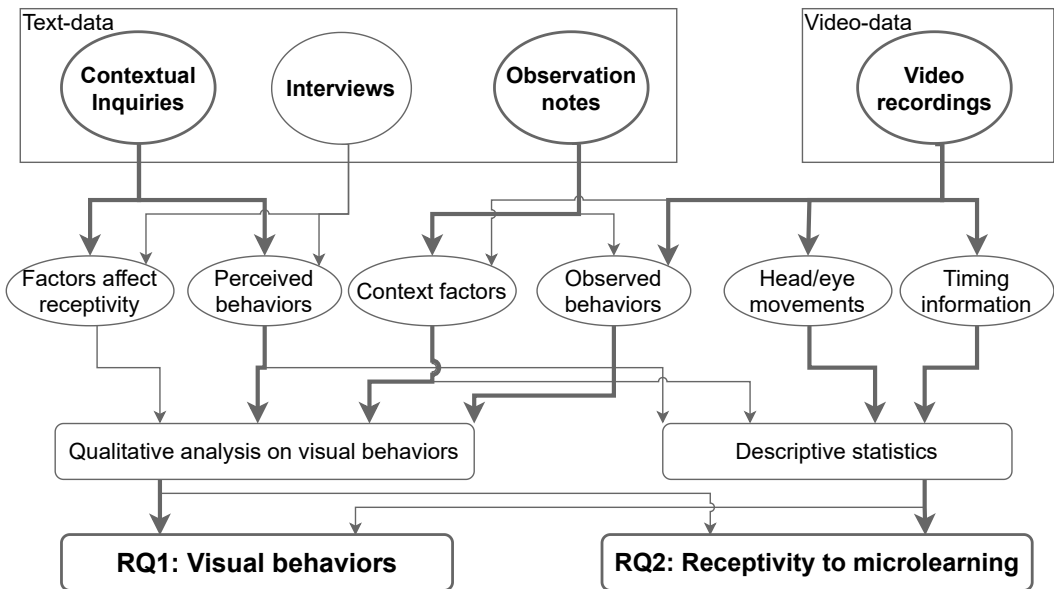

Fig. 11. The triangulation of data sources towards research questions. The thickness of the lines/arrows represents the contribution of data to the analysis.

A.3   Visual behavior distributions

The duration distributions of *inspect* and *drift* were right-skewed and with log transformation, both approximated to normal distributions (*Anderson-Darling*= 0.61, $p = 0.64$ for *inspect* and *Anderson-Darling*= 1.12, $p = 0.30$ for *drift*), while *glance* did not (Table 5). This finding aligns with literature as log-normal distributions commonly represent typical behaviors of humans, economics, biology, and so forth [57].

Table 4. Video coding for a participant (P1) up to 25 minutes. Note: Pattern/visual behavior was added later. Here *Intervene* represents contextual inquiries that did not belong to natural behavior.

