ParaGlassMenu: Towards Social-Friendly Subtle Interactions in Conversations

Runze Cai
runze.cai@u.nus.edu
National University of Singapore
NUS-HCI Lab, School of Computing
Singapore

Nuwan Janaka
nuwanj@comp.nus.edu.sg
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
CNRS@CREATE LTD, 1 Create Way, #08-01 CREATE Tower, 138602
Singapore

Minghui Sun
smh@jlu.edu.cn
Jilin University
College of Computer Science and Technology
Changchun, Jilin, China

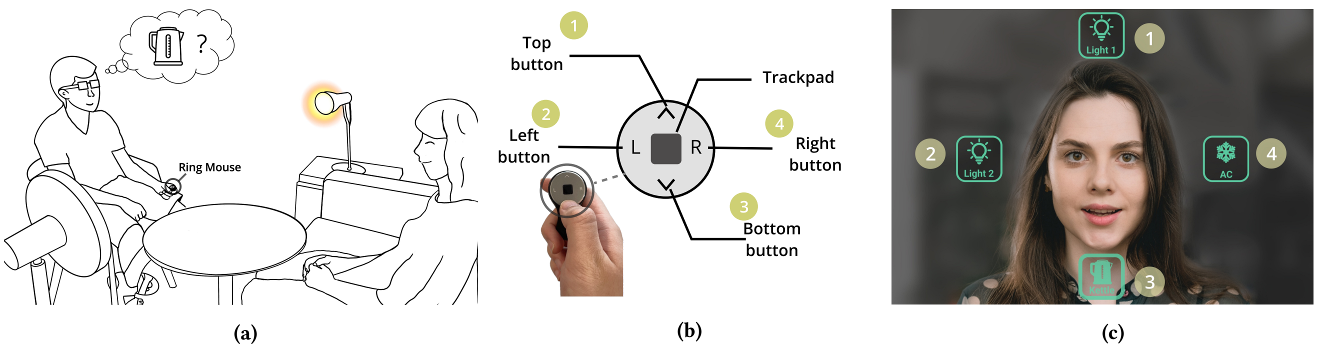


Figure 1: Using *ParaGlassMenu* (consisting of an Optical See-Through Head-Mounted Display and ring mouse) for digital interactions in social settings. (a) A host is conversing with a guest in the living room while he wants to check the water boiling status in the kitchen using *ParaGlassMenu*. (b) The host can operate the menu subtly using a ring mouse, which consists of four clickable buttons and one trackpad. (c) A semi-transparent OHMD menu is displayed circularly surrounding the conversation partner's face (located at the same depth as the face). The yellow labels stand for the mappings between icons and buttons. To check the kettle's menu, the host can click the *bottom* button on the ring mouse.

ABSTRACT

Interactions with digital devices during social settings can reduce social engagement and interrupt conversations. To overcome these drawbacks, we designed *ParaGlassMenu*, a semi-transparent circular menu that can be displayed around a conversation partner's face on Optical See-Through Head-Mounted Display (OHMD) and interacted subtly using a ring mouse. We evaluated *ParaGlassMenu* with several alternative approaches (Smartphone, Voice assistant, and Linear OHMD menus) by manipulating Internet-of-Things (IoT) devices in a simulated conversation setting with a digital partner. Results indicated that the *ParaGlassMenu* offered the best overall performance in balancing social engagement and digital interaction needs in conversations. To validate these findings, we conducted a second study in a realistic conversation scenario involving commodity IoT devices. Results confirmed the utility and social acceptance of the *ParaGlassMenu*. Based on the results, we discuss implications for designing attention-maintaining subtle interaction techniques on OHMDs.

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3581065

CCS CONCEPTS

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

KEYWORDS

IoT manipulation, circular menu, ring interaction, social interaction, conversation, HMD, smart glasses, Internet-of-Things

ACM Reference Format:
Runze Cai, Nuwan Janaka, Shengdong Zhao, and Minghui Sun. 2023. *ParaGlassMenu*: Towards Social-Friendly Subtle Interactions in Conversations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3544548.3581065

1 INTRODUCTION

In an ideal world, face-to-face social interactions are the best when all parties involved give undivided attention to one another. However, real-world situations are often more complex. Considering the following two scenarios: a) John is living alone in his apartment and has decided to host a party in his place. After the arrival of the guests, as the only host, he needs to juggle between the needs of chatting with the guests with the other host duties, including preparing food and drinks, adjusting the environment to make it more comfortable for the guests, etc. b) John is asked to join an ad hoc in-person meeting after work, preventing him from going to a date. His girlfriend, Nicole, unaware of the situation, sends him a message to ask what happened. At this moment, John must choose between ignoring the message, which may upset Nicole [2], or pausing the current conversation to reply to the message, which could impair the face-to-face interaction [19, 40, 56, 78]. Although less desirable, such scenarios are quite common in everyday life as we need to handle multiple requests during social interactions. In such situations, it may be desirable to minimize the interruption of these secondary tasks to the primary social interaction, which leads to the topic of this paper: how to support secondary human-computer interaction with minimal interference to ongoing primary social interactions.

To address this challenge, we propose an interaction technique called *ParaGlassMenu*, which is designed to support seamless subtle interactions [64] in social settings. *ParaGlassMenu* incorporates four important design requirements to support general-purpose subtle interactions in social settings.

- First, it minimizes visual distractions (being non-intrusive [64]) to users during social settings by leveraging the insights of attention-maintaining visualizations [36]. This allows users to focus on their conversational partner while interacting with the menu items displayed in the peripheral area of their vision on an Optical See-Through Head-Mounted Display (OHMD) [34].
- Second, the input mechanism, using a ring mouse, supports discreet manipulations (hiding activities [64]) cross-scenario [71] to minimize distracting others and protect privacy when necessary [54].
- Third, *ParaGlassMenu* supports both discrete and continuous manipulations to accommodate a wider range of interaction needs.
- Fourth, as a hierarchical menu, *ParaGlassMenu* is scalable and can accommodate a larger set of commands than many previously proposed subtle interaction techniques (e.g., Jaw-Teeth interaction [8], etc.).

Note that while many other subtle interaction techniques [5, 8, 26, 54, 68] have been proposed, they are missing some of the above design requirements and have not been tested with a solution that combines all of the above features together in social settings (see Related Work, sec 2.2 for details).

When considering the application scenario to evaluate the effectiveness of *ParaGlassMenu*, we chose Internet-of-Things (IoT) control (e.g., Figure 1c) as it allows us to evaluate *ParaGlassMenu* with a wider range of interaction types [28, 72], including checking information, discrete and continuous manipulation, and searching. This enables us to better generalize the results of our evaluation to applications with similar interactions (e.g., the example of handling remote social inquiries, **Figure 11a**). In addition, the use of IoT control applications in social settings enables us to study the manipulation of digital tasks in a social context, where the manipulation can be either personal or for all involved parties [47], and its visibility can be either opaque (discreet and invisible to others) or transparent (noticeable by others) [47, 54]. This provides insights into how users use *ParaGlassMenu* to manage digital tasks with various purposes in conversations.

We first compare *ParaGlassMenu* with a Phone Interface, a Voice-User Interface (VUI), and an OHMD linear menu, under a simulated conversation setting in the laboratory. Results showed that *ParaGlassMenu* was the most preferred interface with the best interaction performance, and the lowest cognitive load and disengagement. To ecologically validate these findings, we further evaluated *ParaGlassMenu* in a more realistic scenario where users interacted with commodity IoT devices while conversing with a friend in a modeled home. The results confirmed the utility of *ParaGlassMenu* in a more realistic scenario while revealing additional insights about users' manipulation behaviors in real-world contexts. Finally, we discuss implications on how to design attention-maintaining subtle interaction techniques with general purpose on OHMDs.

The contributions of this paper thus are threefold: 1) the design of a novel attention-maintaining subtle menu technique called *ParaGlassMenu* for the emergent OHMD platforms that incorporates four essential design requirements, 2) the empirical validation of the effectiveness of *ParaGlassMenu* compared to other commonly used approaches in social settings, and 3) insights and design recommendations for creating more effective attention-maintaining subtle interaction design.

2 RELATED WORK

Our work is related to the following four areas.

2.1 Digital interactions in social settings

While much work was conducted in this area, we mainly focus on two aspects of it, i.e., categorization and evaluation.

Digital interactions can be categorized in many ways (e.g., interaction modalities), and in social settings, a way to classify them based on their relevance to the people involved in social engagement. From one user's point of view, the digital interactions can either be 1) related to other parties or the common conversation topics, or 2) personal or non-conversation-related [47]. For example, from a host's perspective in a gathering, playing music for everyone can be classified as the former type of digital interaction. While checking email to see whether a message has arrived from her boss (who is not in the gathering) or checking the oven's status in the kitchen can be classified as the latter. Due to the different nature and requirements of digital interactions, their corresponding interaction technique may require a specific design. Literature has suggested they can be treated either transparently or in a hiding manner [47, 54]. Yet how to incorporate these principles of visibility selection into a specific application context is still the job of researchers/designers.

Given that digital interactions can bring potential negative effects on social activities [19, 40, 56, 78], their impacts on conversation quality need to be evaluated in order to determine their feasibility in conversations. Particularly, their abilities to maintain eye contact [6, 7, 31] and minimize impoliteness behaviours [40, 51] should be assessed as they affect users' active engagement in conversation. Besides, we need to evaluate the social acceptability of interactions from both the users' and the observers' perspectives [3, 40, 77], as previous studies have demonstrated positive [58, 75, 81] and negative [33, 41, 78] effects of digital interactions on all involved parties.

2.2 Subtle interactions in social settings

The concept of subtlety has been leveraged to minimize interference in social settings while interacting with digital devices [54, 64]. Based on Pohl's categorization [64], they can be categorized into four types, including 1) being non-intrusive to users' perception, 2) hiding activity from others, 3) doing less while interacting, and 4) nudging users. In conversation settings, being non-intrusive and hiding activities are the most relevant as they can retain eye contact and politeness for the following reasons. Firstly, non-intrusive interactions can reduce distraction [64], thus helping users maintain attention on others. Secondly, hiding activity in conversation can reduce disruption and gain social acceptance [20, 64], especially when users deal with conversation-irrelevant information.

To support non-intrusive interaction in social settings, several subtle interactions are introduced. For example, HiddenHaptics [54] allows users to receive information through vibro-tactile cues on a smartphone without directly looking at it. However, its vibration feedback only supports relatively simple information [54]. On the other hand, attention-maintaining interfaces [36, 69] utilize the eye's peripheral vision to deliver visual feedback on OHMD and help users maintain attention on the central target [36]. However, such interfaces mainly focus on providing system feedback (output, i.e., uni-directional interaction) to users without allowing user input.

Similarly, many subtle interaction mechanisms have been proposed to hide activities while providing inputs in social settings. These include hand gestural interactions [32], foot plantar-based interactions [26], silent-speech interactions [46], and gaze interactions [68]. Additionally, embedding interactions into common objects allow users to hide their activities, which include mug interaction [12], watch interaction [54], and book interactions [5]. While all of them are useful for different scenarios with different capabilities, they cannot completely replace the usability (e.g., accuracy) of touch mechanisms or mechanical sensors [32, 46]. Besides, only supporting a limited number of discrete commands (e.g., jaw-teeth interaction [8]) makes it less suitable as a general and scalable technique. In addition, some of the above-mentioned subtle techniques are less suitable for conversations (e.g., gaze interactions make it hard to maintain eye contact with conversation partners [68]; jaw-teeth interaction [8] and silent-speech interaction [46] are hard to interact with while speaking).

In contrast, thumb-index interactions using a ring mouse [67, 71], supported by mechanical sensors or touch interactions, strike a balance among usability, convenience, and social acceptability [3]. Their tiny shape makes them easier to carry than other hand-held devices, and the one-hand manipulation nature is more flexible than interactions involving two hands (e.g., watch interaction [54]).

2.3 IoT control interfaces for social settings

We applied the *ParaGlassMenu* for IoT device manipulation during conversations to evaluate how *ParaGlassMenu* supports varying complexities and purposes of digital interactions in realistic social settings. Thus, we reviewed existing IoT control interfaces.

Today, users can easily manipulate IoT devices using touch screens on smartphones or dedicated wall-mounted displays [50]. In addition, voice assistants, like Alexa, help users control smart devices [48, 65] in an eyes-free and hands-free manner [66]. Recent research on gesture control interfaces, such as SeleCon [4] and Physical Loci [63], allow users to make intuitive hand gestures to manipulate IoT devices. Social robot [50] also enables users to manipulate IoT devices with tangible icons and expressive gestures, which provides high situational awareness.

Furthermore, the emerging OHMDs allow users to interact with Augmented Reality (AR) menus to control the IoT devices, and several input mechanisms are used alone or together, including gaze, voice, and mid-air gestures [42, 74]. The leverage of OHMDs supports quickly acquiring information in a non-intrusive manner [36, 49] and their different input modalities provide flexible options across different scenarios.

2.3.1 Categorization of existing interfaces. To evaluate the aforementioned interfaces in social settings, Table 1 categorizes them in terms of two dimensions, i.e., attention maintenance (non-intrusive) and manipulation visibility (hiding activities) based on previous frameworks on subtle interactions [5, 64]. We found that because they were designed for different scenarios, supporting eye contact or opaque interaction (i.e., discreet and invisible to others) was not integrated into these interfaces (e.g., gaze interaction can cause attention shift, and voice interaction can make the manipulation transparent). Thus, there is a need for new interfaces to fill this gap. Based on the analysis of subtle interaction in social settings (sec 2.2), we chose a ring mouse supporting thumb-index interaction [67, 71] as the discreet input mechanism. In addition, an OHMD menu placed in the peripheral vision area, thus supporting non-intrusive AR interface [21, 36, 52, 60] was chosen as the output mechanism.

