



# WorkFit: Designing Proactive Voice Assistance for the Health and Well-Being of Knowledge Workers

**Shashank Ahire**  
Human-Computer Interaction  
Leibniz University Hannover  
Hannover, Germany  
shashank.ahire@hci.uni-hannover.de

**Benjamin Simon**  
Human-Computer Interaction  
Leibniz University Hannover  
Hannover, Germany  
bsimon@hci.uni-hannover.de

**Michael Rohs**  
Human-Computer Interaction  
Leibniz University Hannover  
Hannover, Germany  
michael.rohs@hci.uni-hannover.de

## ABSTRACT

Prior research has designed and evaluated Voice Assistance (VA) for different settings such as the home, school, and public spaces. Office environments have been relatively understudied, leaving a gap in understanding the essential factors for designing a VA specifically for work settings. In this study, we developed the *WorkFit* VA specific for the office environment, focusing on the health and well-being of knowledge workers. WorkFit was designed to monitor knowledge workers for sedentary behavior, inconsistent hydration, and stress, and to deliver proactive voice interventions followed by a health recommendation to mitigate those issues. We evaluated WorkFit in a field study with 15 knowledge workers for 5 working days. In the study, we determined challenges and opportunities for voice interactions in work settings. We identified contextual factors for identifying inopportune moments for voice interactions in an office setting. We found that 92% of knowledge workers accepted WorkFit's hydration interventions while 79% of them engaged in walking breaks. Moreover, breathing exercises recommended by WorkFit significantly stabilized the heart rate of knowledge workers during stress. Based on our findings, we propose five design recommendations for the development of VA customized to office settings.

## CCS CONCEPTS

• **Human-centered computing** → **Sound-based input / output.**

## KEYWORDS

voice assistant, conversational user interface, office, proactive, ubiquitous, health, well-being, knowledge worker

## ACM Reference Format:

Shashank Ahire, Benjamin Simon, and Michael Rohs. 2024. WorkFit: Designing Proactive Voice Assistance for the Health and Well-Being of Knowledge Workers. In *ACM Conversational User Interfaces 2024 (CUI '24)*, July 8–10, 2024, Luxembourg, Luxembourg. ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3640794.3665561>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
CUI '24, July 8–10, 2024, Luxembourg, Luxembourg  
© 2024 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-0511-3/24/07.  
<https://doi.org/10.1145/3640794.3665561>

## 1 INTRODUCTION

VA has become pervasive, comprising numerous applications, such as search, shopping, social media, and travel, that offer varying forms of conversational interactions. Leveraging Large Language Models (LLMs), voice assistance exhibits human-like qualities, appearing more persuasive, empathetic, and intelligent. As VA expands across different contextual settings, including homes [1, 31], schools [22], in-vehicle [46] and public spaces [29, 32], it is also essential to study how VA can be integrated into the office setting.

Knowledge workers are often involved in tasks demanding substantial cognitive effort within their fields. As a result, it becomes crucial for VA to take into account the availability of knowledge workers and respond appropriately. Developing VA in work environments is challenging due to the formal nature of work settings. This formality presents a challenge for VA in terms of understanding and following workplace rules when intervening.

Conversational user interfaces (CUIs), comprising text, audio, and virtual assistance have demonstrated potential in promoting physical activity [5, 43] and mental well-being [10, 12]. CUIs can help to improve a person's health and well-being and their likelihood of reaching a desired health-related goal. Previous literature has found that knowledge workers desire assistance to support them for health and well-being at the workplace [2]. As office workers often spend the majority of their workday sitting [28], encouraging physically active breaks is particularly important for them. Moreover, work environments have been associated with health issues like hypertension and heart disease [35]. Additionally, studies have highlighted the adverse effects of dehydration on cognitive functions and mood during work [21]. Hence, it is crucial for knowledge workers to engage in physical activity, manage stress, and maintain consistent hydration during work hours. This presents an opportunity for proactive VA to assist knowledge workers in maintaining their health and well-being during work.

To investigate the usability of VA in office settings and gain insights for developing VA suitable for the office use, we created the WorkFit VA. It delivers proactive voice interventions to remind knowledge workers about sedentary behavior, hydration, and stress. WorkFit continuously monitors sedentary time, water intake, and heart rate, triggering voice interventions upon detecting irregularities. Additionally, it provides health recommendations to address these issues. In a field study, we assessed WorkFit's performance with 15 knowledge workers over a period of 5 days (40 hours per participant) within their workplace.

During the assessment, we found the challenges in designing VA for office settings include lack of agency, identifying inopportune moments, ensuring privacy, and maintaining non-intrusiveness

towards colleagues. On the positive side, deploying VAs in office settings demonstrated benefits such as awareness and reflection on unhealthy behavior, effectively capturing user attention, and facilitating dual-task interactions. Additionally, we identified four essential factors influencing the determination of inopportune moments in office settings: emotional and cognitive availability, social context, shared locations, and user activity. Of the health recommendations delivered by WorkFit, 92% of the hydration interventions and 79% of the break suggestions were accepted by knowledge workers. The '4-7-8' breathing exercise, recommended by WorkFit, was able to ameliorate high heart rates in 30 s and low heart rates in 10 s. Based on our findings, we put forth five design recommendations: empowering users with agency, giving precedence to user privacy during interaction, prioritizing interventions, adhering to office etiquette, and considering the user's social context, location, and availability.

In this study, our contribution is as follows: (1) We present the WorkFit system, which provides VA for the health and well-being of knowledge workers in office settings. (2) We determine the challenges and advantages of deploying VA in the office environment. (3) We identify four contextual factors that are essential for delivering proactive voice interventions within the office context. (4) We investigate whether our health recommendations were effective in mitigating sedentary behavior, stress, and inconsistent hydration of knowledge workers. (5) Lastly, we propose five design recommendations for developing VA for work settings.

## 2 RELATED WORK

We review papers on proactive interventions and opportune moments in CUIs. Further, we discuss how CUIs have been used for health and well-being.

### 2.1 Proactive Interventions and Opportune Moments in CUIs

Proactive interventions are most effective when delivered at opportune moments, enhancing user engagement and avoiding user dissatisfaction. These favorable instances are termed "opportune moments" [7, 42]. Conversely, instances that may lead to user annoyance and result in minimal or no engagement are identified as "inopportune moments." In formal environments like offices, identifying inopportune moments is vital as proactive interventions can prove costly for users. Hence, it is crucial to refrain from delivering voice interventions during inopportune moments to prevent user annoyance.

Proactive health reminders have proven beneficial for urgent and critical issues [5, 36, 43]. However, proactive interruptions can also create problems in multi-party conversations, informal meetings, and social meetings. Proactive smart speakers have been applied successfully in the domain of healthcare. Active reminders via smart speakers, based on different types of triggers, like time and health sensor data have helped users.

In a study, Reicherts et al. [31] investigated the factors that influence user perception of proactive smart speakers. They found that users felt inappropriate when conversational agents interrupted an ongoing conversation. They also found that the timing and content of proactive interactions were important factors in determining

users attitudes towards them. Additionally, the frequency of proactive interactions was found to have a negative impact on users attitudes.

To estimate the opportune moments for voice information service without disturbing users activity Komori et al. [16] developed a prototype that could detect user activity transition based on user location, body movement and ask them about the acceptability of notification at a particular moment. They found that notifications are more acceptable with participants are at bed, while eating, performing household chores, and while working. Similarly, to determine opportune moments for proactive conversational interactions in domestic settings Cha et al. [8] conducted a study with 40 participants. They concluded that opportune moments are dependent on user busyness, complexity of the primary task, user mood, and social availability. They found that opportune moments for smart speaker interaction are those when users are engaged in repetitive activities that require low attention.

With the aim of investigating interaction errors in proactive smart speakers, Wei et al. [45] used the experience sampling method to identify the types of errors that occur in proactive smart-speaker interactions, as well as to find strategies that users apply to tackle those errors. They found that interaction errors can negatively impact the user experience. Further, interaction errors increase the user burden and discourage them from performing further interactions. In another study, Wei et al. [44] found that the availability of participants for a conversation was higher when they were performing entertainment tasks rather than work or study tasks. The self-reported score of boredom and mood were found significantly correlated with participant's availability.

Zargham et al. [48] depicted storyboards of scenarios to users to understand the user perception of proactive smart speaker interventions in everyday situations. The authors considered eight day-to-day scenarios for the interviews. Participants stated that proactive interventions related to health are valuable to them.

### 2.2 CUIs for Supporting Health and Well-Being

Conversational agents have been investigated in workplace reflection and well-being. Kimani et al. [13] introduced the 'Amber' CA, specifically designed to promote well-being and productivity among knowledge workers. This CA effectively supports multiple work-related objectives, such as providing break reminders, minimizing social media distractions, and facilitating task reflection. Similarly, Ahire et al. [2] identified knowledge workers expectations regarding conversational agents. Knowledge worker desired a CA that could assist them for health-oriented breaks and prevent distractions. The authors developed the 'Ubiquitous Work Assistant', comprising a stationary CA intended for the user's work desk and a wearable CA fixed to the user's wrist.

The 'Woebot' [10] chatbot demonstrated its efficacy as a text-based conversational agent by supporting cognitive therapeutic processes and reducing depression and anxiety among students. Lee et al. [18] investigated the role of chatbots in fostering self-compassion. Their study resulted in increased self-compassion through chatbot interaction. While previous studies have explored the use of CAs in office settings, it is worth noting that Kocielnik et al. [14] specifically examined the use of a voice interface in an office context.