P1	9:25-11:05	11/6	Bus/Walking	Home to school	
Time	Posture	Location	Familiarity/Crowdedness	Visual behavior and primary task	Pattern
9:24:35	Walking	Road	Familiar	looking around	*Drift*
9:24:42	Walking	Road	Less crowded	texting on phone	*Inspect*
9:24:53	Walking	Road		looking around (left and right)	*Drift*
9:26:07	Walking	Road		reading on phone	*Inspect*
9:26:13	Walking	Road		looking around	*Drift*
9:26:23	Walking	Road		check time on watch	*Glance*
9:26:24	Walking	Road		looking around	*Drift*
9:26:31	Walking	Road		texting, holding phone closer to eyes	*Inspect*
9:26:49	Walking	Road		looking around	*Drift*
9:27:43	Walking	Road		texting on phone	*Inspect*
9:27:45	Walking	Road		check for a bus	*Glance*
9:27:46	Walking	Road		texting on phone	*Inspect*
9:29:25	Walking	Road		looking around	*Drift*
9:30:02	Walking	Road		engaging with phone	*Inspect*
9:30:07	Walking	Road		check for a bus	*Glance*
9:30:11	Walking	Road		reading on phone	*Inspect*
9:31:01	Walking	Road		looking around	*Drift*
9:31:24	Walking	Road		contextual inquiry	*Intervene*
9:33:24	Standing	Bus-stop	Less-crowded	engaging with mobile	*Inspect*
9:33:46	Standing	Bus-stop		looking around	*Drift*
9:33:49	Standing	Bus-stop		check and notice the bus	*Glance*
9:33:53	Standing	Bus-stop		keep looking at the bus	*Inspect*
9:34:52	Standing	Bus-stop		looking around	*Drift*
9:35:35	Standing	Bus-stop		glance at the coming bus	*Glance*
9:35:37	Walking	Bus-stop		looking around	*Drift*
9:35:41	Walking	Bus-stop		check the opening bus door	*Glance*
9:35:42	Walking	Bus-stop		keep looking at path/passengers, onboarding bus	*Inspect*
9:36:41	Walking	Bus-stop		check for a seat	*Glance*
9:36:45	Walking	Bus-stop		keep looking at the surroundings to avoid bumping	*Inspect*
9:37:59	Sitting	Bus	Many empty seats	looking around after sitting	*Drift*
9:38:14	Sitting	Bus		engaging with phone	*Inspect*
9:39:38	Sitting	Bus		check the crowd	*Glance*
9:39:41	Sitting	Bus		doing nothing, looking ahead	*Drift*
9:41:14	Sitting	Bus		conversing with a neighboring passenger	*Inspect*
9:42:01	Sitting	Bus		check the surrounded crowd	*Glance*
9:42:03	Sitting	Bus		doing nothing, looking ahead	*Drift*
9:47:05	Sitting	Bus		check the outside bus stop	*Glance*
9:47:10	Sitting	Bus		watching the outside scenery	*Inspect*
9:47:28	Sitting	Bus		doing nothing, waiting for alight	*Drift*
9:47:41	Sitting	Bus		check the bus door	*Glance*
9:47:45	Sitting	Bus		looking around	*Drift*

Glance took the least amount of time, i.e., shortest duration (Table 5 and Fig 4 (a) overall), though the frequency to duration (percentage) ratio was the highest. This implies that *glance* has a supporting role for other behaviors given the higher frequency of occurrence.

Table 5. Distribution of visual behavior patterns; duration (in seconds; with order '*percentage*, *mean* (*sd*) [*min*, *max*], *median* (*IQR*)'), and frequency ('*percentage*, *patterns per hour*'). Note: Calculations of duration or frequency percentages were performed after removing the contextual inquiries since they were not part of the observed activities.

Visual behavior pattern	Distribution	Duration		Frequency
Glance	-	1.7%, 1.7 (1.2) [0, 4]	, 1 (1 − 3)	33.2%, 34.3
Inspect	∼ log-normal	72.2%, 53.7 (83.4) [3, 695]	, 27 (12 − 64)	46.8%, 48.3
Drift	∼ log-normal	26.1%, 45.6 (48.4) [4, 302]	, 33 (13 − 60)	20.0%, 20.6

A.4 Statistics of visual behaviors

Table 6 shows the associated data with Fig 4.

A.5 Monitored Context Changes
- Walking (after checking the road, while checking the road, after walking for a long time, after some filler activities such as phone engagements, walking fast/slow, walking in paths with turns and twists, walking in a crowd)
- Waiting for a vehicle (posture: standing/sitting, while checking the vehicle: a vehicle is/is not approaching, after checking the vehicle: correct/wrong vehicle, after checking the vehicle schedule: arriving/not arriving soon)
- On vehicle (posture: standing/sitting, same posture for some time: 1-2min/10min/30min/60min, after/while looking around/outside, doing nothing, after/during filling activities)
- Crowd changes (increase/decrease)
- Transitions (walking to/from the vehicle, walking to/from waiting, vehicle to/from waiting, vehicle to vehicle, few minutes before transitions, few minutes after transitions)

A.6 Contextual Inquiry Topics
- Willingness to learn 3-6 words now (1-5 scale), reasons
- Factors affect the willingness, how they affect, why they affect
- Visual attention paid to the surroundings/context now (1-5 scale), reasons, whether it affects the willingness, if so how and why

A.7 Interview Topics

A.7.1 Pre-study form and interview (for participants selection).
- Demographics (age, gender, native language, education, employment)
- Commuting behaviors (mediums, activities that require commuting, duration/distance of commuting, and schedule)
- Second language learning (second language/s, reasons for learning, time allocated for learning, current fluency with the second language, apps using/used to support learning)
- Second language app usage (usage pattern, usage history, usage pattern during commuting, pain points of app usage in commuting, reasons for app usage/not usage)
- Factors affect language learning during commuting (factors, impact on learning, reasons)

Table 6. Duration (in seconds; in order '*percentage, mean (sd) [min, max]*') and frequency ('*percentage, patterns per hour*') of visual behaviors of participants based on video analysis. All values are rounded to the nearest corresponding decimal place. [a] includes mobile language learning during the commute.