Weigel et al. [80] introduced a flexible input device that can be deformed into various shapes, including a ring. They demonstrated an example of a pie menu on Google Glass and Oculus Rift as one instance of their design space. However, their focus was on the flexible input mechanism rather than interactions in social settings, so they did not provide menu design guidelines or further evaluate their design in social settings. *ParaGlassMenu* fills this gap by introducing a concrete design that satisfies the four requirements mentioned in the introduction and provides empirical validation. In

Table 1: IoT control interfaces in conversations can be evaluated in terms of two dimensions, i.e., attention maintenance and manipulation visibility. Attention maintenance has two levels: enable eye contact or not. Manipulation visibility has two levels: transparent or opaque, depending on whether the manipulation is visible. In addition, transparent manipulation has two sub-levels, i.e., fully transparent and semi-transparent, depending on whether the manipulated digital content is known to others. Besides, ⋆ indicates the interface can support more transparent manipulation naturally by verbally expressing it to conversation partners.

	Transparent Manipulation		Opaque Manipulation
	Fully transparent	Semi-transparent	
Enable Eye Contact	*Voice*, *ParaGlassMenu*⋆	*ParaGlassMenu*⋆	*ParaGlassMenu*
Not Enable Eye Contact	*Phone*⋆, *Wall-Mounted Display*⋆, *Social Robot*, *Gaze*⋆, *Mid-Air Gesture*⋆	*Phone*, *Wall-Mounted Display*, *Gaze*, *Mid-Air Gesture*	

particular, the requirement of displaying the menu non-intrusively around the face of a conversation partner is not mentioned by Weigel et al., but we believe it is a key insight that contributes to the effectiveness of *ParaGlassMenu* in supporting seamless digital interactions during social settings.

2.4 Visual Menu design

Menu is a common interaction technique in modern GUI to explore and execute commands. Despite extensive studies (see [10] for a comprehensive review) investigating how menus should be designed to improve productivity in various contexts [10], less is known about their suitability for social contexts. While some menus designed for multitasking usage (e.g., [83]) could potentially be used in social settings, they have not been evaluated under such contexts. Hence, we look into more social-friendly menu designs that allow interactions without affecting social engagement.

To support better social engagement, one approach is to maximize eye contact during a social engagement. Previous studies verified the advantages of presenting the information in the eye's peripheral region to help users keep attention on conversation partners [36, 68].

In seeking a suitable design, we investigated the various layouts presented in literature [10, 79] and identified four possible categories, including: vertical, horizontal, circular, and rectangular. Among them, a vertically arranged, left-aligned linear menu was the most preferred layout for presenting menus [79]. Nonetheless, the circular presentation has been proven to facilitate eye contact in a realistic conversation setting [36], while a linear layout has been found to encourage attention switching between the side visualization and the conversation partner [36]. The difference in previous studies motivates us to further investigate the most suitable layout for non-intrusive bi-directional (considering both input and output) interaction in social settings.

3 *PARAGLASSMENU*

The design inspiration for *ParaGlassMenu* comes from several branches of prior works. One branch of work is extensive studies on radial style menus, including the Pie Menu [16], Marking Menu [44], Wavelet Menu [25], Flower Menu [9], etc. These menus take advantage of the radial layout, and create menus that are compact and efficient. The second branch of inspiration comes from the recent investigation on OHMD interface design. Given the unique features of OHMD, such as a transparent display that can overlay virtual content on realistic objects, we need to customize the menu designs on OHMD. One particular piece of work that inspired this design is Janaka et al.'s work on the attention-maintaining interface [36], where a circular progress bar is displayed in the para-central and near peripheral vision on the OHMD to enable users to receive digital information while engaged in a social conversation.

We extend the idea of para-central and near peripheral visualization, which only supports uni-directional notification, into an attention-maintaining interaction technique that supports bi-directional information exchange (involving both input and output).

3.1 *ParaGlassMenu* Design

Figure 2 shows the overall visualization of *ParaGlassMenu*. The basic design of the menu follows a hierarchical radial menu with all menu items layout radially around the center, which can facilitate visual search for hierarchical menus [70]. However, it has a number of features that are different from a traditional radial menu.

Position and layout: We designed non-intrusive menus for *ParaGlassMenu* based on the guideline of attention-maintaining visualization proposed by Janaka et al. [36]. Its menu items circularly surround the target (i.e., the conversation partner's face, Figure 1c) and enable users to check the menu using peripheral vision (angle = 13.7° when focusing on the face center) to minimize the attention switching and any occlusion to the conversation.

Item presentation: The menu items include both icons and text, as icons are easy to recognize, and text provides precise information [76].

Color and transparency: The menu items are rendered in green color following recommendations from prior studies [18, 22] to ensure easy recognition in OHMD. For each menu item, icons are rendered in a semi-transparent fashion to minimize occlusion. For the same reason, items are presented without any connection lines.

Supporting continuous feedback: Circular progress bars (e.g., Figure 2, AC) are adopted as indicators for continuous manipulation (e.g., increasing temperature) and selecting from list, as the circular presentation in peripheral vision could provide non-intrusive feedback at a glance [36].

Ring mouse interaction:

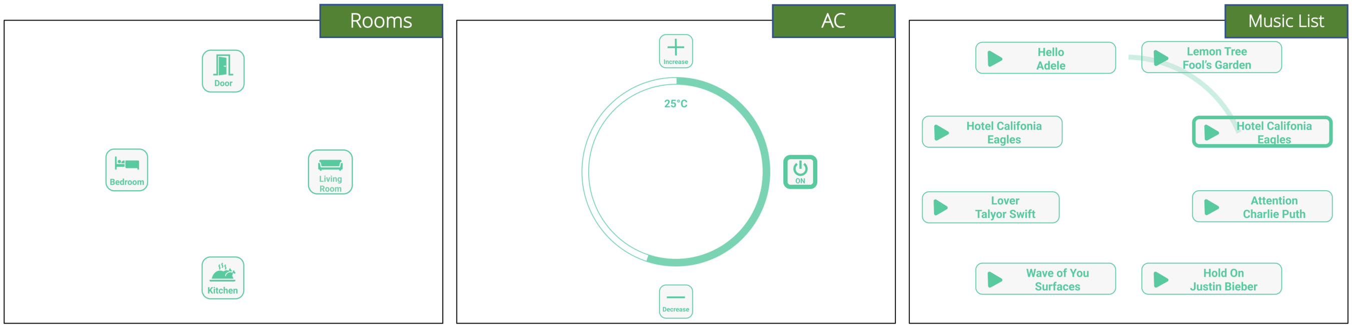


Figure 2: Three instances of menu layouts in *ParaGlassMenu* demonstrating three usage scenarios. The first panel (rooms) consists of 4 icons corresponding to 4 clickable buttons on the ring mouse. The second panel (AC menu) consists of icons and a circular progress bar. As an indicator of continuous manipulation, the circular progress can be controlled by scrolling the ring mouse's trackpad. The third panel shows a (music) list, and the circular bar indicates that users can also scroll the trackpad to select items.

1) Activation: The menu is inactive and not visible on the OHMD by default to minimize unwanted attention during conversations. Hence, users can activate the menus by clicking any of the four buttons on the ring mouse, and consequently, menu items will be shown.

2) Manipulation: The ring mouse supported both clicking and scrolling interaction and followed natural spatial mapping guidelines proposed by Norman [59] to reduce cognitive load. Specifically, as shown in Figure 1b and Figure 1c, users can click the respective button on the ring mouse for item selection, and the selected item will be highlighted with bold-ed icons. Besides, users can return to the previous menu by clicking the left button or swiping from right to left. In addition, continuous manipulation and selecting from list are supported by scrolling circularly on the trackpad.

Additionally, to make the menu compatible with the ring mouse, the maximal number of items in each menu level was set to match the input mechanism of the ring mouse (four items in our prototype, as the ring mouse has four clickable buttons, see Figure 1c).

4 STUDY OVERVIEW

Based on the analysis of prior research, the promising features of OHMD and ring mouse motivated us to design *ParaGlassMenu* to support bi-directional interaction, specifically for IoT manipulation during social settings. In addition, to evaluate the performance and social acceptability of *ParaGlassMenu*, two studies were conducted answering the following research questions with the approval from our university's institutional review board (IRB).

- *RQ*1: How does the *ParaGlassMenu* compare with other interfaces in terms of the quality of conversation and IoT manipulation?
- *RQ*2: How does the *ParaGlassMenu* support IoT manipulation in real social settings?

5 *STUDY 1*: COMPARISON BETWEEN THE *PARAGLASSMENU* AND ALTERNATIVE INTERFACES

To answer *RQ*1, the *ParaGlassMenu* was compared against the smartphone touch-screen interface, voice user interface, and Linear OHMD menu interface. The comparisons were conducted in a simulated conversation setting to eliminate confounding factors in realistic settings and provide a consistent experience.

5.1 Participants

Twenty volunteers (12 females, 8 males, mean age = 22.10 years, SD = 1.65 years) from the university community participated in this study. None of these participants wore spectacles or had any vision deficiencies. Eleven participants reported that they used IoT devices (i.e., smart speakers (9), smart lamps (2)). Each participant was compensated ≈ USD 7.25/h for their time.

5.2 Comparative Interfaces

Three comparative interfaces were selected in this study for the following reasons. Firstly, current IoT manipulation interfaces have two main limitations in terms of 1) maintaining eye contact and 2) conducting opaque (discreet and invisible) IoT manipulation in conversation (see Table 1). Thus, to systematically evaluate *ParaGlassMenu*, it should be compared against interfaces with the two above limitations. In addition, prior studies (e.g.,[5]) compared the subtle interfaces with traditional interfaces in conversation settings to identify the relative advantages and disadvantages. Thus, two traditional interfaces, including phone (hard to maintain eye contact) and VUI (hard to support opaque manipulation), were selected as comparative baselines in this study. Secondly, because of previous results' inconsistency on the preferred interface layout [36, 79], *ParaGlassMenu* was compared with a linear layout menu to further explore the most suitable layout for attention-maintaining bi-directional interactions in conversations. The details of these three comparative interfaces are as follows.

5.2.1 Smartphone touch-screen interface (Phone Interface).

Design: Android Google Home app[1] was used to manipulate IoT devices as it is commonly used in smart homes [27].

Interaction Methods: Users can press the power button to unlock the phone. Then, users can open and manipulate the mobile app (see Figure 3a) by tapping on the app's icons on the phone or scrolling the circular slider on the screen using touch interactions.

5.2.2 Voice user interface (Voice Interface).

Design: Google Home platform, specifically Google Nest Hub 2[2], was used for interacting with IoT devices via voice [66]. Since our focus is voice-based interaction, its visual display was completely covered to support voice-only interactions.

Interaction Methods: Users can speak out "Hey Google" to activate the voice assistant and then speak out the device name, desired function, and the device's location to manipulate an IoT device. For example, users can say, "Hey Google, turn on the light in the living room" to manipulate (turn on) the selected device.

5.2.3 Linear Menu on OHMD with ring interactions (Linear Interface).

Design: This interface consists of a vertical linear menu aligned to the left side of the conversation partner's face, following Vatavu et al. [79]. All the other aspects of this menu were identical to the *ParaGlassMenu*.

Interaction Methods: Similar to the *ParaGlassMenu*, menu items are shown on the OHMD upon activation. However, given its vertical layout, users only use the top and bottom buttons for navigating menus (Figure 3b). Additionally, clicking the right button enables users to select the chosen menu icon, while the left button enables them to return to the previous menu. Lastly, users can do continuous manipulations by scrolling vertically on the central trackpad.

5.3 Apparatus

Figure 4 shows the overall simulated conversation setting of the experiment. A virtual conversation partner (a muted talking head video following [36]) and two virtual rooms (a living room and a kitchen) were displayed on three 27" LCD monitors (refresh rate = 60 Hz, resolution = 1920 x 1080 px) at eye level. The former was modeled after an average female (head height = 24 cm [61], FoV = 9.15° vertical at 1.5 m) and was displayed on the central monitor 1.5 meters away from the participants following the social conversation distance defined by Hall et al.[29, 36]; while the latter were displayed on side monitors to provide an immersive feeling of a home at the same distance. A Python program controlled the virtual conversation partner and other stimuli on desktops. Note that the virtual conversation partner was used with a trade-off consideration between external validity and internal validity [53]. While using realistic conversation partners can enhance external validity, it can significantly reduce internal validity by introducing potential confounding factors, such as inconsistent replies in terms of content and duration, which can affect the users' manipulation behaviors. Thus, we selected a virtual conversation partner to make a fair comparison in this study.

There were a total of eight IoT devices, four in the living room (two lights, an air-conditioner, and a smart speaker) and four in the kitchen (a light, a dishwasher, and two drink machines), following common smart home settings [23, 35, 37]. To manipulate the IoT devices, as shown in Figure 4, participants used either an OHMD (Nreal Light[3], 1920x1080, 60 Hz, FoV $\approx$ 45° horizontal $\times$ 25° vertical) with a ring mouse (Sanwa 400-MAW151BK with 4 buttons and 1 touchpad), a smartphone (Google Pixel 4, 5.7"), or a smart speaker (Google Nest Hub 2) depending on the conditions. A mobile eye tracker (Pupil Core/Pupil Core Addon) was used either attached to OHMD or directly worn on the head. Four April Tags were attached to the central monitor for the eye tracker to register the location of the virtual conversation partner.

For *ParaGlassMenu* and *Linear Interface*, participants wore the Nreal Light along with the ring mouse on their dominant hand. Menus were implemented using Unity[4] and displayed at the same depth as the virtual partner using the mixed reality mode of Nreal. OpenCV plus Unity asset [5] was used to track the target face with the Nreal's camera, which positioned the menu around the face. The size of the menu icons, 5 cm in diameter, was designed based on a pilot study (N=5) where participants could recognize the menu while looking at the virtual face from a 1.5 m distance (Figure 1c and Figure 3b).

For the *Phone Interface*, participants used a Google Pixel 4 phone installed with the Google Home app and YouTube Music app[6]. YouTube Music app was used to select and stream songs to the smart speaker, as Google Home app doesn't allow playing songs directly in the app. The locked phone was placed on the table within hand reach. Moreover, for the *Voice Interface*, participants used integrated Google Voice Assistant in the Google Nest Hub 2. In addition, Google Home Playground[7] was used to generate virtual IoT devices and rooms for Google Home App (*Phone Interface*) and Google Nest Hub 2 (*Voice Interface*).