However, the authors permitted participants to interact with the voice interface in a quiet room or away from their desk. Therefore, it is unclear how proactive health-oriented voice interventions perform in an office environment during work tasks.

### 3 WORKFIT DESIGN AND DEVELOPMENT

While VA has proven beneficial in delivering health reminders across various contexts and user groups, its suitability for delivering such reminders in office settings of knowledge workers remains uncertain. Thus, our research aims to address the following key questions:

- **RQ1:** *What challenges and benefits are associated with the deployment of VA in an office environment?*  
To develop VA that is adapted to office settings, it is important to understand the distinctive features that VA offers, compared to other modalities. Furthermore, it is essential to identify issues that arise when deploying VA in office settings.
- **RQ2:** *Contextual Factors for Determining Inopportune Moments for Voice Assistance?*  
Delivering proactive interventions during inopportune moments can incur high error costs for users. To avoid this, it is crucial to identify contextual factors that play a role in determining inopportune moments. These factors should be taken into account when delivering proactive voice interventions. Previous literature has shown that not taking into account contextual factors can lead to user disengagement and frustration [4, 26]. However, considering contextual factors can help VA to deliver interventions that are timely, relevant, and actionable.
- **RQ3:** *Did WorkFit's health recommendations prove effective in mitigating health concerns and encouraging healthy behavior?*  
Studying the effectiveness of VA in delivering health-based interventions within office settings is crucial. It is essential to determine whether knowledge workers actively engage with voice recommendations or if they ignore them. Moreover, it is important to investigate whether the health recommendations by WorkFit effectively foster behavior towards improving health and well-being.

#### 3.1 Design Choices

We systematically and purposefully made decisions that align with our research questions.

**3.1.1 Health and Well-Being Apps.** Previous studies have shown that knowledge workers express a desire for VA to assist them in managing health and well-being issues [2]. We hypothesized that knowledge workers would be more likely to assess an application designed to support their health and well-being specifically in the context of their office work environment. Drawing insights from existing literature, we identified three prevalent health concerns among knowledge workers at their workplaces: sedentary behavior [33, 40], hydration [2, 27, 39], and stress [9, 47].

*Coping exercise with a defined goal.* Existing research highlights the success of goal-based interventions in motivating users to engage in activities [6]. Our objective was twofold: not only to detect

and deliver health and well-being interventions but also to suggest coping exercises to address these issues. We chose to suggest coping activity as a health recommendation for each intervention. These coping activities need to be actionable and feasible within the work environment, enabling easy execution for knowledge workers. Furthermore, each suggested activity should include a time-based or performance-oriented goal.

*Stress.* Controlled slow breathing exercises, like the '4-7-8 technique', offer benefits for relaxation, heart rate regulation, and blood pressure management [20, 34]. This technique, also known as *pranayama*, is an ancient yogic practice that triggers the parasympathetic state, promoting rest and relaxation. The exercise involves three steps: (1) Inhale for 4 seconds, (2) hold breath for 7 seconds, and (3) exhale for 8 seconds. To ensure sufficient impact, we recommended performing the '4-7-8' breathing exercise for a duration of 2 minutes.

*Sedentary behavior.* Walking offers substantial long-term health benefits, especially after prolonged periods of sitting [40]. As a simple and effortless activity, walking is particularly suitable for knowledge workers in office settings. To accommodate the office environment's constraints, we suggested a goal of minimum of 20 steps per sedentary break.

*Inconsistent Hydration.* To tackle inconsistent hydration, we requested knowledge workers to drink a cup of water during each intervention. A daily intake of 15.5 cups of water is recommended [11, 15], with each cup being about 250 ml. To achieve this goal during the active sixteen hours of the day (excluding eight hours of sleep), a rate of one cup per hour is advisable, hence our reminder for one cup every 45-minutes.

**3.1.2 Conversation Design.** When designing the conversational strategy of WorkFit we decided to adhere to the following three principles: (1) *User-centered dialogue:* Our aim was to consider a user-centered approach and allow knowledge workers to flexibly act upon the interventions. When occupied with a task, it should be possible for the user to decline the given health recommendation and to postpone the recommendations. (2) *Reflect:* Moreover, our aim was to make users reflect on their unhealthy behavior when they declined a recommendation. During the reflective conversation, WorkFit should highlight the importance of healthy behavior. (3) *Motivate:* While performing the health recommendation, WorkFit should update the knowledge worker about their progress and give a motivational message regarding their goal.

**3.1.3 Smartwatch.** We chose to use smartwatches due to their health monitoring and interaction capabilities, driven by factors such as: (1) *Availability of physiological sensors:* Smartwatches offer real-time physiological measurements through built-in sensors. (2) *Proximity:* The close proximity of a smartwatch to its user enhances its efficacy in capturing attention and increasing user engagement. (3) *Multi-modality:* Smartwatches enable multi-modal interactions, including graphical, spoken, and tactile output, allowing users to engage through diverse modalities. For the development and deployment of WorkFit, we selected Samsung Galaxy Watch4<sup>1</sup>

<sup>1</sup><https://www.samsung.com/global/galaxy/galaxy-watch4/specs/>

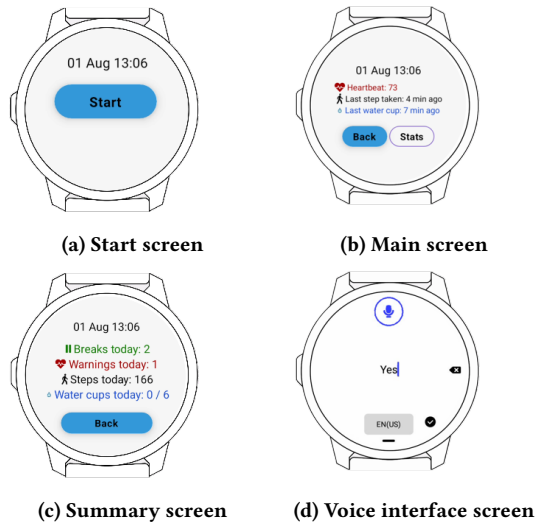


Figure 1: Screens of the WorkFit graphical interface.

### 3.2 Development

We developed WorkFit as an Android app for the smartwatch. For accurate monitoring, WorkFit accesses the heart rate and accelerometer sensors of the smartwatch. WorkFit’s graphical interface comprises four screens: Start (Figure 1a), Main (Figure 1b), Summary (Figure 1c), and Voice Interface (Figure 1d). At the beginning of the work day, the knowledge worker presses the start button on the Start screen (Figure 1a), which activates the timers for sedentary behavior and water intake. The Main screen (Figure 1b) displays key data: ‘Heartbeat’ (updated every second), ‘Last step taken,’ and ‘Last water cup,’ which are updated every minute. This screen also features ‘Back’ and ‘Stats’ buttons. ‘Back’ concludes the work day, while ‘Stats’ provides overall work day details.

The Summary screen (Figure 1c) shows daily statistics: ‘Breaks today,’ ‘Warnings today,’ ‘Steps today,’ and ‘Water cups today.’ ‘Breaks today’ presents the number of breaks taken, ‘Warnings today’ displays number of heart rate warnings, ‘Steps today’ shows the footstep count, and ‘Water cups today’ reveals the number of water cups consumed. A back button returns to the main screen.

The voice interface screen (Figure 1d) appears when voice input is needed. Knowledge workers interact with WorkFit using this screen. The interface has basic buttons: a backspace button for fixing mistakes, a send button to confirm, and a microphone icon for voice input. We adhere to Google’s Conversational Design guidelines<sup>2</sup> and the guidelines of Murad et al. [23] to create voice-based conversations.

**3.2.1 Sedentary Behavior Intervention.** To automatically track sedentary behavior of the knowledge worker, WorkFit relies on accelerometer sensor readings using class ‘TYPE\_STEP\_DETECTOR’<sup>3</sup>. WorkFit monitors activity and inactivity of the knowledge worker based on the readings provided by the step detector. The ‘Last step taken’

<sup>2</sup><https://developers.google.com/assistant/interactivecanvas/design>

<sup>3</sup>[https://developer.android.com/reference/android/hardware/Sensor/#TYPE\\_STEP\\_DETECTOR](https://developer.android.com/reference/android/hardware/Sensor/#TYPE_STEP_DETECTOR)

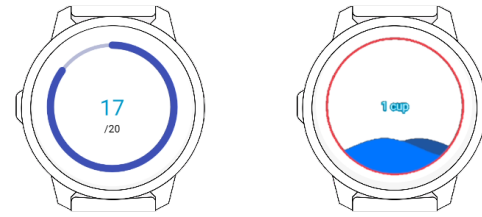


Figure 2: Visualization of step count and water intake animation.

Figure 2: Visualization of step count and water intake animation.

timer starts when WorkFit detects user inactivity. This timer increments until a footstep is detected by the step detector (as shown in Figure 1b). After 60 minutes, WorkFit initiates a voice-based intervention, as described in Table 1. When the user confirms the break, WorkFit prompts the user to take 20 steps (Figure 2a). If a footstep is detected before achieving 60-minute sedentary counter, the timer resets automatically.

Table 1: Example voice interaction for sedentary break intervention.