Behavior	Glance Duration	Frequency	Inspect [a] Duration	Frequency	Drift Duration
Overall					
Overall	1.7%, 1.7 (1.2) [0, 4]	33.2%, 34.3	72.2%, 53.7 (83.4) [3, 695]	46.8%, 48.3	26.1%, 45.6 (48.4) [
P1	1.7%, 1.0 (0.7) [0, 3]	22.8%, 22.4	50.1%, 47.6 (41.9) [3, 183]	38.6%, 37.9	48.2%, 45.8 (66.0) [
P6	1.8%, 1.2 (0.9) [0, 4]	38.7%, 55.2	82.3%, 39.8 (42.9) [3, 220]	52.2%, 74.5	15.9%, 44.0 (31.8) [9
Commuting Stage: Overall					
Walking for commuting*	2.1%, 1.9 (1.4) [2, 4]	29.0%, 39.0	71.4%, 36.4 (37.7) [3, 206]	52.4%, 70.6	26.5%, 38.0 (32.4) [
Waiting for commuting	2.3%, 2.2 (1.2) [1, 4]	29.0%, 36.9	73.6%, 45.4 (48.3) [4, 199]	46.0%, 58.4	24.1%, 27.4 (20.1) [
On-vehicle	1.4%, 1.6 (1.2) [0, 4]	37.2%, 32.0	72.1%, 68.2 (109.9) [3, 695]	44.3%, 38.1	26.5%, 60.3 (61.4) [9
Posture: Overall					
Walking	1.7%, 2.1 (1.4) [0, 4]	24.3%, 27.9	76.1%, 41.6 (45.7) [3, 206]	57.2%, 65.8	22.2%, 37.6 (33.2) [
Standing	1.8%, 1.4 (1.1) [0, 4]	39.0%, 45.2	84.0%, 55.2 (105.8) [3, 695]	47.3%, 54.8	14.2%, 32.3 (20.0) [
Sitting	1.7%, 2.2 (1.3) [0, 4]	34.3%, 26.8	54.8%, 76.9 (79.6) [5, 287]	32.8%, 25.7	43.5%, 60.9 (65.7) [9
P3 only					
Overall	0.6%, 1.8 (1.2) [0, 4]	20.5%, 12.4	65.2%, 94.9 (170.5) [4, 695]	41.0%, 24.7	34.2%, 49.7 (37.7) [8
Road	2.5%, 2.3 (0.7) [2, 4]	28.6%, 35.8	53.7%, 36.0 (32.6) [5, 70]	42.9%, 53.7	43.8%, 44.0 (42.4) [
Station	0.4%, 1.0 (1.0) [0, 2]	9.0%, 14.1	71.8%, 26.1 (21.8) [4, 72]	63.6%, 98.8	27.8%, 50.5 (52.2) [
Metro	0.5%, 1.7 (1.4) [0, 4]	21.9%, 10.3	65.4%, 160.7 (226.2) [6, 695]	31.2%, 14.7	34.1%, 143.8 (55.0)
Walking	1.3%, 2.5 (0.7) [2, 3]	14.3%, 18.8	75.8%, 29.1 (20.1) [4, 70]	71.4%, 93.7	22.9%, 44.0 (42.4) [
Standing	0.1%, 1.0 (1.0) [0, 2]	9.1%, 2.5	83.1%, 294.5 (314.9) [16, 695]	36.4%, 10.2	16.8%, 39.7 (22.3) [8
Sitting	1.1%, 1.7 (1.4) [0, 4]	28.0%, 22.7	38.6%, 71.5 (95.2) [6, 245]	24.0%, 19.4	60.3%, 150.7 (73.5)

A.7.2 Post-study interview.