5.4 IoT Manipulation Tasks Design

IoT manipulation tasks can be divided into two types: 1) *information* task in which users get information about a device and 2) *command* task in which users execute a command on a device [72]. Moreover, our analysis of the IoT tasks in smart home scenarios based on the Google Home Device traits [28] revealed six major sub-tasks related to IoT manipulations: 1) Activation: turning on the manipulation interface; 2) Navigation: going to the corresponding room/device; 3) Selection: selecting the room/device/item; 4) Checking: examining the state of the device; 5) Discrete manipulation: changing the discrete state of the device; and 6) Continuous manipulation: changing the continuous state of the device.

By aligning the two IoT manipulation task types with six subtasks, we found activation, navigation, and selection were common to both types. Besides, an *information* task involves checking; while a *command* task, depending on the capabilities of the device, includes discrete or continuous manipulation. Moreover, the task

[1]https://play.google.com/store/apps/details?id=com.google.android.apps.chromecast.app
[2]https://store.google.com/product/nest_hub_2nd_gen
[3]https://www.nreal.ai/light
[4]https://unity.com
[5]https://assetstore.unity.com/packages/tools/integration/opencv-plus-unity-85928
[6]https://play.google.com/store/apps/details?id=com.google.android.apps.youtube.music
[7]https://developers.google.com/assistant/smarthome/tools/home-playground

(a) **(b)**

Figure 3: Two Comparative interfaces. (a) An instance of menu layout on the phone. The path shows how to control AC. (b) OHMD Linear Menu. The yellow arrows stand for using the *top* button and *bottom* buttons for menu navigation.

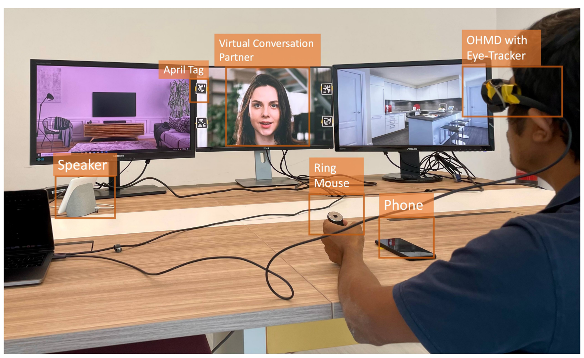


Figure 4: Overall setup of *study 1*. The participant sits 1.5 m away from the virtual conversation partner. Depending on the condition, the participant either wears an OHMD with a ring mouse or only uses the phone or the speaker (VUI) to manipulate virtual IoT devices.

complexity, i.e., the number of steps or the duration required to complete a task, depends on the number of states supported by the device.

Thus, to evaluate the interfaces across different tasks with different complexities, four IoT manipulation tasks (i.e., *IoT Task*s), including one *information* task and three *command* tasks, were selected to cover the full spectrum of sub-tasks. In that regard, *Checking Info* (i.e., check the device's current state) was selected as the information task; while *Discrete Manipulation* (i.e., change the active state of the device), *Continuous Manipulation* (i.e., change the continuous state of the device), and *Selecting From List* (i.e., change the discrete state of the device which has more than two states) were selected as command tasks. Appendix A.1 presents sample IoT tasks and Table 2 summarizes the interaction methods used for the selected *IoT Task* on the selected *Interface*.

5.5 Study Design

A repeated-measures within-subject design was used in which the independent variables were IoT *Interface* (*ParaGlassMenu*, *Linear*, *Phone*, *Voice*) and *IoT Task* (*Checking Info*, *Discrete Manipulation*, *Continuous Manipulation*, *Selecting From List*), resulting in 16 sessions per participant. Furthermore, the IoT *Interface* was counterbalanced using Latin Square across participants, and the *IoT Task*s were presented in a fixed order with increasing complexity, i.e., *Checking Info* followed by *Discrete Manipulation*, followed by *Continuous Manipulation* followed by *Selecting From List* because comparing conversation and IoT manipulation quality across different task types was not in the scope of this research.

To avoid the potential biases due to the menu layouts, three trials for each *IoT Task* were designed, and each trial involved different devices with the same complexity. In summary, the final design involved 960 IoT trials in total, including: 20 participants $\times$ 4 *Interfaces* $\times$ 4 *IoT Tasks* $\times$ 3 trials per task.

5.6 Task and Procedure

After getting consent, participants were first given brief guidance and training sessions to familiarize themselves with each *Interface*; then completed the 16 sessions in the formal experiment.

For each session, the eye-tracker was first calibrated, then three trials were conducted. For each trial, the manipulation commands were first displayed in text form consisting of action, device name, and location (e.g., "Raise the Temperature of the AC Above 27 in the Living Room", see Appendix A.1) on the central monitor for seven seconds to ensure participants can read the commands at least twice [15]. Next, the text "Start" was shown on the monitor for one second to tell participants they could start manipulating the device when the virtual conversation partner showed up; then, the virtual conversation partner was displayed on the central monitor and continuously speaking (moving mouth) until the participant **successfully** completed each trial (see the details of stimuli in Appendix A.2). We asked the participants to act as if they are listening to their conversation partner when manipulating.

Table 2: Interaction methods of *IoT Task* on different *Interfaces*

Interface	*Checking Info*	*Discrete Manipulation*	*Continuous Manipulation*	*Selecting From List*
ParaGlassMenu	Clicking buttons	Clicking buttons	Clicking buttons + Scrolling	Clicking buttons + Scrolling
Linear	Clicking buttons	Clicking buttons	Clicking buttons + Scrolling	Clicking buttons + Scrolling
Voice	Speaking	Speaking	Speaking	Speaking
Phone	Tapping	Tapping	Tapping + Scrolling	Tapping + Scrolling

To ensure consistent experience among all participants, the state of the IoT devices and the status of the *Interface* were reset to the default after each trial. After finishing all three trials for each session, participants filled out questionnaires, detailed in sec 5.7, about their experience with the corresponding *Interface* and *IoT Task* pair.

Moreover, participants were given a 10-minute break upon completing all four sessions for each *Interface*. After completing all sixteen sessions, they filled out a questionnaire with their overall rankings and attended 8-12 minutes of semi-structured post-interview. The entire experiment took approximately 120 minutes per participant.

5.7 Measures

Following our RQs, the quality of simulated conversation and IoT manipulation were evaluated using objective and subjective measures. Additionally, preference rankings for all interfaces and their reasons were collected.

5.7.1 Quality of (simulated) conversation. The *Face Focus* (i.e., the percentage of time the user's gaze within the bounding box between eyebrow and mouth, Figure 5) was used as the objective measure following Janaka et al. [36].

In addition, subjective measures were also collected, such as *Politeness* ('I felt it is polite to use the system during the conversation') and *Naturalness* ('I acted naturally at all times while focusing on the face and manipulating IoT devices') by adapting from previous studies on social setting [40, 57] using a 7-point (1 = Strongly Disagree, 7 = Strongly Agree) Likert scales. Lastly, the perceived task load for maintaining focus on conversation and manipulating IoT devices was collected using Raw NASA-TLX (*RTLX* [30]).

Figure 5: The *Face Focus* area is enclosed in green rectangle. The size of *Face Focus* area is 11cm x 11cm, following the female biocular breadth [61].

5.7.2 Quality of IoT manipulation. The *Task Duration* (i.e., the average time to complete the given IoT task in seconds from starting of *IoT Task* till completion feedback) and *Task Accuracy* (i.e., the ratio of the number of successful manipulation attempts relative to the total number of manipulation attempts) were utilized as objective measures. Furthermore, *Relaxation* ('I felt relaxed while manipulating IoT devices', using 7-point Likert scales [57]), and system usability score (*SUS* [14]) were collected as subjective measures.

5.7.3 Analysis. Factorial repeated measures ANOVAs or factorial repeated measures ANOVAs after Aligned Rank Transform (ART [82]), in cases of violation in ANOVA assumptions, were applied; and the normality and sphericity were tested using the Shapiro-Wilk test and the Mauchly test, respectively. Moreover, paired-sample t-test or Wilcoxon signed-rank test were used as post-hoc tests, and Bonferroni correction was applied to p-values for multiple comparisons. The interview recordings were transcribed and thematically analyzed following Braun and Clarke [13].

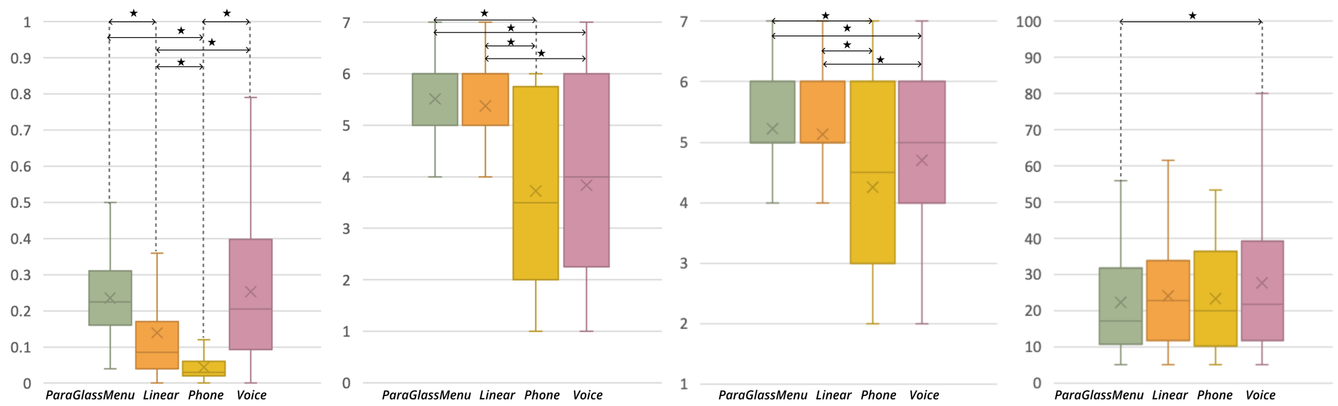
5.8 Results

During the study, a total of 320 data points were collected. Figure 6 and Figure 7 indicate the participants' performance (see Appendix A.3 for details).

5.8.1 Quality of (simulated) conversation. Overall, there was a significant ($p < 0.05$) main effect of the type of *Interfaces* for all measures, and the *ParaGlassMenu* allowed the highest quality of conversation when compared to other interfaces.

Face Focus: A repeated-measures ANOVA after ART indicated significant main effects of *Interface* ($F_{3,285} = 85.155$, $p < 0.001$), *IoT Task* ($F_{3,285} = 9.394$, $p < 0.001$), and interaction effect ($F_{9,285} = 2.583$, $p = 0.007$). Besides, there were simple effects ($p < 0.05$) for individual levels of *Interface* and *IoT Task* except for *Phone Interface*. Moreover, post-hoc analysis revealed *Voice* and *ParaGlassMenu* were significantly higher than *Linear* and *Phone* ($p_{bonf} < 0.05$), with *Linear* significantly higher than *Phone* ($p_{bonf} < 0.05$). There was no significant difference between *ParaGlassMenu* and *Voice*.

Overall, *Voice* enabled the highest *Face Focus* ($M = 0.253$, $SD = 0.192$) on the virtual conversation partner as it did not provide any visual feedback that deviated their visual focus from the conversation partner's face; however, six participants who disagreed with the above mentioned that they could focus better with *ParaGlassMenu* ($M = 0.235$, $SD = 0.119$) over *Voice* as they tended to look at the smart speaker before speaking; while the circular layout of *ParaGlassMenu* helped them concentrate on the face. In contrast, *Phone* had the lowest *Face Focus* ($M = 0.044$, $SD = 0.043$) as IoT manipulation using *Phone* required users to switch between the phone and the face.



(a) ***Face Focus*** **(0-1)** (b) ***Politeness*** **(1-7 Likert)** (c) ***Naturalness*** **(1-7 Likert)** (d) ***RTLX*** **(1-100)**

Figure 6: Quality of simulated conversation. Error bars represent standard error. The middle lines of box plots represent median values. ⋆ represents significant ($p < 0.05$) post-hoc tests and × inside box plot represents the mean value point. See Appendix A.3 for details.

Politeness: There was only a significant main effect of *Interface* ($F_{3,285} = 50.731$, $p < 0.001$) and the post-hoc analysis revealed that *ParaGlassMenu* and *Linear* were significantly higher ($p_{bonf} <$ 0.001) than *Phone* and *Voice*, with no significant difference between other *Interface* pairs.

Overall, OHMD interfaces, particularly *ParaGlassMenu* showed the highest *Politeness* ($M = 5.51$, $SD = 1.12$) as it enabled participants to keep focus on the face. In contrast, participants felt it was "rude" and "impolite" to use the *Phone* ($M = 3.73$, $SD = 1.84$) to manipulate devices during a conversation as it required attention switching between the face and the phone and violated social norms. Similarly, participants felt using *Voice* ($M = 3.84$, $SD = 1.75$) was impolite and "awkward" as it could interrupt and pause the conversation; however, two participants mentioned that using *Voice* was acceptable to play songs when the conversation topics were related to songs as it could increase shared interactions.

Naturalness: There was only a significant main effect of *Interface* ($F_{3,285} = 12.800$, $p < 0.001$) and the post-hoc analysis revealed *ParaGlassMenu* and *Linear Interface*s were significantly higher ($p_{bonf} < 0.02$) than *Phone* and *Voice*, with no significant difference between other *Interface* pairs.

Overall, *ParaGlassMenu* showed the highest *Naturalness* ($M = 5.23$, $SD = 1.04$) indicating that it allowed the manipulation of IoT devices with lesser interruption, according to post-interviews.

RTLX: There were only significant main effects of *Interface* ($F_{3,285} = 4.234$, $p = 0.006$) and *IoT Task* ($F_{3,285} = 4.040$, $p = 0.008$). Moreover, the post-hoc analysis revealed *ParaGlassMenu* was significantly lower ($p_{bonf} = 0.004$) than *Voice*, with no significant difference between other *Interface* pairs.