#### Scenario 1: User Takes a Break

**WorkFit (Voice):** *You haven’t moved for a while. How about taking a break? Say yes to confirm.*

**User:** *Yes.*

**WorkFit (GUI):** *Displays the number of steps performed by the user and the goal of 20 steps (see Figure 2a).*

**WorkFit (Voice, after 10 steps):** *You have completed half of the steps. Keep moving.*

**WorkFit (Voice, after 20 steps):** *Well done! You completed the 20 steps.*

#### Scenario 2: User Delays or Declines the Break

**WorkFit (Voice):** *You haven’t moved for a while. How about taking a break? Say yes to confirm.*

**User:** *No.*

**WorkFit (Voice):** *Do you want to delay your walking break? If yes, for how many minutes? Otherwise, please say No.*

**User (Positive answer):** *10 minutes.*

**User (Negative answer):** *No.*

**WorkFit (Voice):** *It is important to take breaks!*

Following our conversation design principles, in the sedentary behavior intervention conversation (Table 1), we offered knowledge workers a choice to accept, delay, or decline the health recommendation. If the knowledge worker would like to postpone the break, then WorkFit asks for waiting time (in minutes) before it delivers the intervention again. If the knowledge worker declines the interventions, WorkFit gives information that allows the worker to reflect on their sedentary behavior. It does so by highlighting

the importance of taking breaks. If the knowledge worker accepts the intervention and starts performing the health recommendation by walking, then after completing 10 steps, WorkFit VA delivers a motivational message and requests to keep going. After completing the 20 steps, it will update the user about the completion of the exercise.

**3.2.2 Hydration Intervention.** WorkFit triggers a hydration voice intervention every 45 minutes. It begins with a voice reminder (as shown in Table 2), followed by a request message to confirm after two minutes. If the knowledge worker cannot drink water immediately, perhaps due to an empty bottle or due to being away from their desk. In this case, WorkFit gives them an option to postpone the intervention for a specific time. If the knowledge worker declines to postpone the intervention, it will deliver a reflective message highlighting the importance of drinking water. On the other hand, if the user accepts the intervention and drinks a cup of water, WorkFit acknowledges the user and reports the total water intake for the day (see Figure 2b).

**Table 2: Voice interaction for hydration intervention**

---

**Scenario 1: User Drinks Water**

---

**WorkFit (Voice):** *It's time to drink a cup of water.*  
**WorkFit (Voice, after 2 minutes):** *Did you drink a cup of water?*  
**User:** *Yes.*  
**WorkFit (Voice):** *Well done! You've already had x glasses of water. (WorkFit plays water animation, see Figure 2b).*

---

**Scenario 2: User Delays Drinking**

---

**WorkFit (Voice):** *It's time to drink a cup of water*  
**WorkFit (Voice, after 2 minutes):** *Did you drink a cup of water?*  
**User:** *No.*  
**WorkFit (Voice):** *Do you want to delay your drinking break? If yes, for how many minutes? Otherwise, please say no.*  
**User (Positive answer):** *10 minutes.*  
**WorkFit (Voice):** *Your water drinking break has been delayed by 10 minutes.*  
**WorkFit (Voice, 10 minutes later):** *Would you like to drink water now?*  
**WorkFit (Voice, 2 minutes later):** *If you drank a cup of water, please say yes.*  
**User:** *Yes.*  
**User (Negative answer):** *No.*  
**WorkFit (Voice, if no):** *It is important to drink water while working!*

---

**3.2.3 Stress Intervention.** To monitor the knowledge worker's heart rate, we use the heart rate sensor available in the smartwatch. We fetched heart rate reading using the class 'TYPE\_HEART\_RATE'<sup>4</sup>. We monitored the heart rate reading every second and displayed it

on the interface (see Figure 1b.) A standard resting heart rate for adults typically ranges between 60 and 100 bpm [19, 25]. WorkFit activates the 'high heart rate' warning when the heart rate exceeds 105 bpm. Correspondingly, the 'low heart rate' warning appears when the heart rate falls below 55 bpm. When any of these warnings are triggered, users are prompted to perform the '4-7-8 breathing exercise' to regulate their heart rate back to a healthy resting range.

Following our conversation design principles, when users confirm their intent to perform the breathing exercise, WorkFit initiates the guidance process (Table 3). It is depicted by an animation consisting of a contracting and expanding bubbles, replicating the inhale and exhale of breathing (Figure 3). Throughout the voice guidance, graphical cues are provided as follows: during the 4-second inhale phase, the label 'Inhale' (Figure 3a) is displayed. During the 7-second breath-hold phase, the label 'Hold' (Figure 3b) is displayed. While exhaling, the label 'Exhale' (Figure 3c) is shown. The complete exercise comprises two cycles of these stages. The whole breathing exercise lasts 120 seconds. WorkFit initiates the conversation concerning the breathing exercise upon the occurrence of high or low heart rate warnings, as illustrated in Table 3. If the knowledge worker declines performing the breathing exercise, it provide the option of postponing the breathing exercise.

**Table 3: Voice interaction during high/low heart rate warning**

---

**Scenario 1: User Accepts the Breathing Exercise**

---

**WorkFit (Voice):** *Your heartbeat is too high/low. Would you like to perform a breathing exercise? If so, please say yes.*  
**User:** *Yes.*  
**WorkFit (Voice):** *The name of this breathing technique is '4-7-8.' This exercise takes two minutes.*  
**WorkFit (Voice):** *The exercise starts now. Breathe in through your nose for 4 seconds. (The animation for inhale is shown, see Figure 3a.)*  
**WorkFit (Voice):** *(Counts down 4 seconds.)*  
**WorkFit (Voice):** *Hold your breath for 7 seconds. (The animation for holding the breath is shown, see Figure 3b.)*  
**WorkFit (Voice):** *(Counts down 7 seconds.)*  
**WorkFit (Voice):** *Exhale for 8 seconds. (The breathing animation for exhale is shown, see Figure 3c.)*  
**WorkFit (Voice):** *(Counts down 8 seconds.)*  
**WorkFit (Voice):** *We completed the first cycle. Now, we will perform the second cycle of the breathing exercise.*

---

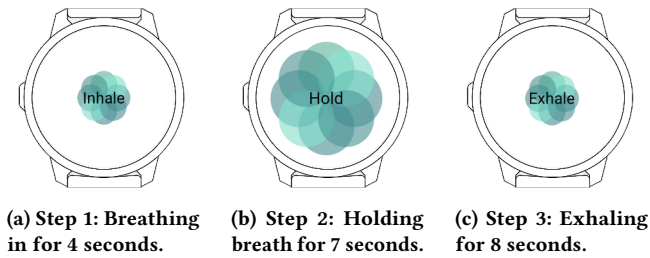
**Scenario 2: User Delays the Breathing Exercise**

---

**WorkFit (Voice):** *Your heartbeat is too high/low. Would you like to perform a breathing exercise? If so, please say yes.*  
**User:** *No.*  
**WorkFit (Voice):** *Do you want to delay your breathing exercise? If yes, for how many minutes?*  
**User:** *10 minutes.*  
**WorkFit (Voice):** *Your breathing exercise has been delayed by 10 minutes.*  
**WorkFit (After 10 minutes):** *Let's start the breathing exercise. (WorkFit starts the breathing exercise.)*

---

<sup>4</sup>[https://developer.android.com/reference/android/hardware/Sensor/#TYPE\\_HEART\\_RATE](https://developer.android.com/reference/android/hardware/Sensor/#TYPE_HEART_RATE)



**Figure 3: Graphical guidance animations for the ‘4-7-8’ breathing exercise.**

## 4 PILOT STUDY

Following the creation of the WorkFit VA, we conducted an evaluation over five working days, involving five lab members to identify technical, usability, interaction, and deployment issues. In the lab deployment, we discovered that WorkFit voice interventions interrupted lab members during their scheduled meeting.

### 4.1 Learnings from the Pilot Study

In feedback, lab members emphasized that meetings are inappropriate moments to deliver a voice intervention. Voice interventions during a meeting were regarded as unprofessional and distracting by knowledge workers. They stated that WorkFit should not interrupt them during an inopportune moment, such as in meetings. Delivering a well-being related or health-related voice message during an inopportune moment was regarded as particularly embarrassing by the participants of the pilot study. Thus, there is a high cost related for WorkFit to deliver voice messages in such moments.

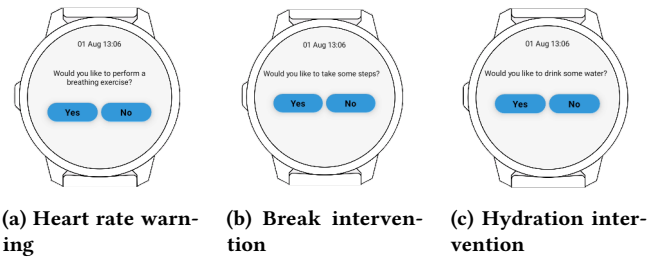
### 4.2 WorkFit Interaction at Inopportune Moments

During inopportune moments WorkFit will not deliver a proactive voice intervention. Instead, WorkFit delivers a combination of graphical user interface (GUI) and vibration notification (as shown in Figure 4). A combination of GUI and vibration intervention was designed to consume minimal user attention during an inopportune moment. To consider an inopportune moment during work, participants of the field study were asked to add an inopportune moment to their Google calendar. Before delivering a voice intervention, WorkFit checks the calendar events.