- Reasons for visual behavior changes (explain the scenarios, if requires show video recording)
- Visual attention changes with crowdedness, path complexity, movements/motion, unfamiliarity, and posture
- Suitable/appropriate moments for microlearning during commuting, reasons/factors
- Unsuitable/inappropriate moments for microlearning during commuting, reasons/factors
- Why not use moment (1,2...x) today for vocabulary learning (when they seem good for shadower), reasons
- Expectations for microlearning (minimum duration, number of words)

B STUDY 2: TECHNOLOGY PROBE

B.1 Procedure

Before starting the probe, the experimenters provided a training session for the participants to familiarize themselves with the microlearning app on their respective platforms. Once each participant started commuting, two experimenters comprising the main experimenter and assistant experimenter followed them. After noting down context factors such as signals (Appendix A.5), the main experimenter would trigger a microlearning session when a *drift* of more than 10-seconds was observed during commuting. A 10-second threshold was used to minimize recognition errors. Then the main experimenter would conduct a contextual inquiry after the participant completes the microlearning session (see Appendix B.2 for inquiry topics). The main experimenter kept the gap between the two triggers at a minimum of 5 minutes to reduce interruptions. On occasion, inquiries related to 2-3 consecutive sessions were conducted together (within 2-3 minutes). The assistant experimenter video recorded the whole study, mainly focusing on participant context and visual behaviors.

At the end of the commute, the main experimenter carried out a 30-40 minute semi-structured interview with audio recording to assess the participants' experience with microlearning on their platform (see Appendix B.3 for interview topics). Whenever participants experienced difficulty recalling certain details, the experimenter replayed the specific instances from the video recordings to help them remember.

B.2 Contextual Inquiry Topics

- Willingness to microlearning (1-5 scale), reasons
- Concentration on words (1-5 scale) and surroundings (1-5 scale), reasons, distractions from surroundings
- Experience related devices after microlearning (comfortableness, issues, switching to learning, appearance/notifications), reasons
- Preference for presentation (appearing, disappearing, duration), reasons
- (For OHMD users) Experience/expectation with the mobile phone under similar situations, reasons

B.3 Interview Topics

B.3.1 Pre-study form and interview (for participants selection).

- Same questions used in A.7.1
- Mobile phone operating system
- Vision-related questions (vision corrections, usage of spectacles/contact lens)

B.3.2 Post-study interview.

- Experience of microlearning on the platform (overall, comfortableness of device usage, issues: social & technical, automatic triggering, switching to learning), reasons
- Effects of situational factors and environmental interruptions on microlearning (situations easy/difficult to remember words, easy/difficult to concentrate on words), reasons
- Preference for presentation (appearing as a group, duration, number of words, learning vs. reviewing), reasons
- Expected interactions for the platform (controlling microlearning session, manual vs. automatic triggering)
- (For mobile phone users) Experience with notifications

- (For OHMD users) Experience/expectation/comparisons with the mobile phone under similar situations, reasons

B.4 Utilizing *drift*s for microlearning

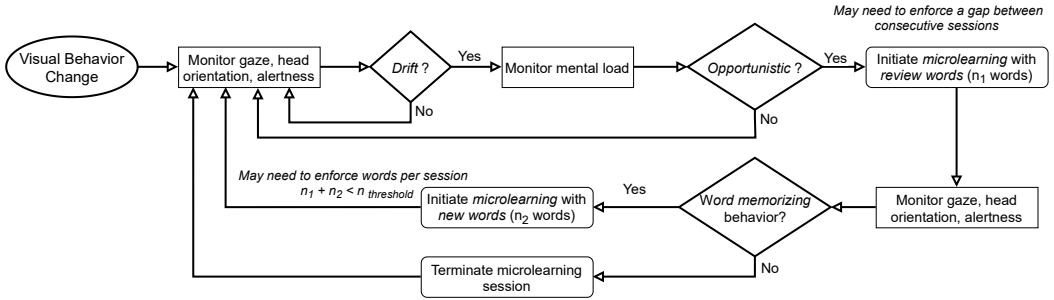Fig 12 shows our proposed solution for utilizing *drift*s for microlearning.

Fig. 12. A flowchart for automatic detection and utilization of opportunistic *drift*s for microlearning. *'Opportunistic' drift*s are determined by mental workload (sec 8.1.1) and sufficient duration (sec 5.2.1). Hence, before triggering the microlearning session, the duration of *drift* behavior should exceed a customizable threshold. Additionally, conducting multiple sessions consecutively is potentially suboptimal since users may feel tired, affecting their ability to learn attentively. The maximum consecutive session count and the minimum time gap between sessions should be customized accordingly. Moreover, the number of words per session (learning or reviewing) should be customizable to cater to different user learning preferences. When the user requests to terminate the session in order to act upon dynamic *signal*s (sec 8.1.3), the system should reset to the initial monitoring state. Practical implementation will also depend on the mobile platform since *'word memorizing behavior'* is influenced by the platform (sec 7.1.2). Further user studies are required to determine the subjective parameters, such as the time gap between sessions.