Overall, the *ParaGlassMenu* had the lowest *RTLX* ($M = 22.23$, $SD = 14.34$) as it enabled easier IoT devices multi-tasking while focusing on the face. Additionally, *ParaGlassMenu*, *Linear*, and *Phone* provided visual cues, which reduced the burden of remembering the commands or making mistakes compared to *Voice*. In contrast, *Voice* caused the highest *RTLX* ($M = 27.67$, $SD = 19.99$) due to command recognition errors which made participants repeat voice commands. Moreover, as expected, it made users "wait" for confirmation feedback which took longer time-demand than other *Interface*s.

5.8.2 *Quality of IoT manipulation.*

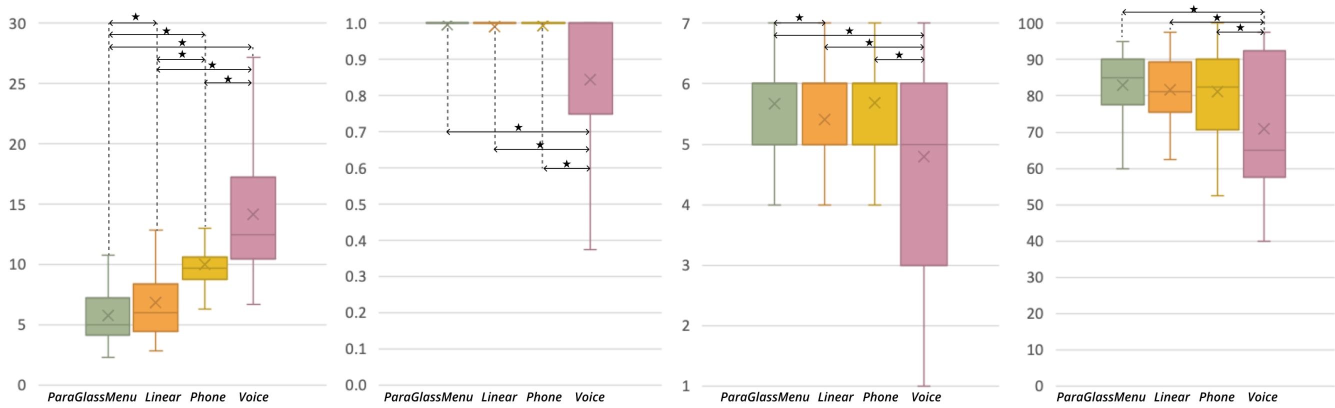
Overall, there was a significant ($p < 0.05$) main effect of *Interface* for all measures, and the *ParaGlassMenu* increased the quality of IoT manipulation over others.

Task Duration: There were significant main effects of *Interface* ($F_{3,285} = 321.711$, $p < 0.001$), *IoT Task* ($F_{3,285} = 58.370$, $p < 0.001$), and interaction effect ($F_{9,285} = 15.496$, $p < 0.001$). Besides, there were simple effects ($p < 0.05$) for all individual levels of *Interface* and *IoT Task*. The post-hoc analysis revealed significant differences ($p_{bonf} < 0.001$) between all *Interface* pairs with the *ParaGlassMenu* having the lowest duration and *Voice* having the highest.

Overall, *ParaGlassMenu* had the lowest *Task Duration* ($M = 5.75$, $SD = 2.28$) as it enabled to locate and navigate individual devices easily while maintaining focus on the face, provided "more intuitive" manipulation compared to *Linear*, and reduced attention switching between the face and the menu compared to *Phone*. On the contrary, as expected, *Voice* had the highest *Task Duration* ($M = 14.18$, $SD = 5.60$) due to the longer time to provide voice commands and get feedback, and multiple attempts due to voice recognition errors.

Task Accuracy: There were significant main effects of *Interface* ($F_{3,285} = 64.194$, $p < 0.001$), *IoT Task* ($F_{3,285} = 100.873$, $p < 0.001$), and interaction effect ($F_{9,285} = 20.279$, $p < 0.001$). Besides, there were simple effects ($p < 0.05$) for *Voice* and *IoT Task*s except for *Discrete Manipulation*. The post-hoc analysis revealed *ParaGlassMenu*, *Linear*, and *Phone* were significantly higher ($p_{bonf} < 0.001$) than *Voice*, with no significant difference between other *Interface* pairs.

Overall, *Voice* had the lowest accuracy ($M = 0.844$, $SD = 0.183$) due to the speech recognition inaccuracy, which led to repeated commands. On the contrary, *ParaGlassMenu* has the highest accuracy ($M = 0.997$, $SD = 0.028$) due to its intuitive spatial mapping, and *Phone* has the second highest accuracy ($M = 0.994$, $SD = 0.039$) due to its familiar UI designs with touch interaction.



(a) ***Task Duration*** **(seconds)** (b) ***Task Accuracy*** **(0-1)** (c) ***Relaxation*** **(1-7 Likert)** (d) ***SUS*** **(1-100)**

Figure 7: Quality of IoT manipulation. Error bars represent standard error. The middle lines in the box plot represent median values. ⋆ represents significant ($p < 0.05$) post-hoc tests and × inside box plot represents the mean value point. See Appendix A.3 for details.

Relaxation: There was only a significant main effect of *Interface* ($F_{3,285} = 12.523$, $p < 0.001$) and the post-hoc analysis revealed *Voice* was significantly lower ($p_{bonf} < 0.05$) than other *Interface*s, and *Linear* was significantly lower ($p_{bonf} < 0.05$) than *ParaGlassMenu*. There was no significant difference between other *Interface* pairs.

As expected, *Phone* had the highest *Relaxation* ($M = 5.69$, $SD = 0.91$) due to device familiarity. *ParaGlassMenu* has the second highest *Relaxation* ($M = 5.68$, $SD = 1.13$) due to its quick and intuitive manipulation. While *Voice* was felt the least relaxed ($M = 4.80$, $SD = 1.69$) as incorrect recognition of voice commands caused repeated attempts and delays in feedback.

SUS: A one-way repeated-measures ANOVA with Greenhouse-Geisser correction ($\epsilon = 0.71$) revealed a significant effect of *Interface* ($F_{2.155,40.952} = 5.288$, $p = 0.008$, $\eta^2 = 0.218$; Note: *SUS* was calculated only for each *Interface*). The post-hoc analysis revealed that *Voice* was significantly lower ($p_{bonf} < 0.05$) than *ParaGlassMenu*, *Linear*, and *Phone*, with no significant difference between the pairs of the three *Interface*s.

Overall, *ParaGlassMenu* was perceived as the most usable system ($M = 83.00$, $SD = 9.82$) to manipulate IoT devices in a conversation setting as it was "intuitive", "easy to use", "polite", "faster than others", and "help[ed] to concentrate on people's face". In contrast, *Voice* had the lowest *SUS* score ($M = 70.88$, $SD = 18.84$), which was below the threshold (i.e., 80 [11]) for good usability.

5.8.3 Preference rankings. Figure 8 indicates the overall preference ranking of *Interface*s.

The majority of participants (12) ranked *ParaGlassMenu* as their most preferred *Interface*, while *Voice* is the least preferred one (11). They reported that *ParaGlassMenu* was intuitive, easy to use, polite, and less distracting to the conversation than the other *Interface*s, while *Voice* could interrupt conversations as voice commands could pause the conversation and speech recognition errors cause repeated attempts.

The participants (5) who selected the *Phone* as their first preference mentioned that familiarity helped them to control the IoT devices easily and conveniently, and it was acceptable as "most people have gotten used to people occasionally checking their phones". At the same time, two participants who chose *Voice* as their first preference mentioned that it took them less effort, did not affect their focus on the partner, felt more natural, and was easier to use when compared to ring mouse or phone. Lastly, the remaining participant who chose *Linear* as the first preference mentioned that the 1D nature of *Linear* was simpler and easier to locate than the 2D nature of *ParaGlassMenu*.

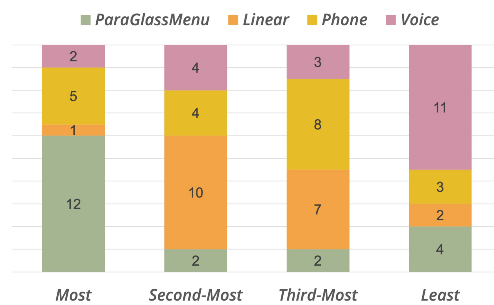


Figure 8: Preference ranking for different *Interfaces* (N=20).

5.9 Discussion

Overall, the *ParaGlassMenu* achieved the highest conversation quality in terms of more focus on the conversation partner ($M = 23.5\%$, $SD = 11.9\%$), the highest politeness ($M = 5.51$, $SD = 1.12$ / 7) and naturalness ($M = 5.23$, $SD = 1.04$ / 7), and the lowest cognitive load ($M = 22.23$, $SD = 14.34$ / 100). *ParaGlassMenu* also enables the most effective IoT manipulation measured with the lowest IoT manipulation time ($M = 5.75$, $SD = 2.28$ s), the highest accuracy ($M = 99.7\%$, $SD = 2.8\%$) and best usability score ($M = 83.00$, $SD = 9.82$ / 100) in a relaxed manner ($M = 5.68$, $SD = 1.13$ / 7). Thus, manipulation of IoT devices with

ParaGlassMenu demonstrated the lowest interference to the conversation. Furthermore, it was also the most preferred *Interface*. *Linear Interface* is recommended as the second choice due to its familiar linear layout, but interacting with it requires much higher attention, which causes a noticeably lower focus on the conversation partner.

On the other hand, as expected, *Phone* and *Voice Interface*s have limitations in a conversation setting, given the *Phone* failed to support high conversation quality and *Voice* failed to support both high conversation quality and high usability in social interactions. But this does not mean *Phone* and *Voice Interface*s should be excluded. *Phone Interface*s are the most accessible interface today, and it is the most familiar, making them the easiest and default choice for most users. *Voice Interface* has the ability to maintain visual attention on a target and be accessed ubiquitously, which can be particularly useful in other non-social settings, such as single-user scenarios and driving scenarios.

6 *STUDY 2*: VALIDATE THE *PARAGLASSMENU* IN A REALISTIC SETTING

The results of *study 1* indicated the suitability of *ParaGlassMenu* during conversation settings in terms of quality of conversation and IoT manipulation in a controlled simulated laboratory setting. To further verify the external validity of this finding and answer *RQ*2, we conducted a second study in a modeled realistic setting to evaluate the usage of *ParaGlassMenu*.

6.1 Participants

Twenty four participants (who did not participate in *study 1* or pilot studies) in twelve pairs (6 female pairs, 5 male pairs, 1 mixed gender pair, mean age = 23.00 years, SD = 2.22, $P1-P12_{Host/Guest}$) from the university community participated in this study. Following previous studies [57, 60], we chose the pairs of individuals who were familiar with each other to generate natural conversations. For each pair of participants, one was randomly chosen as the host while the other acted as the guest. Moreover, all participants self-reported to be fluent in English. They had normal or corrected to normal visual acuity without color deficiencies. None of them had prior experience using OHMDs, while ten of the participants had smart devices in their homes and self-reported to spend 1-10 minutes per day controlling IoT devices via phones or voice assistants. Each participant (host/guest) was compensated ≈ USD 7.25/h for their time.

6.2 Apparatus

Figure 1a shows the overall setting of the experiment. Similar to *study 1*, a local apartment was modeled in the lab with a living room and a kitchen. A sofa, a chair, and a table were placed in the living room, and a dining table was placed in the kitchen.

A total of four IoT devices were selected following past field studies on smart homes [23, 35, 37]. Three of them were in the living room: a fan (Xiaomi Mi Smart Standing Fan 2), a lamp (Philips Hues and Bridge), and a music player (Macbook Pro's speaker); and one was in the kitchen: a water heater (with Xiaomi Mi Smart Plug 2). To manipulate IoT devices, the host wore the OHMD (Nreal Light, with circular menu, enabling face-tracking) and the ring mouse (Sanwa 400-MAW151BK), which were identical to *study 1*.

6.3 Study Design

A repeated-measures within-subject design was used. Each pair of participants had two conversations, i.e., with or without IoT manipulation (*IoT*, *No_IoT*), and the order was counterbalanced. Among them, the *No_IoT* condition, as a baseline, measured the conversation's quality without the usage of *ParaGlassMenu* and minimized potential confounding factors due to current limitations with OHMDs, such as weight and appearance [38, 43]. In summary, the design was as follows: 12 participant pairs x 2 types of conversation = 24 conversation sessions in total.

6.4 Tasks

Participants had two tasks during the conversation. The primary task required hosts and guests to converse casually on any topic for around 15-20 minutes [55] in both *IoT* and *No_IoT* conditions.

No specific secondary tasks were required for participants in the *No_IoT* condition, but they could adjust the room conditions before starting the conversation; while in the *IoT* condition, the secondary task for hosts was to freely manipulate the IoT devices based on their preferences or suggestions from guests (who were aware that they could express their needs for IoT control to the hosts) if they had any, which included turning on/off the lights/fans, changing the speed of fan/brightness of light, playing/pausing music in the living room; remotely turn on the kettle, and preparing tea/milk/water in the kitchen.

6.5 Procedure

After getting consent, the hosts were given a briefing and training session to familiarize the *ParaGlassMenu* and the IoT devices in the modeled home, while the guests were briefed about the task in a separate room. When both participants were ready, the guest was guided into the modeled home and started the conversation at a distance of 1.5 meters from the host. The experimenters monitored (and video recorded) the conversation and IoT manipulations via teleconferencing in a separate room. To provide natural opportunities to control IoT devices, at the beginning of the study, the room temperature was set to 28°C, and lighting was set to 50 lux representing dark indoors [45].

After finishing each conversation session, the host and the guest filled out separate questionnaires (sec 6.6) about their experiences. They were given a five-minute break between the conversation sessions. After completing the two conversation sessions, they separately attended 10-15 minutes semi-structured interviews. The entire experiment took approximately 75 minutes per participant pair.