During an intervention at an inopportune moment, upon selecting ‘Yes’ as an option (Figure 4), depending on the intervention WorkFit displays the screen shown in Figure 2a or 2b to display the goal. On selecting ‘No’, the interface switches to the main screen. WorkFit will again remind the user in the next intervention. When WorkFit detects a high/low heart rate condition, it notifies the user with a warning in graphical form accompanied by a vibration. As a breathing guide, WorkFit displays the graphics only (see Figure 3).

## 5 FIELD STUDY

To assess WorkFit’s effectiveness in supporting knowledge workers within their work context, we conducted a field study. We carried



**Figure 4: Graphical plus vibration interventions during an inopportune moment.**

out a total of 75 days of field study, with each user spending 5 work days (40 hours) participating.

### 5.1 Participants

Through advertising on the university’s online notice board, WhatsApp groups, and student forums, we received 38 applications for the experiment. Selecting based on profession, voice interaction experience, concern about sedentary behavior and inconsistent hydration, and prior health monitoring experience with smartwatches or smartphones, we selected a cohort of 15 participants (11 male, 4 female). All participants were knowledge workers, working a self-reported average of 9 hours daily (SD = 5 hours). Their ages ranged from 23 to 33 years (M = 26.6, SD = 2.9). Professions included software engineers, researchers, and machine learning engineers. The participant’s information and their experience with VUIs is described in Table 4. As participants spent a number of days working from home each week, we granted them flexibility to evaluate WorkFit in both their home office and professional office settings. As compensation, participants received €25 Amazon vouchers at the study’s conclusion.

### 5.2 Method

We invited participants to the lab for the orientation session. We briefed the participants about our objectives, data collection procedures, and data processing specifics. The participants were first asked to complete a consent form and then proceeded to fill out the pre-study form (outlined in section 5.2.1). We connected the mobile phones of the participants to a Samsung Galaxy Watch4, followed by a demo and training them on interacting with WorkFit. The participants were instructed to start WorkFit at the beginning of each workday and to end it at the day’s conclusion. To identify inopportune moments, we requested participants to add placeholders in the Google calendar representing office events, such as, meeting 1, meeting 2 and so on. Finally, we requested that participants utilize WorkFit daily, while working from home or in the office.

**5.2.1 Pre-Evaluation Questionnaire.** The pre-study questionnaire was administered using an online survey to gain insights into participants backgrounds. In total, the pre-study questionnaire consisted of 15 questions focusing on age, profession, gender, work habits, voice interaction familiarity, health priority during work, break frequency, water consumption, experience with fitness tracking tools.

**Table 4: Participant Information**

Participant ID	Gender	Profession	Experience with VUIs
U01	F	Researcher	No experience with VUIs.
U02	F	Software developer	Regularly uses Alexa and Google Assistant smart speakers for various tasks.
U03	M	Coding and testing	No experience with VUIs.
U04	M	Simulation Engineer	Utilizes Google Assistant primarily for reminders and internet searches.
U05	F	Data analyst	Has beginner-level experience with VUIs.
U06	M	Researcher	Uses Alexa for entertainment purposes and Siri for hands-free interaction.
U07	M	Software Engineer	No experience with VUIs.
U08	F	Computer Engineer	Uses Google Assistant and Alexa frequently to check weather, news, and search for information.
U09	M	Researcher	No experience with VUIs.
U10	M	Researcher	Relies on Siri for setting schedules and searching for information.
U11	M	Machine learning Engineer	No experience with VUIs.
U12	M	Software engineer	Uses Google Assistant mainly for setting up reminders.
U13	M	Software engineer	Interacts with Alexa for various tasks.
U14	M	Software Developer	Uses Alexa primarily at home.
U15	M	Researcher	Occasionally uses Siri and Google Assistant for navigation.

**5.2.2 Daily Questionnaire.** Participants evaluated WorkFit over five consecutive workdays and completed a daily questionnaire at the end of each day. We created the questionnaire aiming to capture their overall daily experience, encompassing user inputs, impressions, and suggestions. The daily questionnaire facilitated the identification of their preferences, dislikes, inopportune instances, and improvement suggestions. Additionally, participants provided subjective assessments of break interventions, water intake prompts, and breathing exercises. They rated WorkFit’s utility, helpfulness, and satisfaction on a scale of 1 to 10.

**5.2.3 Post-Evaluation Questionnaire and Interview.** After the longitudinal study, participants were asked to complete a post-evaluation survey to capture their overall user experience. The survey inquired about app usage positive and negative experiences with WorkFit as a workplace conversational interface, and subjective evaluations of heart rate warnings, water intake, and break intervention accuracy based on their physiological conditions. Additionally, participants were encouraged to provide suggestions for enhancing the next version of WorkFit. The semi-structured post-evaluation interview centered on participant’s insights gathered from their responses to the daily questionnaire. On an average, the interview lasted for 45 minutes. In the interview, the participants were probed with respect to the overall user experience, daily usage, preferred features, factors for voice interaction, behavioral adaptations during interaction and contextual awareness during events

Post-study interview audio recordings were transcribed and coded. The coding process involved generating codes to capture the data’s essence and represent its concepts and ideas. Inductive thematic analysis was used to identify emerging themes. Patterns, similarities, and connections between codes were then identified to delineate distinct themes. Each theme was defined and labeled descriptively. To ensure rigor, the second and third authors conducted independent cross-checks on the identified themes [17].

## 6 RESULTS

Quantitative and qualitative findings are described with respect to the research questions stated above (section 3). We describe the challenges and benefits for designing and developing VA in work settings. We report contextual factors for identifying inopportune moments. Lastly, we focus on assessing the effectiveness of WorkFit recommendations. In a total of 75 days of field evaluation, WorkFit delivered in total 394 voice interventions. Participants assessed WorkFit in the office setting for a total of 54 days and at home for 21 days.

### 6.1 What challenges and benefits are associated with the deployment of VA in an office environment?

It is imperative to identify and address the challenges that may arise when deploying VA in restricted environments, such as office spaces. Addressing these challenges is essential for enhancing the overall usability and effectiveness of VA interactions in sensitive setting like office.

#### 6.1.1 Challenges.

*Lack of Agency Over Proactive Voice Interventions.* A few participants were annoyed by the lack of agency when the voice interactions were triggered. They desired agency over voice interventions depending on the situation, availability, and people around them. “If your colleagues are not aware of WorkFit, then it is embarrassing. It is like you didn’t put your phone on silent mode. This is considered unprofessional.” [P05] Participants desired that WorkFit should privately notify them before delivering a voice intervention. They desired a new form of interaction which could give them control and make them aware about an upcoming intervention.

In the study, some of the participants expressed a negative sentiment towards the proactive voice interventions. The instances of

these interventions occurring during opportune moments, particularly when there were no scheduled events in the participant's calendar, sometimes resulted in uncomfortable situations. This was attributed to the lack of control over proactive voice interventions to adhere to professional mannerisms in the office setting.

*Identifying Inopportune Moments.* Our investigation revealed that relying solely on calendar events falls short in effectively managing inopportune moments. While knowledge workers are typically organized and adhere to calendar events, we identified a range of contextual factors (see section 6.2) that they value for interaction considerations within WorkFit. Although minimizing inconvenient instances is beneficial, the complexity lies in addressing multiple contextual factors for each interaction. Additionally, determining these moments using various contextual factors necessitates continuous monitoring of knowledge worker activities, leading to potential privacy concerns.

Moreover, as we observed calendar events to determine inopportune moments, it might appear technically suitable to intervene the user. However, various factors, including social context, relationships with involved individuals (such as close friends, colleagues, or acquaintances), and the nature of the intervention, make it challenging to accurately predict the user's readiness and comfort for intervention.

*Managing Privacy.* Although, participants found voice interventions to be effective in gaining attention, they also criticised them with respect to privacy in the office. Some participants were unhappy, because WorkFit delivered indiscreet health warnings in a social setting in their office. The spoken health warning led to a privacy violation. The lack of encoding while delivering a health warning was a common complaint from the participants. *"It was a very obvious wording. Everyone around you knows you are stressed. It should not give personal information so obviously. It could rephrase and ask me: How about an oxygen break?"* [P09] Further, participant P02 criticised: *"If you are with colleagues, you don't want them to know that your heart rate is running high."* [P2]

*Non-Intrusive for Colleagues.* The primary medium of VA is audio, which easily captures the attention of people present in close proximity. The loud nature of audio often creates problems while interacting in a social setting. Many participants complained that voice interventions in office settings distracted their colleagues. Participants reported that some instances of voice interaction were attention-grabbing for their colleagues. For this reason many participants reduced the volume of the smartwatch. Participant P2 reported that, while working from home volume was at 80%, and while working from the office volume was at 20%. The primary reason for decreasing the volume was that the interventions should only be audible to themselves and not to their colleagues. *"To detect inopportune moments it can synchronize with my calendar, but not with my colleague's calendar."* [P14]

### 6.1.2 Benefits.

*Effective in Capturing User Attention.* Participants noted that voice interventions captured their attention effectively, particularly when they lost track of time. Compared to other notification forms like GUIs and vibration, which participants were familiar

with, voice interventions were more proficient at highlighting unhealthy behaviors. This auditory approach not only brought their awareness to these behaviors but also motivated them to follow the recommended activities. As one participant remarked, *"The voice interventions stand out in comparison to other notifications. Other notifications are just plain sound and no sentences. Audio sentences cannot be ignored and make you to perform the action."* [P09]

*Awareness, Self Check, and Reflection on Unhealthy Behavior.* WorkFit conversational interaction enhanced user awareness and encouraged reflection on their work behavior. Statements like *"You haven't moved for a while, how about taking a break?"* increased participant's awareness of sedentary patterns, motivating them to contemplate their behavior. If users opted to cancel a break, conversational statements reinforced the significance of hydration and breaks. This conversational approach effectively convinced users to adhere to the recommended activities. *"Voice interventions not only reminded me, but it also made me reflect and perform a self-check to drink water or take a break. It acted very well as a self-check mechanism. I almost got used to it, today I had to remind myself on my own."* [P10]

*Dual-Tasking and Eyes-Free Interaction.* Voice notifications should clearly convey the notification's intent, which is crucial for health and well-being notifications that require attention. Unlike GUI notifications, where users need to actively interpret the notification's intent, WorkFit's voice notifications provide the intent and action without demanding continuous monitoring by the user. This enables the knowledge workers to concentrate on their tasks and determine the feasibility of the suggested activity. WorkFit ensures essential attention, minimizing cognitive load for having to check the display. *"Unlike other healthcare applications, WorkFit didn't expect me to check the app continuously."* [P03]

## 6.2 Contextual Factors for Determining Inopportune Moments for Voice Assistance?

We exclusively used calendar events to identify inopportune moments when users would prefer to avoid voice interactions. Nonetheless, there might be several instances in the work environment where users would rather not engage in a voice interaction. To comprehend the contextual factors related to these situations, we requested participants to record such inopportune moments along with their contextual details in our daily survey. Thematic analysis revealed four contextual factors that knowledge workers expected WorkFit to consider prior to initiating interactions. Here are the four identified factors that are relevant for inopportune moments in both office and home settings:

*6.2.1 Emotional and Cognitive Availability.* Serendipitous meetings are spontaneous and could happen at any location in the office setting such as in the office corridor, in the kitchen, at the cubicle. Participants encountered serendipitous meeting scenarios, such as colleagues or their boss reaching them at their desk, coincidental encounters in the corridor, or students or colleagues reaching them. During an on-going engagement with their colleagues, it is difficult for knowledge workers to interact with WorkFit. Therefore, they are cognitively unavailable at those moments. Further, sometimes during those situations knowledge workers could also be going

through emotions such as stress, or feeling nervous or tense. “I was in an informal meeting with my colleague at my desk, which was not scheduled in the calendar. I was discussing about problems I have been encountering in my project, so I was bit stressed. Suddenly it said: You seem to be tensed, would you like to perform a breathing exercise? It broke my flow, I was disrupted in my work conversation.” [P01]

**6.2.2 Social Context.** Participants mentioned that they desired WorkFit to recognize and adapt to social scenarios in their office. Although social moments are informal, participants preferred to refrain from engaging with voice interventions in such moments. “Meetings, lunch break, when I am not at my work station. It should not serve me any voice interventions.” [P09] During the social event, delivering a voice intervention could lead to unintended attention from colleagues. In addition to this, delivering information by voice may lead to privacy issues.

**6.2.3 Shared Spaces.** Participants stated that they would abstain from having a voice interaction at some locations in their office. These locations are voice sensitive. A voice intervention, even at a low volume, is bound to catch attention of other persons in the vicinity. Participants stated that some locations should be marked as “no voice intervention zones” and mentioned office toilets, laboratories, kitchens, elevators, and office corridors as examples. “When I am at the office kitchen or in the pantry and when I am with my colleagues or at any office common place, WorkFit should not serve me any voice interventions.” [P13]

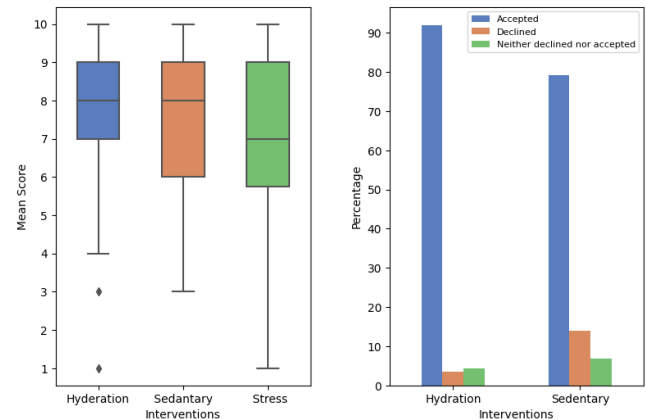
**6.2.4 Embracing Time and Work Flexibility.** The COVID-19 pandemic has forced knowledge workers to work from home but enabled them to manage their time according to their convenience. Work from home has offered greater time flexibility and convenience to knowledge workers. Due to flexibility of work time, work from home has enabled knowledge workers to balance between personal and professional time. This convenient work pattern has allowed knowledge workers to opt for relaxing breaks during work. Informal events such as power nap, family time, gym, and playful activities are now considered normal in a work day. “While working from home, I tend to sleep for sometime after lunch. WorkFit should consider my personal time.” [P10].

### 6.3 Did WorkFit’s health recommendations prove effective in mitigating health concerns and encouraging healthy behavior?

**6.3.1 Hydration Intervention.** Overall, WorkFit generated a total of 275 hydration voice interventions. Among these interventions, participants accepted 253 (92%), declined 10 (3.6%), and 12 (4.4%) interventions remained in a state of neither acceptance nor decline (see Figure 5b). The mean rating for perceived effectiveness of the hydration intervention was 7.9 (as shown in Figure 5a).

Participants indicated that the hydration interventions encouraged them to hydrate more frequently with small sips of water, in contrast to their prior habit of consuming larger quantities at once or maintaining irregular water consumption patterns. “Earlier I used to drink 400 ml in one go, after a long interval. But WorkFit helped me to divide my water consumption into smaller chunks.” [P13]. Also, participant P04 revealed that hydration interventions

were immensely beneficial, because the participant was suffering from the Gilbert syndrome<sup>5</sup>, in which an increase in the level of bilirubin leads to dehydration in the body. “I have a genetic problem of Gilbert syndrome, due to which my liver discharges lots of bil [bilirubin]. So, drinking water frequently is essential. If I don’t drink water, I get water cramps [muscle cramps caused by dehydration]. So for me it was really very beneficial” [P04]. Nevertheless, certain participants desired WorkFit to update accurate water consumption rather than assuming a standard consumption of 1 cup of water with every intervention.



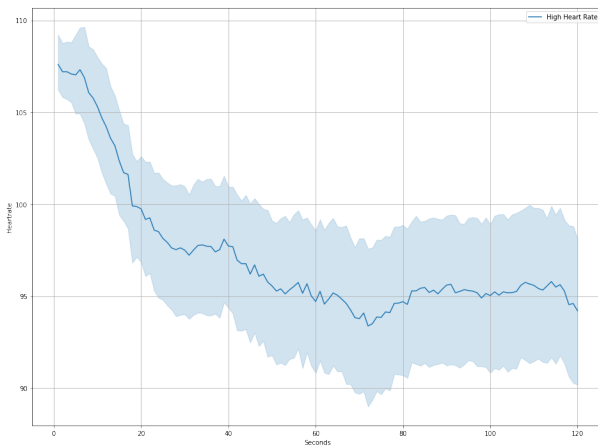
(a) Mean perceived effectiveness score of WorkFit’s health recommendations. (b) Comparison of intervention states (accepted, declined, ignored).

**Figure 5: WorkFit’s effectiveness scores and comparative analysis of the interventions.**

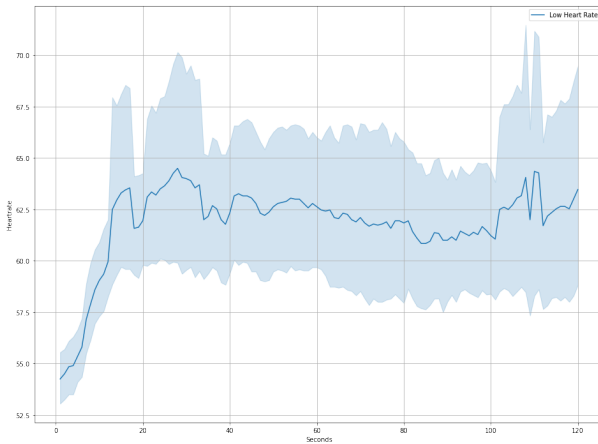
**6.3.2 Stress Intervention and Breathing Exercise.** Participants expressed their appreciation for the stress interventions as it increased their awareness of stressful moments when they experienced stress while working. Also they commended on the ‘4-7-8’ breathing technique as a coping exercise to stabilize their heart rate. Through the WorkFit real-time heart tracking (Figure 1b), the participants noticed that their heart rate returned to a more desirable level while performing the breathing exercise. In total, 56 HHR (high heart rate) warnings were received by 11 participants. In the daily survey questionnaire, participants were asked to rate if they judged that the breathing exercise helped them in ameliorating their heart rate. The question received mean rating of 7.1 with a standard deviation of 2.1.

To determine the impact of the breathing exercise, we recorded the participant’s heart rate for 120 seconds, during the breathing exercise. Figure 6a shows the percentile median graph with a 95% confidence interval of heart rate readings across all the participants for each second. The graphs exhibit a constant decline in the heart rate while performing the breathing exercise. To determine if the breathing exercise helped participants to reduce their heart rate significantly, we performed a paired t-test on the percentile value, comparing the initial 5 s (1-5) and the heart rate at 116-120 s. We

<sup>5</sup><https://medlineplus.gov/genetics/condition/gilbert-syndrome/>



(a) High heart rate (HHR) normalization graph.