6.6 Measures

Similar to *study 1*, the quality of conversation and IoT manipulation were measured using objective and subjective measures. In addition, both the host's and guest's perceptions of their conversation, IoT manipulation behavior, and associated advantages/disadvantages of using the *ParaGlassMenu Interface* were captured in the interviews.

6.6.1 Quality of conversation. For both *IoT* and *No_IoT* conditions, subjective measures from hosts and guests on attention and concentration, eye contact, naturalness, and perceived impact of OHMD were collected following measures developed by McAtamney et al. [57] (see **Table 3**).

6.6.2 Quality of IoT manipulation. For the *IoT* condition, *Task Duration* and the usage of IoT manipulation (e.g., the number of times they used each device) were collected to assess how IoT manipulation performance was in the modeled setting.

In addition, subjective measures, such as *Relaxation* and *SUS* were collected from the hosts, similar to *study 1*. Besides, *IoT Interruption* ("Device manipulation by (me/my partner) did interrupt the conversation", 1 = Strongly Disagree, 7 = Strongly Agree) and *Politeness* were collected from both the hosts and guests, following [40]. Lastly, the rating of *Hospitality* ("My partner treated me well during the conversation, satisfying my needs by controlling some appliances") was collected from guests using a 7-point Likert Scale.

6.6.3 Analysis. Wilcoxon signed-rank test was used when the baseline data was present, or else descriptive statistics statistically were used to analyze the results. Lastly, the interview recordings were thematically analyzed in a manner similar to *study 1* (sec 5.7.3).

6.7 Results

Overall, the results indicated that hosts could manipulate IoT devices in a relaxed and polite manner using the *ParaGlassMenu Interface* to cater to both the hosts' and guests' needs with low interference to conversations. Figure 10a and Figure 10b show both the objective and subjective mean performance measures (see Appendix B.1 for details).

6.7.1 Overall experience. On average, the *IoT* sessions lasted 19.1 minutes (SD = 1.4) and each host carried out 7.5 IoT tasks (SD = 2.65). While the *No_IoT* sessions lasted 18.8 minutes (SD = 1.5). Note that the number of IoT tasks performed by each participant is higher than normal (≈ 3 voice commands per hour with "Alexa" [73, Fig 2]), which helps us to validate the *ParaGlassMenu* in more extreme usage.

In addition, as expected and shown in Figure 9, most IoT tasks were carried out by the hosts at the start of the conversation ($M = 4.58$, $SD = 2.39$ within the first 8 min; $M = 2.92$, $SD = 1.73$ for the rest of time) to set up the environment, such as turn on the light and fan. Furthermore, hosts carried out IoT tasks to entertain guests (e.g., play music) or to satisfy their guests' on-demand needs (e.g., adjust the light when the guest was scanning a magazine).

6.7.2 Quality of conversation.

Impact of using ParaGlassMenu: Overall, as shown in **Figure 10a**, both the hosts and the guests rated the conversation related scores as high (i.e., 25th percentile value above 5 out of 7) in terms of $AC1$ (*listening to guest*), $AC2$ (*concentration on conversation*), $AC3$ (*attention on guest*), $EC1$ (*eye contact with guest*), $NB1$ (*host acting naturally*), and $NB2$ (*host feeling relaxed*) for both *IoT* and *No_IoT* conditions. These results support that IoT manipulation with *ParaGlassMenu* had a low impact on the overall quality of conversation because it enabled the hosts to maintain attention and concentration on the guests and behave naturally during the conversation. Note that the relatively low scores on $IO1$ (*ignoring glasses*) largely came from the limitation of Nreal glasses, such as their non-negligible weight and semi-transparent lens (sec 6.8.3). We expect such drawbacks can be largely eliminated with lighter and more transparent glasses in the future [1].

Influence of multitasking: As expected, due to the need to multitask, *IoT* condition scored lower in certain measures: $AC2$ (*concentration on conversation*) for both hosts and guests, $AC3$ (*attention on guest*), $NB1$ (*host acting naturally*) for guests, as indicated by the significant differences ($p < 0.05$) labelled in Figure 10a. However, all these scores in *IoT* condition were still high (i.e., 25th percentile value above 5 out of 7). In addition, the feedback in the interview suggested that while participants could notice the above difference between the two conditions, the magnitude of the difference was acceptable as "we [both hosts and guests] maintained good conversations in both sessions. ($P2_{Guest}$)"

Comparison with other interfaces: A few participants (2 hosts and 3 guests) felt that the *ParaGlassMenu* helped them to be much more engaged in conversation as compared with past experiences of using either the *Phone* or *Voice Interface* to control IoT devices, because "the ring mouse is subtle enough that you can use it without interrupting the flow of the conversation. ($P12_{Host}$)", "If the host uses phone and voice to control the devices, the conversation could be paused during his operations. ($P6_{Guest}$)"

6.7.3 Quality of IoT manipulation.

Task duration: Table 4 shows the descriptive statistics of the IoT tasks. Compared with *study 1*, the average IoT task duration in this study was generally higher because most tasks were not urgent, and the hosts attempted to make a "cozy environment" for guests in a relaxed manner.

SUS, Politeness, Interruption, and Relaxation: Hosts found that the *ParaGlassMenu* had high usability (*SUS* score: $M = 84.583$, $SD = 8.45$), and both hosts and guests rated high *Politeness* for *ParaGlassMenu* as it enabled "convenient and fast manipulation ($P3_{Host}$)" with low *IoT Interruption* to the conversation (see Figure 10b); yet, they still faced several issues, which are described below (sec 6.8.3).

Besides, all hosts felt relaxed when they manipulated the devices during their conversation and they elaborated in the interviews that controlling devices via the *ParaGlassMenu* can "reduce [users'] manual labor, increase their conversation time, and make the conversation smoother." ($P1_{Host}$)"

Hospitality: All guests rated high *Hospitality* from hosts, as hosts could attend to their requests fluently without interrupting the conversation, which made them feel more "welcomed and comfortable".

6.8 Discussion

The results verify the usability of *ParaGlassMenu* in realistic conversations. In the aspect of subtle interactions, the non-intrusiveness of *ParaGlassMenu* helped hosts focus easily on their partner's face and reduced the frequency in which "missing important information [non-verbal cues] from the partner's conversation ($P6_{Host}$)". And the host's subtle interactions not only provided a comfortable conversation environment but also avoided unnecessary interruption and made the conversation flow smooth.

In addition, users' manipulation behaviors in realistic conversations and the current prototype's limitations are discussed.

Table 3: Aspects and measures adopted from McAtamney et al [57] on conversation behavior of the host from hosts' and guests' points of view. Each measure is rated using a 7-point Likert scale.

Aspect on conversation	Measures
Attention and concentration	**AC1** (*listening to guest*): [host] 'When the other person was speaking, I was always listening to them' / [guest] 'When I was speaking, I think the other person was always listening to me' **AC2** (*concentration on conversation*): [host] 'I was always concentrating on the conversation' / [guest] 'I think the other person was always concentrating on the conversation' **AC3** (*attention on guest*): [host]'When I was speaking, my attention was towards the other person' / [guest] 'When the other person was speaking their attention was towards me'
Eye contact	**EC1** (*eye contact with guest*): [guest] 'When I was speaking the other person maintained eye contact'
Naturalness	**NB1** (*host acting naturally*): [host] 'I acted naturally at all times during the conversation' / [guest] 'The other person acted naturally at all times during the conversation' **NB2** (*host feeling relaxed*): [host] 'I felt relaxed during the conversation'/ [guest] ' The other person appeared relaxed during the conversation'
Impact of OHMD	**IO1** (*ignoring glasses*): [host] It was easy to ignore the fact that I was wearing smart glasses / [guest]'It was easy to ignore the fact that the other person was wearing smart glasses'

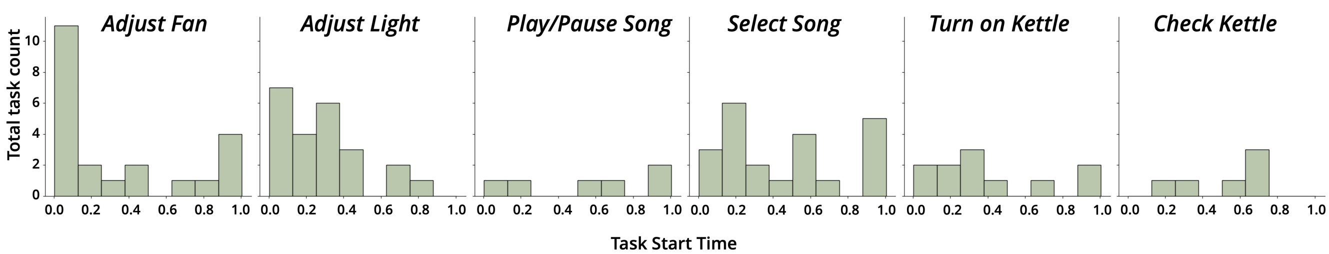


Figure 9: IoT task distribution during the conversation. The X-axis represents the start time with respect to total conversation duration, while Y-axis represents the total IoT task count for all 12 participant pairs.

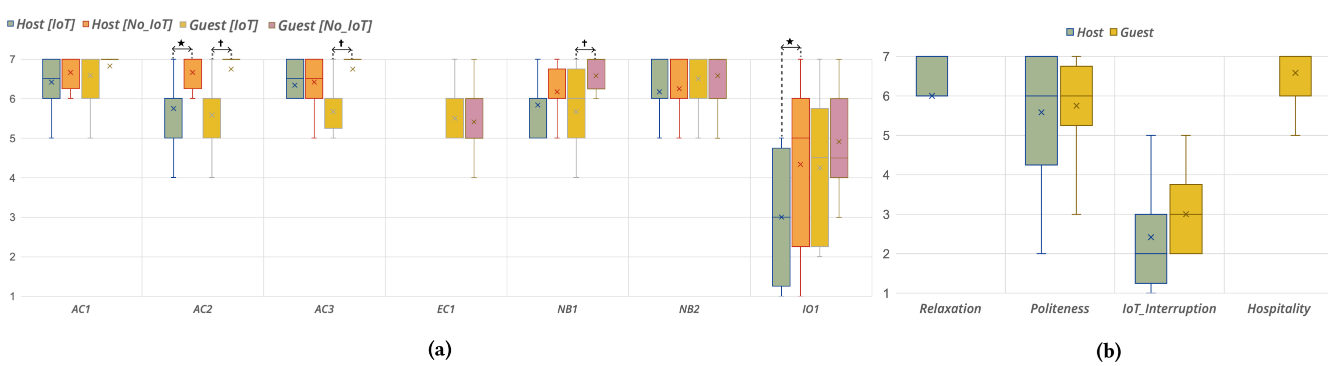


Figure 10: Quality of Conversation and IoT manipulation. (a) Conversation Quality ratings by hosts and guests during conversations for both *IoT* and *No_IoT* conditions (N = 12 x 2). The maximum score for each measure is 7. * and † represent significant ($p < 0.05$) post-hoc tests and $\times$ inside box plot represents the mean value point. The definition of the abbreviations used in the x-axis (e.g., AC1) can be found in Table 3. See Appendix B.1 (Table 7 and Table 8) for details. (b) Subjective ratings of Quality of IoT Manipulation by hosts and guests during conversations (N = 12 x 2). $\times$ inside box plot represents the mean value point. See Appendix B.1 (Table 9) for details.

6.8.1 Menu deactivation. Although manual menu deactivation was supported to minimize the unnecessary visual intrusiveness during conversations, the majority of the hosts (8) did not deactivate it. There are two reasons. Firstly, when participants focused on the conversation partner, the menu items circularly displayed in the peripheral vision remained non-intrusive to their focus on the partner. Secondly, participants tended to keep the menu on after starting pending tasks (e.g., hosts turn on the kettle, but it's not

Table 4: The average IoT task count and task duration (in seconds; in order '$mean$ (sd) $[min,\ max]$') per participant pair.

IoT Task	Manipulation Type	Average Task Count	Task duration
Adjust Light	*Continuous Manipulation*	1.92	9.04 (4.48) [2, 20]
Adjust Fan	*Continuous Manipulation*	1.83	9.77 (5.58) [3, 26]
Play/Pause Song	*Discrete Manipulation*	0.5	7.00 (6.13) [1, 18]
Select Song	*Selecting From List*	1.83	16.59 (9.69) [5, 43]
Turn on Kettle	*Discrete Manipulation*	0.92	6.45 (3.11) [3, 13]
Check Kettle	*Checking Info*	0.5	3.17 (0.41) [3, 4]

yet boiled) as it "saved time" on checking the task's status without being noticed by others. This suggests auto menu deactivation for non-pending tasks. In addition, upon menu activation, quickly resuming pending tasks helps to reduce manipulation time.

6.8.2 Manipulation visibility. Besides immediately satisfying the IoT manipulation requests from the guests, two patterns of visibility of IoT manipulation were observed for all the hosts. First, hosts attempted to manipulate devices discreetly when the guest was speaking or during a pause in the conversation. Second, they verbally highlighted their manipulation task to their partner before the manipulation (e.g., "let me turn on the light").

The different manipulation visibility indicated that users make decisions on awareness and social etiquette management in conversations; specifically, the visibility of hosts' manipulations depended on whether the guest could notice the **effects of manipulation** (e.g., light brightness' change) and whether the manipulation is **relevant to the conversation**. Generally, hiding digital interactions irrelevant to the conversation (e.g., turning on the kettle in the kitchen) could help hosts avoid distractions to others. However, if the guest could notice the effects of manipulation without being informed, they would realize that the host was distracting from the conversation, which violated the social norms [24]. In this case, if the manipulation is relevant to the conversation topic or involved parties, hosts would inform the guests in advance to avoid unexpected distractions and impoliteness to the conversations. The design implications for the manipulation visibility are discussed in sec 7.2.2.

6.8.3 Limitations with current prototype. Several hardware issues were reported by the participants. Similar to *study 1*, a few hosts (2) found that using 'scrolling' to select songs or change the brightness/speed of light/fan was not precise enough, which led them to select the wrong options. Besides, guests could sometimes notice the button click sound, but all of them mentioned it was acceptable when they knew manipulations were done to cater to their needs. In addition, all hosts noted that the OHMD was a bit heavy and slippery, thus not convenient to wear for a long time. All guests mentioned that the black lenses made OHMDs harder to ignore.