(b) Low heart rate (LHR) normalization graph.

**Figure 6: Heart rate recovery graph during a breathing exercise.**

found that the heart rate was significantly reduced at the end of the breathing exercise ( $t(52) = 6.1, p < 0.01$ ). We investigated further to determine the minimum duration that the breathing exercise required to significantly reduce the heart rate. We performed paired t-tests comparing the initial 5 s to subsequent five-second blocks (6-10 s, 11-15 s, ... 116-120 s). The results show a significant difference at 26-30 s ( $t(56) = 5.4, p < 0.01$ ) and all further blocks. Hence, the '4-7-8' breathing exercise had a desirable effect on the heart rate after performing it for only 30 s.

WorkFit issued 20 Low Heart Rate (LHR) warnings. Similar to the high heart rate median percentile graph, we calculated a percentile median graph with 95% confidence interval for the low heart rate (Figure 6b). The graph depicts a consistent increase in heart rate while performing a breathing exercise. The paired t-test shows a significant difference at the end of breathing exercise ( $t(17) = 2.7, p < 0.05$ ). In fact, a significant difference was already found at 6-10 s compared to the initial heart rate ( $t(19) = 3.0, p < 0.01$ ). Thus

after performing the '4-7-8' breathing exercise for only 10 s, the participants were able to improve their heart rate.

**6.3.3 Sedentary Intervention.** In total WorkFit recommended 43 breaks to all participants, of which 79.1% (34) breaks were accepted, 14.0% (6) were declined, and 6.9% (3) were neither declined nor accepted (as illustrated in Figure 5b). During the 34 accepted breaks, the participants performed 2438 steps in total, with a mean of 71 steps per break.

Participants valued WorkFit's recommendation of 20 steps during a break, finding it achievable and effective in countering sedentary behavior, whether at home or in the office. "In the office setting, you are not expected to have long walks so I just have to do the bare minimum" [P10]. Moreover, participants appreciated the sensor-based step counting feature, recognizing and valuing WorkFit's ability to not only propose a step count goal but also ensure participant's successful achievement of this goal. "Many times I was lazy to get up, but the step counter was counting steps and displaying '0/20'. Because it was sensor-based, it could not be gamed [cheated]. So, I had to get up and walk those steps. The step counter encouraged me to get up and perform those steps." [P03]. In general, the participants noticed the influence of WorkFit interventions on their sedentary behavior. WorkFit's interventions targeting sedentary behavior motivated them to consistently take breaks and promoted physical activity.

Some participants expressed the view that a 60-minute interval between breaks was excessively long and suggested the option to personalize the time between breaks. Users indicated a preference for customizing the break reminder timer to intervals such as 45 or 50 minutes, aligning better with their individual preferences and behavior.

## 7 DISCUSSION

### 7.1 Voice Assistance at the Workplace

WorkFit's voice interaction addresses knowledge worker's well-being by offering timely reminders about sedentary behavior, eliminating the need for constant health checks. Voice alerts during critical events, like heart rate warnings, were attention grabbing. Voice-guided breathing exercises were also precise and beneficial. This is similar to findings of Kocielnik et al. [14] and Tseng et al. [41]: Conversational interactions enabled users to reflect on their unhealthy behavior and be mindful of their unhealthy work practices. Our findings add to the findings noted by Kocielnik et al. [14], which were: voice is easier for input and interactive and engaging. The challenges that we identified (Section 6.1.1) also resonate with challenges identified by Kocielnik et al. [14] and Pinder et al. [30], i.e., context detection, ethics and privacy. Similar to Reicherts et al. [31], knowledge workers found it inappropriate when WorkFit interrupted an ongoing conversation.

The use of wireless earphones is prevalent among knowledge workers during work hours. When connected to a smartwatch, utilizing earphones as the primary means for voice interaction could alleviate privacy and non-intrusiveness concerns for colleagues. While earphones may offer a convenient method for delivering voice interventions and minimizing awkward moments, challenges persist regarding submitting voice input to WorkFit. Furthermore,

even when using earphones, it remains crucial to recognize inopportune moments such as during online meetings and user's cognitive availability and giving agency to the user during proactive voice interactions.

**7.1.1 Redesigning VA: Meeting the Needs of the Workplace.** Although voice interaction was helpful in capturing attention and reflection with the proactive nature of WorkFit's voice interventions, the proactive interventions occasionally led to discomfort, as it could not be tailored to professional etiquette norms in the office environment. Participants expressed a desire for agency over voice interactions, particularly in consideration of their surroundings and the presence of colleagues. Many participants viewed the lack of control as a drawback, emphasizing the need for intervention adaptability, based on the situation and the people present. Additionally, voice interventions like heart rate warnings raised privacy concerns among participants. Although these interventions were urgent and required instant attention, some users desired new forms of proactive voice interaction, in order to give them agency and also capture their attention instantly.

We propose *adaptive VA*, i.e., which refers to a voice interface design that dynamically adjusts its interaction and behavior based on user preferences, such as the user's location, social settings, and cognitive availability. The advantages of an adaptive user interface have been established in previous studies [3, 37], highlighting enhancements in user experience and interface usability. In the case of an *adaptive VA*, it should dynamically switch between modalities (GUI, vibration and voice) based on factors such as notification type, intent, contextual elements, criticality and urgency. The *adaptive VA*, has the capability to consider multiple factors and subsequently deliver interventions in a format suitable for the current context. Unlike traditional approaches, the *adaptive VA*, eliminates the need to explicitly track whether the moment is opportune or inopportune. Instead, it tailors the intervention style based on relevant factors. Investigating the suitability of an *adaptive VA*, in an office setting holds significant value.

**7.1.2 Inopportune Moments in Work Settings.** Although it has been reported that participants would like to have health interventions proactively [48], our study found that daily health interventions should also consider opportune and inopportune moments for proactive interaction. Lastly, while working from home, the participants desired WorkFit to be able to recognize personal moments sleep, outdoor walks, and time with their kids. Like Reicherts et al. [31], we identified that participants accepted getting interrupted while at work compared to interrupting an ongoing conversation with their colleague.

In the study, we tried to tackle the problem of interruption during an inopportune moment by synchronizing the user's calendar with WorkFit interventions. Although this approach was successful, there were many inopportune moments, which user desired to be recognized by WorkFit. Considering only events placed in the calendar as inopportune moments and events not placed in the calendar as opportune moments were inadequate for successful user engagement with WorkFit. For some participants events depended on their colleagues availability, which was unpredictable for them. Also, in office setting of knowledge workers it was acceptable to walk into another colleague's cubicle/office if something urgent

needed to be discussed. Also, participants desired to avoid having a voice interaction in a social setting which could place them at the center of attention. Furthermore, participants also wanted to consider location while serving a voice intervention. For example, office pantry, common shared places and toilets. Participants also wanted WorkFit to consider the surrounding and the environment they are in.

## 7.2 Design Recommendations

Derived from our findings, we propose the subsequent design recommendations for the development of proactive VA:

**7.2.1 Empowering User Agency: Dual-Interventions.** It is widely advised in various human-computer interaction guidelines to empower users with agency or control during interactions [23, 24, 38]. However, proactive VA has no provision to enable users with agency during the interaction. In proactive interventions, users lack control over voice interactions, which can result in occasional embarrassment and annoyance, ultimately diminishing the overall user experience and making it challenging to maintain sustained user attention and engagement. Even after tracking the user's schedule, it is difficult to predict an opportune moment for the interaction. Tracking their schedule may suggest it is an opportune moment to intervene, there are multiple factors such as the presence of colleagues, location, and cognitive availability, all of which play a crucial role in determining if the moment is opportune or inopportune. In a setting like an office, which consists of formal and social interaction, it is essential to give users control to decide if the particular moment is an opportune or inopportune moment for interaction. Allowing users to decide if the moment is appropriate for interaction will lower the instances of inconvenience. We propose "dual-mode intervention" design, the VA may choose to intervene non-intrusively (ring, auditory icon, vibration, or GUI) as the primary mode, reserving voice as a secondary intervention. For example, the intervention could be presented through a GUI to convey the notification's intent, complemented by a vibration to capture the user's attention. If the user does not respond to the intervention, meaning they neither accept, reject, nor postpone it, the system subsequently may deliver a voice intervention.

**7.2.2 Delivering Encoded Voice Interventions for Privacy.** Privacy is an important aspect of an interaction. Not disclosing private information during the interaction is essential. The fundamental characteristic of voice interaction involves conveying information through audio. In VA, the inherent nature of information delivery lacks discretion, potentially drawing the attention of those near the user. Consequently, it is imperative for VA to refrain from delivering any private information during voice interventions, as this could lead to a violation of user privacy. In case of WorkFit, users did not flag a privacy concern for sedentary behavior and hydration intervention. However, they raised concerns about unambiguous stress interventions through voice. For instance, stress interventions should be discreet, potentially initiated with a vibration or GUI intervention and further, voice intervention could be in encoded form (e.g., "take some deep breaths"). Also, they were comfortable receiving all notifications when at home in comparison with an office. They highlighted that not all voice interventions can violate

privacy, but it depends on the type, nature, and setting in which the information is delivered. Therefore, it is essential to adopt measures to encode a voice interventions considering the setting and nature of intervention.