Moreover, unfamiliarity issues also existed at the start of the study. For example, a few hosts (2) mentioned that they forgot how to "clear" the menu (i.e., deactivate *ParaGlassMenu Interface*) and needed several attempts at the beginning of the conversation. To minimize such issues and support novice users, help 'hints' indicating the ring mouse mapping with the next possible tasks in the current menu can be used.

7 OVERALL DISCUSSION

7.1 Using *ParaGlassMenu* in conversation

7.1.1 Quality of conversation and IoT manipulation. *ParaGlassMenu* has been found to facilitate higher-quality conversation when interacting with a secondary digital task. In comparison with other interfaces, it has been shown to have comparable *Face Focus* with *Voice* and significantly higher *Face Focus* than that of *Linear Interface* and *Phone*. Furthermore, *ParaGlassMenu* has demonstrated the highest level of *Politeness* and *Naturalness*, as well as the lowest *RTLX* score.

ParaGlassMenu has also been found to improve the quality of IoT manipulation. In comparison with other interfaces, it has been shown to have the lowest *Task Duration*, highest *Task Accuracy*, highest *SUS* score, and second-highest *Relaxation* score. These findings suggest that *ParaGlassMenu* is a promising interface for facilitating IoT manipulation during social conversations.

Our results were further validated in more realistic conversation settings, as participants showed that *ParaGlassMenu* can support digital interaction with low interference to conversations.

7.1.2 Ability to multitask during social interactions. The majority of participants (10/12 in *study 2*) reported that they were able to effectively manage both conversation and IoT manipulation using *ParaGlassMenu* due to the non-intrusive nature of its visual feedback and discreet interactions. This is in contrast to previous research on subtle interaction, such as jaw-teeth [8] and silent-speech interaction [46], which is challenging to use during social conversations. However, there were instances where participants struggled to maintain attention when manipulation of IoT tasks became more complex (e.g., selecting songs). In these cases, the increased cognitive demands of visual searching and decision-making interfered with their ability to speak and maintain eye contact and resulted in a slight decrease (7% on average) in conversation quality.

Despite these challenges, we observed that participants employed various strategies to mitigate the negative impacts of complex IoT manipulation on conversation quality. One approach (used by 6 out of 12 participants) involved manipulating the IoT device while listening, or during pauses in the conversation, in order to reduce competition for mental resources and minimize interference with speaking, aligned with the theories of capacity sharing and bottleneck model [62]. Another strategy (used by 3 participants) was to delay and slow down manipulation until a less important moment in the conversation, rather than immediately reacting to interaction needs. Additionally, one participant ($P11_{Host}$) reported that the negative impact of digital interaction on conversation quality can be reduced with practice. Previous research has shown that with sufficient training, a task can be performed "automatically" without consuming significant mental resources [17, 39].

7.1.3 Visibility of Manipulation. Given users have different interaction needs in different scenarios (sec 2.1), the ability to flexibly select the visibility of IoT manipulation can help users to follow social norms. The selection is determined by the visibility of effects (sec 6.8.2, e.g., whether producing transparent or discreet manipulation effects) and type of manipulation (i.e., whether it is relevant or irrelevant to the conversation). For example, in order to satisfy the needs of conversation partners and avoid unexpected environmental changes, users conducted over half of the IoT manipulations (51 out of 90 manipulations) transparently by verbally mentioning their actions in advance. Transparent manipulations were also necessary for tasks that involved shared control, such as selecting background music according to guests' preferences (14 out of 51 transparent manipulations).

In contrast, when the digital manipulation tasks were irrelevant to the conversation topics (39 out of 90 manipulations), users chose to perform them discreetly (i.e., opaque) to minimize interference with conversations. Furthermore, participants reported that privacy concerns [54] can also influence the visibility of digital interaction. For example, $P10_{Guest}$ discussed a scenario in which he was engaged in a conversation with a guest while his baby was sleeping in another room. He needed to check on the baby's status periodically using a monitor app on their phone, and suggested that *ParaGlassMenu* could be particularly useful in this scenario. Overall, participants found the ability to discreetly access urgent private information using *ParaGlassMenu* to be useful and appealing.

Considering users' needs for both transparent and opaque manipulations, designing an interface that supports both types of visibility is necessary (see details in sec 7.2.2). This allows users to select the appropriate level of visibility for the current interaction context and social norms. By providing this level of flexibility, the interface can better support the varied interaction needs of users in different scenarios.

7.1.4 Supporting other application scenarios. In addition to facilitating IoT control, *ParaGlassMenu* can be used in a variety of application scenarios. For example, in the scenario mentioned in the introduction, John could use *ParaGlassMenu*, as shown in Figure 11a, to discreetly select a default message response to reply to Nicole (e.g., "Can I call you later?") without significantly interrupting his face-to-face conversation.

While *ParaGlassMenu* is designed for social interaction, it can also be used for other scenarios where users have to focus on a visual target while performing digital operations simultaneously. An example scenario is shown in Figure 11b, where one can utilize the *ParaGlassMenu* to record the important points in a class while remaining concentrated on the lecturer. With a few clicks, a student can start/stop recording the video and highlight important moments, which provides convenience for reviewing the lecture. Besides, the circular progress bar with text presents the recording time non-intrusively for students. Another application scenario can be drone interaction. With *ParaGlassMenu*, the user can perform commands such as controlling the flight and taking photos while maintaining attention to the drone's position and motion in the sky.

7.2 Limitations and potential enhancements for the *ParaGlassMenu*

Based on the findings from the two studies, several issues with the current prototype of *ParaGlassMenu* were identified, including hardware and implementation-related concerns. As such, the following recommendations are suggested to enhance the usability of the *ParaGlassMenu*.

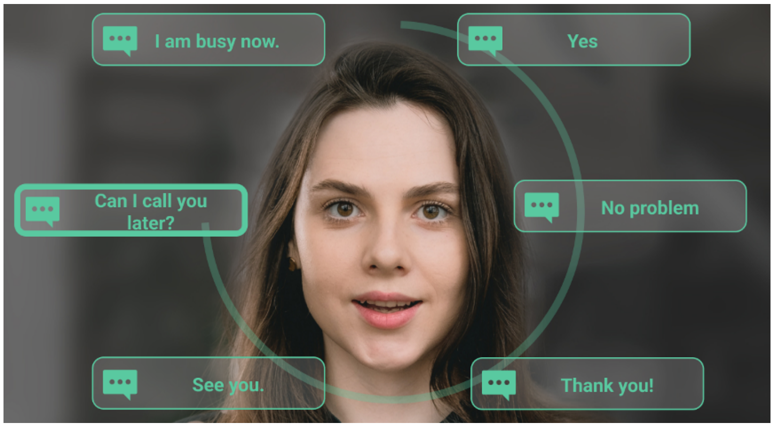
7.2.1 Enhancing thumb-index interaction. We used an off-the-shelf ring mouse for the *ParaGlassMenu* prototype, but it did not always provide precise scrolling and only supported a maximum of four clickable items. To improve the interface, a new ring mouse can be designed with a click wheel similar to iPod Classic[8] or earPod [83] or using DeformWear's technology [80] for multiple item selection. This would enable smooth scrolling and support a larger number of clickable items, improving the usability and functionality of the ParaGlassMenu prototype.

7.2.2 Supporting two visibility of manipulation. *ParaGlassMenu* is designed to support discreet rather than transparent interactions. Currently, transparent interactions are achieved by the host verbally informing the guest of their intentions beforehand. In the future, the design of *ParaGlassMenu* could consider introducing a transparent mode that provides visual or audio cues to the conversation partner during the interaction. For tasks that involve shared decision-making (e.g., selecting a suitable song), using natural language processing to analyze the conversation between the host and guest may be a more natural way to achieve this task. This could involve identifying relevant keywords or phrases in the conversation and using them to filter the options for song selection, reducing the search space and enabling more efficient decision-making. These enhancements could improve the support for transparent tasks and further enhance the user experience in social settings.

7.3 Trade-off of subtle interactions in social settings

By using discreet manipulation with attention-maintaining visualization, users can attend to their digital interaction needs with minimum distraction to their primary social interaction. This can be particularly useful for handling the social needs of multiple relationships, such as maintaining a physical conversation with one person while also responding to remote social inquiries. While we highlighted the benefits of using subtle interactions in the social context, it nevertheless has trade-offs. One potential downside of subtle interactions with digital information is that they may be misused, leading to increased distractions during social interactions. For example, if users find it easier to interact with their digital devices in a subtle manner, they may be tempted to do so more frequently, even in situations where it is not appropriate. Another potential downside of subtle interactions with digital information is that they can cause misperceptions between conversation partners. In *study 2*, one guest reported a higher estimated number of manipulations by the host than was actually the case. This suggests that when people are aware of the possibility of subtle interactions,

[8] https://manuals.info.apple.com/MANUALS/0/MA630/en_US/iPod_classic_120GB_en.pdf



(a)

(b)

Figure 11: Other application scenarios for *ParaGlassMenu*. (a) Users can receive a message in a physical conversation and subtly select a default response to reply to the remote social inquiry with *ParaGlassMenu*. (b) A student can record videos using *ParaGlassMenu* while maintaining attention on the teacher.

they may misinterpret natural behaviors, such as head movements or shifts in eye focus, as manipulations of digital information.

To prevent potential misuse, it's essential to make it easy to limit the usage of technology according to social contexts and provide an option to switch to "focused" mode, where attention can be solely devoted to the conversation partner. Additional visualizations, such as using low opacity for menu items that are not relevant to the conversation, can also help to restrict misuse.

To avoid misperception, users may need to pay more attention to their behaviors or be more transparent. For instance, when performing digital tasks that are related to the conversation, users can use the transparent mode to make it clear to their conversation partner that they're not disengaged. Alternatively, users can explain their motivation for discreet manipulation at the start of the interaction, so that their partner understands why they may need to use technology during the conversation.

In summary, while the use of subtle interaction techniques such as *ParaGlassMenu* in social contexts can offer many benefits, it is important to consider the potential drawbacks and take steps to mitigate them. Responsible design practices and attention to behaviors and transparency can help prevent misuses and misperceptions. Overall, the use of subtle interactions can enhance social interactions when used thoughtfully and with awareness of potential impacts.

8 LIMITATIONS

Despite the results obtained from these studies supporting *ParaGlassMenu*, there are some limitations that need to be considered. First, in *study 1*, all participants didn't wear spectacles due to eye tracking. However, the overall preference could be affected by the spectacle-wearing experience. In addition, participants' past experience and lack of familiarity with selected interfaces might also affect the task completion at the beginning. Furthermore, while a virtual conversation partner and immersive virtual home environment were utilized after considering trade-offs, the external validity of *Naturalness* and *Politeness* in such conversation settings is limited compared with real conversations, and thus, further investigation is required.

Second, in *study 2*, although participants had sufficient time for training, users may still not be fully familiar with the interface and modeled room at the beginning of the experiment. In addition, only a limited number of IoT devices were provided in the modeled room due to space limitation.

Third, the OHMD used in our studies is still not ideal for long-term everyday usage. Future OHMDs with lighter-weight, transparent lenses, better computational capabilities, and longer battery life can significantly improve their comfort and social acceptability.

Finally, the study participants were selected from the local university community, as this tech-savvy group has more experience with smart device usage and is more likely to be early adopters of OHMDs and our technique. However, it's important to note that social acceptance and device usability may vary across different cultural backgrounds and age groups. To generalize the results and better understand the long-term effects, longitudinal studies with different user groups using various IoT devices and OHMD prototypes need to be conducted. This will help to identify any differences in user experiences and preferences, and inform the development of more inclusive and user-friendly technology.

9 CONCLUSION AND FUTURE WORK

We utilized non-intrusive circular OHMD menus with discreet thumb-index interactions to support digital interactions in social settings and studied its usage as a socially friendly IoT manipulation interface. By comparing the proposed *ParaGlassMenu* with *Phone*, *Voice*, and *Linear* interfaces and testing the usage of the *ParaGlassMenu* in both simulated and more realistic conversation settings, it was verified that *ParaGlassMenu* could largely support a variety of interactions during face-to-face social conversations, making it an effective attention-maintaining subtle interface. Future work can explore other application scenarios of *ParaGlassMenu* designs and incorporate additional feedback in audio to further enhance humans' abilities to handle their interaction needs in social settings.

ACKNOWLEDGMENTS

This research is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-RP-2020-016). It is also supported in part by the Ministry of Education, Singapore, under its MOE Academic Research Fund Tier 2 Programme (MOE-T2EP20221-0010), and by a research grant #22-5913-A0001 from the Ministry of Education of Singapore. In addition, this research is part of the programme DesCartes and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. We thank all members of the NUS-HCI lab who helped to complete this project, Zhuoya Yang for helping draw figures, and all reviewers for their valuable feedback.

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] Naresh Kumar Agarwal and Wenqing Lu. 2020. Response to non-response: How people react when their smartphone messages and calls are ignored. *Proceedings of the Association for Information Science and Technology* 57, 1 (2020), e260.

[3] Fouad Alallah, Ali Neshati, Yumiko Sakamoto, Khalad Hasan, Edward Lank, Andrea Bunt, and Pourang Irani. 2018. Performer vs. observer: whose comfort level should we consider when examining the social acceptability of input modalities for head-worn display?. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3281505.3281541

[4] Amr Alanwar, Moustafa Alzantot, Bo-Jhang Ho, Paul Martin, and Mani Srivastava. 2017. SeleCon: Scalable IoT Device Selection and Control Using Hand Gestures. In *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation* (Pittsburgh, PA, USA) *(IoTDI '17)*. Association for Computing Machinery, New York, NY, USA, 47–58. https://doi.org/10.1145/3054977.3054981

[5] Fraser Anderson, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2015. Supporting Subtlety with Deceptive Devices and Illusory Interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 1489–1498. https://doi.org/10.1145/2702123.2702336

[6] Michael Argyle and Mark Cook. 1976. Gaze and mutual gaze. (1976).