**7.2.3 Prioritization of Interventions: Considering Criticality and Urgency.** Previous research has found that users do not mind delivering health and well-being interventions directly without considering opportune moments for delivery [48]. In our experiment, users desired to prioritize interventions. They felt some interventions were urgent while some were not. Some interventions require immediate action while others can be acted upon later. Interventions such as hydration and sedentary reminders possibly can be postponed for some time and still be effective, however, stress interventions may require instant action. Some participants stated that they would like to scale the interventions based on an urgency and criticality parameter. It is essential to consider whether interventions are urgent or not. If the interventions are urgent, then direct voice interventions are appropriate.

**7.2.4 Mindful of Setting's Etiquette.** When VA is deployed in an office setting, it is essential to consider the rules of the office and interact accordingly. As WorkFit is ubiquitous and not stationary, it is compulsory to follow the etiquette of the place. When delivering interventions it is imperative to consider the fact that the office setting may have informal unscheduled meetings and smalltalk at the corridor. Audio should be kept at a moderate volume and not be disturbing to colleagues during the interaction. Also, users like to refrain from interacting with VA when they are in shared places.

**7.2.5 Considering social, location, and user availability factors.** Before delivering an intervention, it is important to identify if a particular intervention should be delivered in this setting. Before delivering interventions is essential to consider social, location and availability factors. Taking into account these factors will help to minimize the negative experiences of VA in office settings. We found that participants were comfortable having interaction in the home in comparison to office. The availability of the user is primary factor, as knowledge workers are often working from home. Due to work flexibility, they are switching between leisure activities and work, such as relaxing, and child-care. Hence, it is essential to consider user availability of while delivering intervention.

### 7.3 Limitations and Future Work

We would like to address the limitations of our paper. Firstly, it is important to note that we did not set up a baseline for participants daily water intake and break-time step counts. Our primary aim in this study is to assess the effectiveness of VA in office settings of knowledge workers. We opted for a healthcare-oriented approach solely as a means to evaluate the viability of voice interactions in the office setting. Consequently, the primary objective of our study did not involve a critical comparison of health recommendations with a baseline.

Additionally, we did not customize heart rate warnings for individual users. Instead, we employed fixed thresholds for low (55 bpm) and high (105 bpm) heart rate triggers [19, 25]. Lastly, WorkFit did not accommodate spontaneous input from participants (e.g., to

enable participants to record input outside the scope of interventions) resulting in participants relying on regular interventions to provide their input. In future, we intend to develop *adaptive VA* and a *dual-mode intervention* design specific for office settings, using contextual factors and different modalities.

## 8 CONCLUSION

In conclusion, this work addressed the imperative need for VA systems designed explicitly for office environments, focusing on the health and well-being of knowledge workers. The developed WorkFit voice assistance successfully monitored and intervened in sedentary behavior, inconsistent hydration, and stress issues, demonstrating its effectiveness in mitigating health concerns among knowledge workers. The field study involved 15 participants over five working days. We found four contextual factors for identifying inopportune moments (emotional and cognitive availability, social context, shared location, and user activity) in the office setting. Moreover, we identified that while VA does offer advantages for deployment in office settings, they also come with certain challenges that need to be addressed. Therefore, we propose the development of “dual-mode intervention” design that are specifically tailored for office settings. In light of our research outcomes, we present five design recommendations for the development of office-specific voice assistants. Lastly, WorkFit’s health recommendations demonstrated a high acceptance rate, with 92% of knowledge workers embracing hydration interventions and 79% engaging in recommended walking breaks. Additionally, the proposed breathing exercises significantly stabilized the participant’s high heart rate within 30 s and low heart rate within 10 s. Our study emphasizes the importance of designing voice interfaces for sensitive settings such as the office.

The findings of our research, including the identification of inopportune moments, empowering user agency, privacy management, recognition of social context, and assessment of emotional and cognitive availability, are relevant not only within sensitive environments like offices, but also extend to various other sensitive settings such as hospitals and universities. Moreover, this study has implications not only for WorkFit itself but also for the broader design of similar systems. For instance, the emphasis on health interventions tailored to real-time tracking suggests that future wellness programs could benefit from adopting a more targeted and adaptive approach.

## REFERENCES

- [1] Shashank Ahire. 2023. Designing a Smart Speaker for Emergent Users: Human Plus AI Response (*IndiaHCI '22*). Association for Computing Machinery, New York, NY, USA, 67–72. <https://doi.org/10.1145/3570211.3570217>
- [2] Shashank Ahire, Michael Rohs, and Simon Benjamin. 2022. Ubiquitous Work Assistant: Synchronizing a Stationary and a Wearable Conversational Agent to Assist Knowledge Work. In *2022 Symposium on Human-Computer Interaction for Work* (Durham, NH, USA) (*CHIWORK 2022*). Association for Computing Machinery, New York, NY, USA, Article 3, 9 pages. <https://doi.org/10.1145/3533406.3533420>
- [3] David Benyon. 1993. Adaptive systems: A solution to usability problems. *User Modelling and User-Adapted Interaction* 3, 1 (1993), 65–87. <https://doi.org/10.1007/bf01099425>
- [4] Ananya Bhattacharjee, Joseph Jay Williams, Jonah Meyerhoff, Harsh Kumar, Alex Mariakakis, and Rachel Kornfield. 2023. Investigating the Role of Context in the Delivery of Text Messages for Supporting Psychological Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 494, 19 pages. <https://doi.org/10.1145/3544548.3580774>