[7] Michael Argyle and Janet Dean. 1965. Eye-contact, distance and affiliation. *Sociometry* (1965), 289–304.

[8] Daniel Ashbrook, Carlos Tejada, Dhwanit Mehta, Anthony Jiminez, Goudam Muralitharam, Sangeeta Gajendra, and Ross Tallents. 2016. Bitey: An Exploration of Tooth Click Gestures for Hands-Free User Interface Control. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 158–169. https://doi.org/10.1145/2935334.2935389

[9] Gilles Bailly, Eric Lecolinet, and Laurence Nigay. 2008. Flower Menus: A New Type of Marking Menu with Large Menu Breadth, within Groups and Efficient Expert Mode Memorization. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (Napoli, Italy) *(AVI '08)*. Association for Computing Machinery, New York, NY, USA, 15–22. https://doi.org/10.1145/1385569.1385575

[10] Gilles Bailly, Eric Lecolinet, and Laurence Nigay. 2016. Visual Menu Techniques. *ACM Comput. Surv.* 49, 4, Article 60 (dec 2016), 41 pages. https://doi.org/10.1145/3002171

[11] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Studies* 4, 3 (may 2009), 114–123.

[12] Ahmet Börütecene, Idil Bostan, Ekin Akyürek, Alpay Sabuncuoglu, Ilker Temuzkusu, Çaglar Genç, Tilbe Göksun, and Oguzhan Özcan. 2018. Through the Glance Mug: A Familiar Artefact to Support Opportunistic Search in Meetings. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction* (Stockholm, Sweden) *(TEI '18)*. Association for Computing Machinery, New York, NY, USA, 674–683. https://doi.org/10.1145/3173225.3173236

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[14] John Brooke. 1996. SUS - A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 7.

[15] Marc Brysbaert. 2019. How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language* 109 (2019), 104047. https://doi.org/10.1016/j.jml.2019.104047

[16] J. Callahan, D. Hopkins, M. Weiser, and B. Shneiderman. 1988. An Empirical Comparison of Pie vs. Linear Menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Washington, D.C., USA) *(CHI '88)*. Association for Computing Machinery, New York, NY, USA, 95–100. https://doi.org/10.1145/57167.57182

[17] L Mark Carrier, Larry D Rosen, Nancy A Cheever, and Alex F Lim. 2015. Causes, effects, and practicalities of everyday multitasking. *Developmental Review* 35 (2015), 64–78.

[18] Isha Chaturvedi, Farshid Hassani Bijarbooneh, Tristan Braud, and Pan Hui. 2019. Peripheral vision: a new killer app for smart glasses. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 625–636. https://doi.org/10.1145/3301275.3302263

[19] Varoth Chotpitayasunondh and Karen M. Douglas. 2018. The effects of "phubbing" on social interaction. *Journal of Applied Social Psychology* 48, 6 (2018), 304–316. https://doi.org/10.1111/jasp.12506 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jasp.12506.

[20] Enrico Costanza, Samuel A. Inverso, and Rebecca Allen. 2005. Toward Subtle Intimate Interfaces for Mobile Devices Using an EMG Controller. 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, 481–489. https://doi.org/10.1145/1054972.1055039

[21] Enrico Costanza, Samuel A. Inverso, Elan Pavlov, Rebecca Allen, and Pattie Maes. 2006. Eye-q: Eyeglass Peripheral Display for Subtle Intimate Notifications. In *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services* (Helsinki, Finland) *(MobileHCI '06)*. Association for Computing Machinery, New York, NY, USA, 211–218. https://doi.org/10.1145/1152215.1152261

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

[23] Rajib Dey, Sayma Sultana, Afsaneh Razi, and Pamela J. Wisniewski. 2020. Exploring Smart Home Device Use by Airbnb Hosts. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3382900

[24] Sarah Diefenbach and Daniel Ullrich. 2019. Disrespectful Technologies: Social Norm Conflicts in Digital Worlds. In *Advances in Usability, User Experience and Assistive Technology*, Tareq Z. Ahram and Christianne Falcão (Eds.). Springer International Publishing, Cham, 44–56.

[25] Jérémie Francone, Gilles Bailly, Eric Lecolinet, Nadine Mandran, and Laurence Nigay. 2010. Wavelet Menus on Handheld Devices: Stacking Metaphor for Novice Mode and Eyes-Free Selection for Expert Mode. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Roma, Italy) *(AVI '10)*. Association for Computing Machinery, New York, NY, USA, 173–180. https://doi.org/10.1145/1842993.1843025

[26] Koumei Fukahori, Daisuke Sakamoto, and Takeo Igarashi. 2015. Exploring Subtle Foot Plantar-Based Gestures with Sock-Placed Pressure Sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3019–3028. https://doi.org/10.1145/2702123.2702308

[27] Google. 2022. Google home - apps on Google Play. https://play.google.com/store/apps/details?id=com.google.android.apps.chromecast.app

[28] Google. 2022. Smart Home Device Traits | Actions on Google Smart Home. https://developers.google.com/assistant/smarthome/traits

[29] Hall and Edward Twitchell. 1966. *The hidden dimension*. Vol. 609. Anchor.

[30] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

[31] Roy S. Hessels. 2020. How does gaze to faces support face-to-face interaction? A review and perspective. *Psychonomic Bulletin & Review* 27, 5 (Oct. 2020), 856–881. https://doi.org/10.3758/s13423-020-01715-w

[32] Yi-Ta Hsieh, Antti Jylhä, Valeria Orso, Luciano Gamberini, and Giulio Jacucci. 2016. Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4203–4215. https://doi.org/10.1145/2858036.2858436

[33] Shamsi T. Iqbal, Jonathan Grudin, and Eric Horvitz. 2011. Peripheral computing during presentations: perspectives on costs and preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 891–894. https://doi.org/10.1145/1978942.1979073

[34] Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2021. Towards Indistinguishable Augmented Reality: A Survey on Optical See-through Head-mounted Displays. *Comput. Surveys* 54, 6 (July 2021), 120:1–120:36. https://doi.org/10.1145/3453157

[35] Timo Jakobi, Corinna Ogonowski, Nico Castelli, Gunnar Stevens, and Volker Wulf. 2017. The Catch(Es) with Smart Home: Experiences of a Living Lab Field Study. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1620–1633. https://doi.org/10.1145/3025453.3025799

[36] Nuwan Janaka, Chloe Haigh, Hyeongcheol Kim, Shan Zhang, and Shengdong Zhao. 2022. Paracentral and near-peripheral visualizations: Towards attention-maintaining secondary information presentation on OHMDs during in-person social interactions. (2022), 14. https://doi.org/10.1145/3491102.3502127

[37] Rikke Hagensby Jensen, Yolande Strengers, Jesper Kjeldskov, Larissa Nicholls, and Mikael B. Skov. 2018. Designing the Desirable Smart Home: A Study of Household Experiences and Energy Consumption Impacts. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173578

[38] 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

[39] Paul A. Kirschner and Aryn C. Karpinski. 2010. Facebook® and academic performance. *Computers in Human Behavior* 26, 6 (2010), 1237–1245. https://doi.org/10.1016/j.chb.2010.03.024

[40] Marion Koelle, Swamy Ananthanarayan, and Susanne Boll. 2020. Social Acceptability in HCI: A Survey of Methods, Measures, and Design Strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–19. https://doi.org/10.1145/3313831.3376162

[41] Marion Koelle, Matthias Kranz, and Andreas Möller. 2015. Don't look at me that way!: Understanding User Attitudes Towards Data Glasses Usage. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Copenhagen Denmark, 362–372. https://doi.org/10.1145/2785830.2785842

[42] Quan Kong, Takuya Maekawa, Taiki Miyanishi, and Takayuki Suyama. 2016. Selecting Home Appliances with Smart Glass Based on Contextual Information. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 97–108. https://doi.org/10.1145/2971648.2971651

[43] Nallapaneni Manoj Kumar, Neeraj Kumar Singh, and V. K. Peddiny. 2018. Wearable Smart Glass: Features, Applications, Current Progress and Challenges. In *2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT)*. 577–582. https://doi.org/10.1109/ICGCIoT.2018.8753047

[44] Gordon Kurtenbach and William Buxton. 1993. The Limits of Expert Performance Using Hierarchic Marking Menus. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. Association for Computing Machinery, New York, NY, USA, 482–487. https://doi.org/10.1145/169059.169426

[45] Recommended Light Levels. 2020. Recommended light levels (illuminance) for outdoor and indoor venues. *The Engineering Toolbox. Recommended Light Levels* (2020).

[46] Richard Li, Jason Wu, and Thad Starner. 2019. TongueBoard: An Oral Interface for Subtle Input. In *Proceedings of the 10th Augmented Human International Conference 2019* (Reims, France) *(AH2019)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3311823.3311831

[47] Ziyang Li and Pei-Luen Patrick Rau. 2021. Talking with an IoT-CA: Effects of the Use of Internet of Things Conversational Agents on Face-to-Face Conversations. *Interacting with Computers* 33, 3 (10 2021), 238–249. https://doi.org/10.1093/iwc/iwab024 arXiv:https://academic.oup.com/iwc/article-pdf/33/3/238/41062328/iwab024.pdf

[48] Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2019. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science* 51, 4 (2019), 984–997. https://doi.org/10.1177/0961000618759414

[49] Feiyu Lu, Shakiba Davari, Lee Lisle, Yuan Li, and Doug A. Bowman. 2020. Glanceable AR: Evaluating Information Access Methods for Head-Worn Augmented Reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 930–939. https://doi.org/10.1109/VR46266.2020.00113

[50] Michal Luria, Guy Hoffman, and Oren Zuckerman. 2017. Comparing social robot, screen and voice interfaces for smart-home control. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 580–628. https://doi.org/10.1145/3025453.3025786

[51] William G. Lycan. 1977. Conversation, politeness, and interruption. *Paper in Linguistics* 10, 1-2 (1977), 23–53. https://doi.org/10.1080/08351819709370438

[52] Kent Lyons. 2003. Everyday Wearable Computer Use: A Case Study of an Expert User. In *Human-Computer Interaction with Mobile Devices and Services (Lecture Notes in Computer Science)*, Luca Chittaro (Ed.). Springer, Berlin, Heidelberg, 61–75. https://doi.org/10.1007/978-3-540-45233-1_6

[53] I. Scott MacKenzie. 2013. *Human-computer interaction: an empirical research perspective* (first edition ed.). Morgan Kaufmann is an imprint of Elsevier, Amsterdam.

[54] Ville Mäkelä, Johannes Kleine, Maxine Hood, Florian Alt, and Albrecht Schmidt. 2021. Hidden Interaction Techniques: Concealed Information Acquisition and Texting on Smartphones and Wearables. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 248, 14 pages. https://doi.org/10.1145/3411764.3445504

[55] Adam M. Mastroianni, Daniel T. Gilbert, Gus Cooney, and Timothy D. Wilson. 2021. Do conversations end when people want them to? *Proceedings of the National Academy of Sciences of the United States of America* 118, 10 (March 2021), e2011809118. https://doi.org/10.1073/pnas.2011809118

[56] Sven Mayer, Lars Lischke, Paweł W. Woźniak, and Niels Henze. 2018. Evaluating the Disruptiveness of Mobile Interactions: A Mixed-Method Approach. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–14. https://doi.org/10.1145/3173574.3173980

[57] Gerard McAtamney and Caroline Parker. [n. d.]. An examination of the effects of a wearable display on informal face-to-face communication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2006-04-22) *(CHI '06)*. Association for Computing Machinery, 45–54. https://doi.org/10.1145/1124772.1124780

[58] Tien T. Nguyen, Duyen T. Nguyen, Shamsi T. Iqbal, and Eyal Ofek. 2015. The Known Stranger: Supporting Conversations between Strangers with Personalized Topic Suggestions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 555–564. https://doi.org/10.1145/2702123.2702411

[59] Donald A. Norman. 2013. *The design of everyday things* (revised and expanded edition ed.). Basic Books.

[60] Eyal Ofek, Shamsi T. Iqbal, and Karin Strauss. [n. d.]. Reducing disruption from subtle information delivery during a conversation: mode and bandwidth investigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013-04-27) *(CHI '13)*. Association for Computing Machinery, 3111–3120. https://doi.org/10.1145/2470654.2466425

[61] Julius Panero and Martin Zelnik. 1979. *Human dimension & interior space: a source book of design reference standards*. Watson-Guptill.

[62] Harold Pashler. 1994. Dual-task interference in simple tasks: data and theory. *Psychological bulletin* 116, 2 (1994), 220.

[63] Simon T. Perrault, Eric Lecolinet, Yoann Pascal Bourse, Shengdong Zhao, and Yves Guiard. 2015. Physical Loci: Leveraging Spatial, Object and Semantic Memory for Command Selection. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 299–308. https://doi.org/10.1145/2702123.2702126

[64] Henning Pohl, Andreea Muresan, and Kasper Hornbæk. 2019. *Charting Subtle Interaction in the HCI Literature*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300648

[65] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174214

[66] Aung Pyae and Tapani N. Joelsson. 2018. Investigating the Usability and User Experiences of Voice User Interface: A Case of Google Home Smart Speaker. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Barcelona, Spain) *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, 127–131. https://doi.org/10.1145/3236112.3236130

[67] Mikko J. Rissanen, Samantha Vu, Owen Noel Newton Fernando, Natalie Pang, and Schubert Foo. 2013. Subtle, Natural and Socially Acceptable Interaction Techniques for Ringterfaces — Finger-Ring Shaped User Interfaces. In *Distributed, Ambient, and Pervasive Interactions*, Norbert Streitz and Constantine Stephanidis (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 52–61.

[68] Radiah Rivu, Yasmeen Abdrabou, Ken Pfeuffer, Augusto Esteves, Stefanie Meitner, and Florian Alt. 2020. StARe: Gaze-Assisted Face-to-Face Communication in Augmented Reality. In *Symposium on Eye Tracking Research and Applications*. ACM, Stuttgart Germany, 1–5. https://doi.org/10.1145/3379157.3388930 ZSCC: 0000013.