- [5] Timothy W. Bickmore, Rebecca A. Silliman, Kerrie Nelson, Debbie M. Cheng, Michael Winter, Lori Henault, and Michael K. Paasche-Orlow. 2013. A Randomized Controlled Trial of an Automated Exercise Coach for Older Adults. *Journal of the American Geriatrics Society* 61, 10 (2013), 1676–1683. <https://doi.org/10.1111/jgs.12449> arXiv:<https://agsjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/jgs.12449>
- [6] Scott A. Cambo, Daniel Avrahami, and Matthew L. Lee. 2017. BreakSense: Combining Physiological and Location Sensing to Promote Mobility during Work-Breaks. 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, 3595–3607. <https://doi.org/10.1145/3025453.3026021>
- [7] Narae Cha, Auk Kim, Cheul Young Park, Soowon Kang, Mingyu Park, Jae-Gil Lee, Sangsu Lee, and Uichin Lee. 2020. Hello There! Is Now a Good Time to Talk? Opportune Moments for Proactive Interactions with Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 74 (sep 2020), 28 pages. <https://doi.org/10.1145/3411810>
- [8] Narae Cha, Auk Kim, Cheul Young Park, Soowon Kang, Mingyu Park, Jae-Gil Lee, Sangsu Lee, and Uichin Lee. 2020. Hello There! Is Now a Good Time to Talk? Opportune Moments for Proactive Interactions with Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 74 (sep 2020), 28 pages. <https://doi.org/10.1145/3411810>
- [9] Don Samitha Elvitigala, Philipp M. Scholl, Hussel Suriyaarachchi, Vipula Disanayake, and Suranga Nanayakkara. 2021. StressShoe: A DIY Toolkit for Just-in-Time Personalised Stress Interventions for Office Workers Performing Sedentary Tasks. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse ; Virtual, France) (MobileHCI '21). Association for Computing Machinery, New York, NY, USA, Article 38, 14 pages. <https://doi.org/10.1145/3447526.3472023>
- [10] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. 2017. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health* 4, 2 (06 Jun 2017), e19. <https://doi.org/10.2196/mental.7785>
- [11] Healthline. 2021. How Much Water Should You Drink Per Day. <https://www.healthline.com/nutrition/how-much-water-should-you-drink-per-day> Accessed on January 19, 2023.
- [12] Junhan Kim, Yoojung Kim, Byungjoon Kim, Sukyung Yun, Minjoon Kim, and Joongseek Lee. 2018. Can a Machine Tend to Teenagers' Emotional Needs? A Study with Conversational Agents. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI EA '18). Association for Computing Machinery, New York, NY, USA, 1–6. <https://doi.org/10.1145/3170427.3188548>
- [13] Everlyne Kimani, Kael Rowan, Daniel McDuff, Mary Czerwinski, and Gloria Mark. 2019. A Conversational Agent in Support of Productivity and Wellbeing at Work. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. 1–7. <https://doi.org/10.1109/ACII.2019.8925488>
- [14] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for Workplace Reflection: A Chat and Voice-Based Conversational Agent. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) (DIS '18). Association for Computing Machinery, New York, NY, USA, 881–894. <https://doi.org/10.1145/3196709.3196784>
- [15] Kathryn M. Kolasa, Carolyn J. Lackey, and Ann C. Grandjean. 2009. Hydration and Health Promotion. *Nutrition Today* 44, 5 (Sep 2009), 190–201. <https://doi.org/10.1097/nt.0b013e3181b9c970>
- [16] Masahiro Komori, Yasuhiro Fujimoto, Jing Xu, Ken Tasaka, Hironori Yanagihara, and Koichi Fujita. 2019. Experimental Study on Estimation of Opportune Moments for Proactive Voice Information Service Based on Activity Transition for People Living Alone. In *Human-Computer Interaction. Perspectives on Design. HCI 2019 (Lecture Notes in Computer Science, Vol. 11566)*, M. Kurosu (Ed.). Springer, Cham. [https://doi.org/10.1007/978-3-030-22646-6\\_39](https://doi.org/10.1007/978-3-030-22646-6_39)
- [17] J. Lazar, J.H. Feng, and H. Hochheiser. 2010. *Research Methods in Human-Computer Interaction*. Wiley. [https://books.google.de/books?id=H\\_r6prUFpC4C](https://books.google.de/books?id=H_r6prUFpC4C)
- [18] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselstein. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300932>
- [19] Hao Liu, Jun Hu, and Matthias Rauterberg. 2010. iHeartRate: A Heart Rate Controlled in-Flight Music Recommendation System. In *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research* (Eindhoven, The Netherlands) (MB '10). Association for Computing Machinery, New York, NY, USA, Article 26, 4 pages. <https://doi.org/10.1145/1931344.1931370>
- [20] SA Manandhar and T Pramanik. 2019. Immediate Effect of Slow Deep Breathing Exercise on Blood Pressure and Reaction Time. *Mymensingh medical journal : MMJ* 28, 4 (October 2019), 925–929. <http://europepmc.org/abstract/MED/31599262>
- [21] Natalie A. Masento, Mark Golightly, David T. Field, Laurie T. Butler, and Carien M. van Reekum. 2014. Effects of hydration status on cognitive performance and mood | British Journal of Nutrition | Cambridge Core. *Cambridge Core* (Jan 2014). <https://doi.org/10.1017/S0007114513004455>
- [22] Oussama Metatla, Alison Oldfield, Taimur Ahmed, Antonis Vafeas, and Sunny Miglani. 2019. Voice User Interfaces in Schools: Co-designing for Inclusion with Visually-Impaired and Sighted Pupils. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3290605.3300608>
- [23] Christine Murad, Cosmin Munteanu, Leigh Clark, and Benjamin R. Cowan. 2018. Design Guidelines for Hands-Free Speech Interaction. 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, 269–276. <https://doi.org/10.1145/3236112.3236149>
- [24] Jakob Nielsen. 2024. 10 Usability Heuristics for User Interface Design. <https://www.nngroup.com/articles/ten-usability-heuristics/>
- [25] Yechiam Ostchega, Kathryn S. Porter, Jeffery Hughes, Charles F. Dillon, and Tatiana Nwankwo. 2011. Resting Pulse Rate Reference Data for Children, Adolescents, and Adults; United States, 1999–2008. (August 24 2011). <https://stacks.cdc.gov/view/cdc/12363>
- [26] Pablo Paredes, Ran Gilad-Bachrach, Mary Czerwinski, Asta Roseway, Kael Rowan, and Javier Hernandez. 2014. PopTherapy: Coping with Stress through Pop-Culture. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare* (Oldenburg, Germany) (PervasiveHealth '14). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Brussels, BEL, 109–117. <https://doi.org/10.4108/icst.pervasivehealth.2014.255070>
- [27] D. Parry, R.S. Oeppen, H. Gass, and P.A. Brennan. 2017. Impact of hydration and nutrition on personal performance in the clinical workplace. *British Journal of Oral and Maxillofacial Surgery* 55, 10 (Dec 2017), 995–998. <https://doi.org/10.1016/j.bjoms.2017.10.017>
- [28] Sharon Parry and Leon Straker. 2013. The contribution of office work to sedentary behaviour associated risk. *BMC Public Health* 13, 1 (2013). <https://doi.org/10.1186/1471-2458-13-296>
- [29] Jennifer Pearson, Simon Robinson, Thomas Reitmaier, Matt Jones, Shashank Ahire, Anirudha Joshi, Deepak Sahoo, Nimish Maravi, and Bhakti Bhikne. 2019. StreetWise: Smart Speakers vs Human Help in Public Slum Settings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300326>
- [30] Charlie Pinder, Jo Vermeulen, Benjamin R. Cowan, and Russell Beale. 2018. Digital Behaviour Change Interventions to Break and Form Habits. *ACM Trans. Comput.-Hum. Interact.* 25, 3, Article 15 (jun 2018), 66 pages. <https://doi.org/10.1145/3196830>
- [31] Leon Reicherts, Nima Zargham, Michael Bonfert, Yvonne Rogers, and Rainer Malaka. 2021. May I Interrupt? Diverging Opinions on Proactive Smart Speakers. In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) (CUI '21). Association for Computing Machinery, New York, NY, USA, Article 34, 10 pages. <https://doi.org/10.1145/3469595.3469629>
- [32] Simon Robinson, Jennifer Pearson, Shashank Ahire, Rini Ahirwar, Bhakti Bhikne, Nimish Maravi, and Matt Jones. 2018. Revisiting "Hole in the Wall" Computing: Private Smart Speakers and Public Slum Settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Montreal QC</city>, <country>Canada</country>, </conf-loc>) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3173574.3174072>
- [33] Anna Rudnicka, Dave Cook, Marta E. Cecchinato, Sandy J. J. Gould, Joseph W. Newbold, and Anna L. Cox. 2022. The End of the Active Work Break? Remote Work, Sedentariness and the Role of Technology in Creating Active Break-Taking Norms. In *2022 Symposium on Human-Computer Interaction for Work* (Durham, NH, USA) (CHIWORK 2022). Association for Computing Machinery, New York, NY, USA, Article 1, 13 pages. <https://doi.org/10.1145/3533406.3533409>
- [34] Samiksha Sanjiv Sathe, Tejal Rajandekar, Kirti Thodge, Amol Bhawane, and Utkarsh Thatere. 2020. Immediate Effect of Buteyko Breathing and Bhrumari Pranayama on Blood Pressure, Heart Rate and Oxygen Saturation in Hypertensive Patients: A Comparative Study. *Indian Journal of Forensic Medicine & Toxicology* (2020).
- [35] Steven Sauter, Lawrence Murphy, Michael Colligan, Naomi Swanson, Jr. Joseph Hurrell, Jr. Frederick Scharf, Raymond Sinclair, Paula Grubb, Linda Goldenhar, Toni Alterman, Janet Johnston, Anne Hamilton, and Julie Tisdale. 2014. STRESS...At Work. *The National Institute for Occupational Safety and Health (NIOSH)* (Jun 2014). <https://doi.org/10.26616/NIOSH/PUB99101>
- [36] Korok Sengupta, Sayan Sarcar, Alisha Pradhan, Roisin McNaney, Sergio Sayago, Debaleena Chattopadhyay, and Anirudha Joshi. 2020. Challenges and Opportunities of Leveraging Intelligent Conversational Assistant to Improve the Well-Being of Older Adults. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/3334480.3381057>

- [37] Anil Shankar, Sushil J. Louis, Sergiu Dascalu, Linda J. Hayes, and Ramona Houmanfar. 2007. User-Context for Adaptive User Interfaces. In *Proceedings of the 12th International Conference on Intelligent User Interfaces* (Honolulu, Hawaii, USA) (IUI '07). Association for Computing Machinery, New York, NY, USA, 321–324. <https://doi.org/10.1145/1216295.1216357>
- [38] Ben Shneiderman. 1997. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (3rd ed.). Addison-Wesley Longman Publishing Co., Inc., USA.
- [39] Sarah Silcox. 2015. Why hydration is a workplace issue. <https://www.personneltoday.com/hr/hydration-workplace-issue/>
- [40] WHO Team. 2020. WHO guidelines on physical activity and sedentary behaviour. <https://www.who.int/publications-detail-redirect/9789240015128>
- [41] Vincent W.-S. Tseng, Matthew L. Lee, Laurent Denoue, and Daniel Avrahami. 2019. Overcoming Distractions during Transitions from Break to Work Using a Conversational Website-Blocking System. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300697>
- [42] Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2015. Interruptibility Prediction for Ubiquitous Systems: Conventions and New Directions from a Growing Field. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) (UbiComp '15). Association for Computing Machinery, New York, NY, USA, 801–812. <https://doi.org/10.1145/2750858.2807514>
- [43] Alice Watson, Timothy Bickmore, Abby Cange, Ambar Kulshreshtha, and Joseph Kvedar. 2012. An Internet-Based Virtual Coach to Promote Physical Activity Adherence in Overweight Adults: Randomized Controlled Trial. *J Med Internet Res* 14, 1 (26 Jan 2012), e1. <https://doi.org/10.2196/jmir.1629>
- [44] Jing Wei, Tilman Dingler, and Vassilis Kostakos. 2022. Understanding User Perceptions of Proactive Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 185 (dec 2022), 28 pages. <https://doi.org/10.1145/3494965>
- [45] Jing Wei, Benjamin Tag, Johanne R Trippas, Tilman Dingler, and Vassilis Kostakos. 2022. What Could Possibly Go Wrong When Interacting with Proactive Smart Speakers? A Case Study Using an ESM Application. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 276, 15 pages. <https://doi.org/10.1145/3491102.3517432>
- [46] Tong Wu, Nikolas Martelaro, Simon Stent, Jorge Ortiz, and Wendy Ju. 2021. Learning When Agents Can Talk to Drivers Using the INAGT Dataset and Multisensor Fusion. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 133 (sep 2021), 28 pages. <https://doi.org/10.1145/3478125>
- [47] Mengru Xue, Rong-Hao Liang, Jun Hu, Bin Yu, and Loe Feijs. 2022. Understanding How Group Workers Reflect on Organizational Stress with a Shared, Anonymous Heart Rate Variability Data Visualization. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 27, 7 pages. <https://doi.org/10.1145/3491101.3503576>
- [48] Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Voelkel, Johannes Schoening, Rainer Malaka, and Yvonne Rogers. 2022. Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) (CUI '22). Association for Computing Machinery, New York, NY, USA, Article 3, 14 pages. <https://doi.org/10.1145/3543829.3543834>