[69] Rufat Rzayev, Susanne Korbely, Milena Maul, Alina Schark, Valentin Schwind, and Niels Henze. 2020. *Effects of Position and Alignment of Notifications on AR Glasses during Social Interaction*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3419249.3420095

[70] Krystian Samp and Stefan Decker. 2010. Supporting Menu Design with Radial Layouts. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Roma, Italy) *(AVI ’10)*. Association for Computing Machinery, New York, NY, USA, 155–162. https://doi.org/10.1145/1842993.1843021

[71] Shardul Sapkota, Ashwin Ram, and Shengdong Zhao. 2021. Ubiquitous Interactions for Heads-Up Computing: Understanding Users’ Preferences for Subtle Interaction Techniques in Everyday Settings. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction (MobileHCI ’21)*. Association for Computing Machinery, New York, NY, USA, Article 36, 15 pages. https://doi.org/10.1145/3447526.3472035

[72] B. Schilit, N. Adams, and R. Want. 1994. Context-Aware Computing Applications. In *1994 First Workshop on Mobile Computing Systems and Applications*. 85–90. https://doi.org/10.1109/WMCSA.1994.16

[73] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "Hey Alexa, What’s Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS ’18)*. Association for Computing Machinery, New York, NY, USA, 857–868. https://doi.org/10.1145/3196709.3196772

[74] Ludwig Sidenmark, Dominic Potts, Bill Bapisch, and Hans Gellersen. 2021. RadiEye: Hands-Free Radial Interfaces for 3D Interaction Using Gaze-Activated Head-Crossing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI ’21)*. Association for Computing Machinery, New York, NY, USA, Article 740, 11 pages. https://doi.org/10.1145/3411764.3445697

[75] Norman Makoto Su and Lulu Wang. 2015. From Third to Surveilled Place: The Mobile in Irish Pubs. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI ’15)*. Association for Computing Machinery, New York, NY, USA, 1659–1668. https://doi.org/10.1145/2702123.2702574

[76] Charles Tijus, Javier Barcenilla, Brigitte Cambon de Lavalette, and Jean-Guy Meunier. 2007. The Design, Understanding and Usage of Pictograms. *Written Documents in the Workplace* (Jan. 2007), 17–31. https://doi.org/10.1163/9789004253254_003

[77] 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. https://doi.org/10.1145/2702123.2702214

[78] Mariek M.P. Vanden Abeele, Marjolijn L. Antheunis, and Alexander P. Schouten. 2016. The effect of mobile messaging during a conversation on impression formation and interaction quality. *Computers in Human Behavior* 62 (2016), 562–569. https://doi.org/10.1016/j.chb.2016.04.005

[79] Radu-Daniel Vatavu and Jean Vanderdonckt. 2020. Design Space and Users’ Preferences for Smartglasses Graphical Menus: A Vignette Study. In *19th International Conference on Mobile and Ubiquitous Multimedia* (Essen, Germany) *(MUM 2020)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://dl.acm.org/doi/10.1145/3428361.3428467

[80] Martin Weigel and Jürgen Steimle. 2017. DeformWear: Deformation Input on Tiny Wearable Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (June 2017), 28:1–28:23. https://doi.org/10.1145/3090093

[81] Kristin Williams, Karyn Moffatt, Denise McCall, and Leah Findlater. 2015. Designing Conversation Cues on a Head-Worn Display to Support Persons with Aphasia. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 231–240. https://doi.org/10.1145/2702123.2702484

[82] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[83] Shengdong Zhao, Pierre Dragicevic, Mark Chignell, Ravin Balakrishnan, and Patrick Baudisch. 2007. Earpod: Eyes-Free Menu Selection Using Touch Input and Reactive Audio Feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI ’07)*. Association for Computing Machinery, New York, NY, USA, 1395–1404. https://doi.org/10.1145/1240624.1240836

A *STUDY 1*

A.1 Samples of IoT Manipulation Task

Table 5 lists task samples we used in *study 1*.

A.2 Stimuli in *study 1*

Figure 12 shows the stimuli of study 1’s experiment.

Table 5: Samples of *IoT Tasks* (CI = *Checking Info*, DM = *Discrete Manipulation*, CM = *Continuous Manipulation*, SL = *Selecting From List*).

IoT Task	Sample instructions
CI	Check whether “Light 2” is On in the Living Room Check whether “Dishwasher” is Off in the Kitchen
DM	Turn On the “Coffee Machine” in the Kitchen Turn Off the “Top Light” in the Kitchen
CM	Raise the Temperature of “AC” Above 27 in the Living Room Decrease the Brightness of “Top Light” Below 20 in the Kitchen
SL	Play Taylor Swift’s “Willow” in the Living Room Play Justin Bieber’s “Intentions” in the Living Room

A.3 Measures in *study 1*

Table 6 indicates participants’ mean performance (‘mean (sd)’) related to the quality of conversation and IoT manipulation.

B *STUDY 2*

B.1 Measures in *study 2*

Table 7 and Table 8 show the quality of conversation during *study 2*. Table 9 shows the subjective ratings on quality of IoT manipulation.

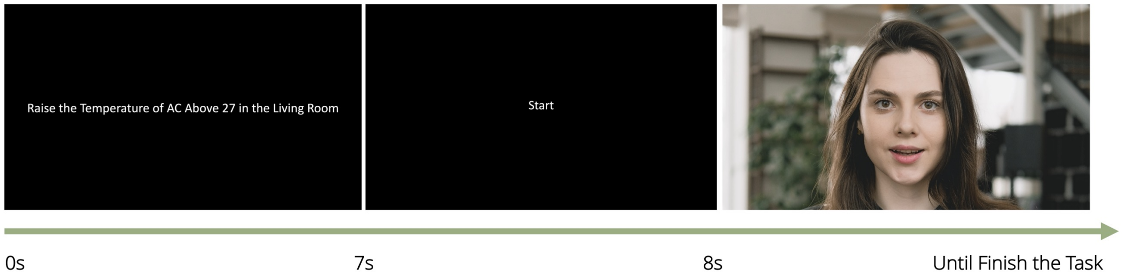


Figure 12: The instruction procedure of one manipulation trial

Table 6: Measures in simulated conversation setting (N = 20). The first column represent the *Interface-IoT Task* combination using the first letters of each (C = *ParaGlassMenu*, L = *Linear*, P = *Phone*, V = *Voice*; CI = *Checking Info*, DM = *Discrete Manipulation*, CM = *Continuous Manipulation*, SL = *Selecting From List*, Avg = Average across all *IoT Tasks*). Highlighted text in each column indicates the best average value across *Interfaces*. Note: *SUS* is measured for *Interface*.

	Quality of conversation				Quality of IoT manipulation			
	Face Focus	*Politeness*	*Naturalness*	*RTLX*	*Task Duration*	*Task Accuracy*	*Relaxation*	*SUS*
C-CI	0.275 (0.104)	5.55 (1.317)	5.15 (0.988)	21.733 (11.832)	3.834 (1.062)	0.988 (0.056)	5.65 (1.089)	
C-DM	0.281 (0.121)	5.55 (1.317)	5.35 (1.089)	20.733 (14.180)	5.075 (1.255)	1.000 (0.000)	5.80 (1.056)	
C-CM	0.252 (0.123)	5.45 (1.468)	5.20 (1.152)	24.192 (16.563)	5.686 (1.980)	1.000 (0.000)	5.80 (1.105)	
C-SL	0.133 (0.050)	5.50 (1.357)	5.20 (1.005)	22.250 (15.280)	8.398 (1.823)	1.000 (0.000)	5.45 (1.317)	
C Avg	0.235 (0.119)	5.51 (1.12)	5.23 (1.04)	22.23 (14.34)	5.75 (2.28)	0.997 (0.028)	5.68 (1.13)	83.00 (9.82)
L-CI	0.129 (0.142)	5.50 (1.192)	5.35 (0.933)	23.592 (13.067)	4.772 (1.881)	0.988 (0.056)	5.30 (1.129)	
L-DM	0.183 (0.203)	5.40 (1.046)	4.95 (1.317)	23.400 (13.469)	5.685 (1.552)	0.988 (0.056)	5.50 (1.100)	
L-CM	0.157 (0.161)	5.40 (1.142)	5.15 (1.089)	22.900 (14.040)	6.016 (1.230)	1.000 (0.000)	5.60 (0.883)	
L-SL	0.088 (0.102)	5.20 (1.152)	5.10 (0.912)	26.275 (16.767)	10.735 (2.203)	0.988 (0.056)	5.25 (1.118)	
L Avg	0.139 (0.157)	5.38 (1.12)	5.14 (1.06)	24.04 (14.20)	6.80 (2.90)	0.991 (0.048)	5.41 (1.05)	81.75 (10.04)
P-CI	0.030 (0.022)	3.60 (1.818)	4.05 (1.468)	25.542 (15.107)	8.777 (1.362)	0.988 (0.056)	5.75 (0.967)	
P-DM	0.061 (0.068)	3.60 (1.875)	4.30 (1.625)	23.833 (14.854)	9.839 (1.649)	1.000 (0.000)	5.50 (0.946)	
P-CM	0.045 (0.031)	3.75 (1.943)	4.20 (1.609)	23.267 (15.751)	10.547 (2.108)	1.000 (0.000)	5.60 (0.940)	
P-SL	0.040 (0.034)	3.95 (1.820)	4.50 (1.504)	20.983 (13.689)	10.874 (1.905)	0.988 (0.056)	5.90 (0.788)	
P Avg	0.044 (0.043)	3.73 (1.84)	4.26 (1.53)	23.41 (14.68)	10.01 (1.92)	0.994 (0.039)	5.69 (0.91)	81.25 (13.32)
V-CI	0.237 (0.164)	3.60 (1.698)	4.35 (1.348)	32.708 (18.492)	16.064 (5.545)	0.718 (0.204)	4.45 (1.731)	
V-DM	0.316 (0.216)	3.90 (1.774)	5.10 (1.210)	23.733 (19.321)	9.306 (1.949)	0.968 (0.103)	5.35 (1.348)	
V-CM	0.224 (0.147)	3.70 (1.838)	4.35 (1.348)	28.500 (19.812)	13.947 (2.857)	0.887 (0.128)	4.40 (1.875)	
V-SL	0.235 (0.229)	4.15 (1.755)	5.00 (1.170)	25.750 (22.488)	17.398 (6.939)	0.802 (0.183)	5.00 (1.686)	
V Avg	0.253 (0.192)	3.84 (1.75)	4.70 (1.30)	27.67 (19.99)	14.18 (5.60)	0.844 (0.183)	4.80 (1.69)	70.88 (18.84)

Table 7: Conversation quality ratings by hosts for both *IoT* and *No_IoT* conditions.

	AC1		AC2		AC3		NB1		NB2		IO1	
	IoT	*No_IoT*	*IoT*	*No_IoT*	*IoT*	*No_IoT*	*IoT*	*No_IoT*	*IoT*	*No_IoT*	*IoT*	*No_IoT*
Mean	6.417	6.667	5.750	6.667	6.333	6.417	5.833	6.167	6.167	6.250	3.000	4.333
SD	0.669	0.651	0.886	0.651	0.888	0.669	0.718	0.577	0.718	0.622	1.537	1.923
Median	6.500	7.000	6.000	7.000	6.500	6.500	6.000	6.000	6.000	6.000	3.000	5.000
25th percentile	6.000	6.750	5.000	6.750	6.000	6.000	5.000	6.000	6.000	6.000	1.750	2.750
75th percentile	7.000	7.000	6.000	7.000	7.000	7.000	6.000	6.250	7.000	7.000	4.250	6.000
Wilcoxon (p-value)	$p = 0.074$		$p = 0.005$		$p = 0.386$		$p = 0.065$		$p = 0.445$		$p = 0.010$	
Wilcoxon (Z score)	$Z = -1.604$		$Z = -2.521$		$Z = -0.535$		$Z = -1.468$		$Z = -0.270$		$Z = -2.369$	

Table 8: Conversation quality ratings by guests for both *IoT* and *No_IoT* conditions.

	AC1		AC2		AC3		EC1		NB1		NB2		IO1	
	IoT	*No_IoT*	*IoT*	*No_IoT*	*IoT*	*No_IoT*	*IoT*	*No_IoT*	*IoT*	*No_IoT*	*IoT*	*No_IoT*	*IoT*	*No_IoT*
Mean	6.583	6.833	5.583	6.750	5.667	6.750	5.500	5.417	5.667	6.583	6.500	6.583	4.250	4.917
SD	0.669	0.389	0.996	0.622	1.155	0.622	1.314	1.311	1.073	0.793	0.798	0.669	1.712	1.443
Median	7.000	7.000	6.000	7.000	6.000	7.000	6.000	6.000	6.000	7.000	7.000	7.000	4.500	4.500
25th percentile	6.000	7.000	5.000	7.000	5.750	7.000	5.000	5.000	5.000	6.750	6.000	6.000	2.750	4.000
75th percentile	7.000	7.000	6.000	7.000	6.000	7.000	6.000	6.000	6.250	7.000	7.000	7.000	5.250	6.000
Wilcoxon (p-value)	$p = 0.117$		$p = 0.006$		$p = 0.007$		$p = 0.528$		$p = 0.031$		$p = 0.425$		$p = 0.084$	
Wilcoxon (Z score)	$Z = -1.214$		$Z = -2.497$		$Z = -2.395$		$Z = 0.000$		$Z = -1.886$		$Z = -0.365$		$Z = -1.437$	

Table 9: Quality of IoT manipulation ratings by hosts (H) and guests (G).

	[H] Relaxation	[H] Politeness	[H] Interruption	[H] SUS	[G] Interruption	[G] Politeness	[G] Hospitality
Mean	6.000	5.583	2.417	84.583	3.000	5.750	6.583
Std. Deviation	1.279	1.730	1.379	8.450	0.953	1.215	0.669
Median	6.000	6.000	2.000	87.50	3.000	6.000	7.000
25th percentile	6.000	4.750	1.750	79.375	2.00	5.750	6.000
75th percentile	7.000	7.000	3.000	90.000	3.250	6.250	7.000