Dual-Mode Interventions: Giving Agency to Knowledge Workers in Proactive Health Interventions

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

Saeid Othman
Human-Computer Interaction
Leibniz University Hannover
Hannover, Germany
othman2@stud.uni-hannover.de

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

ABSTRACT

In the domain of health and well-being, proactive voice interventions have demonstrated their efficacy. However, users often encounter privacy concerns and social embarrassment due to the lack of control over these proactive interventions, especially in formal and social settings. This study introduces a novel approach called "dual-mode intervention." It begins with primary interventions using different modalities (like graphical, tactile, or auditory). If users do not respond to these primary interventions, the system delivers voice interventions after a short interval. We conducted a study employing a within-subjects design, which involved 15 participants. The study compared dual-mode interventions with direct voice interventions in office settings, focusing on addressing health and well-being issues. Our findings indicate that knowledge workers preferred dual-mode interventions over direct voice interventions. Moreover, direct voice interventions received significantly lower ratings compared to dual-mode interventions. Also, we identify user preferences for different dual-intervention modalities. Our findings reveal that the user preferences depend on the type of health intervention. Vibration emerged as the preferred modality, followed by graphical output, auditory icons, and ringing interventions.

CCS CONCEPTS

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

KEYWORDS

voice assistant, office, proactive, voice interface, feedback modalities, health, well-being, knowledge worker

ACM Reference Format:
Shashank Ahire, Saeid Othman, and Michael Rohs. 2024. Dual-Mode Interventions: Giving Agency to Knowledge Workers in Proactive Health Interventions. In *ACM Conversational User Interfaces 2024 (CUI '24), July 8–10, 2024, Luxembourg, Luxembourg*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3640794.3665578

1 INTRODUCTION

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. It is widely advised in various human-computer interaction guidelines and design recommendations to empower users with agency or control during voice interactions [1, 10]

Proactive voice reminders have proven beneficial for communicating urgent and critical health and well-being issues [4, 12, 13]. Voice User Interface (VUI) are highly effective at delivering information and capturing user attention instantly [2]. Furthermore, it also enables users to perform the ongoing task and simultaneously perform an eyes-free and hands-free interaction with VUI seamlessly. This makes them invaluable for delivering critical health interventions and prompting user action. Knowledge workers expressed a desire for voice interfaces to address their health and well-being needs [3]. However, VUI can also be intrusive and can create problems in multi-party conversations, informal meetings, and social settings. In an office context, which consists of formal and social interaction, it is essential to give users control to decide if the particular moment is opportune for interaction. Allowing users to decide if the moment is appropriate for interaction will lower the chances of inconvenience.

This study builds upon a previous study conducted in an office setting [1], in which we developed *WorkFit*, a voice assistant designed for supporting the health and well-being of knowledge workers. WorkFit monitors a knowledge worker's behaviour and intervenes in instances of sedentary behavior, inconsistent hydration, and stress-related issues in office environments. The subsequent field study involved 15 participants over the course of five working days. Our findings revealed challenges, such as empowering users with agency, avoiding intrusiveness for colleagues, and privacy management.

To enable knowledge workers with agency during proactive voice interaction. We propose a "dual-mode intervention" design. The Voice Assistant (VA) may choose to intervene non-intrusively using different modalities for primary intervention, reserving voice as a secondary intervention. Before delivering a voice intervention, it is important to introduce the upcoming voice intervention. For example, the health intervention could be presented through a GUI, to convey the notification's intent, complemented by a vibration, voice, ringing, or auditory icon to draw 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. This gives the user the opportunity to have agency over the proactive voice intervention.

In this study we designed and developed *DeskCare*, a voice assistance for a smartwatch using dual-mode intervention. DeskCare deliveres mental, physical, and nutritional interventions in five modalities: graphical, voice, vibration, auditory icon, and ring. We

CUI '24, July 8–10, 2024, Luxembourg, Luxembourg

ACM ISBN 979-8-4007-0511-3/24/07.
https://doi.org/10.1145/3640794.3665578

evaluated DeskCare with 15 knowledge workers in the lab and in a field study. The findings show that knowledge workers rated dual-interventions significantly higher in-comparison to direct voice interventions. Also, in dual-mode interventions knowledge workers preferred vibration over graphical, auditory icon, and ring for primary mode for health interventions.

Our paper makes three contributions: (1) Designed and implemented dual-mode interventions to empower users with agency over proactive interventions. (2) Conducted a study of dual-mode interventions against proactive voice interventions. (2) Examined modality preferences within dual-mode interventions based on intervention type.

2 RELATED WORK

In previous research, Cha et al. [5] investigated opportune moments for smart speaker interaction among 40 participants. They determined that these moments depend on factors such as user busyness, task complexity, mood, and social availability, identifying repetitive activities requiring low attention as conducive to interaction. Similarly, Reicherts et al. [11] examined proactive interaction timing and content, revealing their significance in shaping user attitudes. They also found that excessive proactive interactions negatively impacted user perceptions.

Wei et al. [14] explored participant availability for conversation, noting higher availability during entertainment tasks than during work or study. They observed a significant correlation between self-reported boredom, mood, and participant availability. In another study, Zargham et al. [15] used storyboards to gauge user perceptions of proactive smart speaker interventions. Participants expressed appreciation for proactive interventions related to health.

Furthermore, Kocielnik et al. [7] compared voice and chat-based conversational agents (CAs) for workplace reflection. They found voice interaction more convenient in office settings, but noted participants felt pressured to respond instantly. Additionally, interaction with CAs increased awareness of unhealthy work practices. Moreover, Kocielnik et al. [8] introduced the "Reflection Companion," a conversational system encouraging reflection on physical activity. Dialogues prompted self-reflection and behavior change adoption among users.

3 PROTOTYPE AND DESIGN

The "dual-mode intervention" comprises two levels of interventions: the primary intervention followed by the secondary intervention. The primary intervention involves utilizing modalities such as vibration, graphical displays, auditory icons, and ring tones, while the secondary intervention relies solely on voice. The primary intervention aims to deliver health and well-being concerns subtly, bringing them to the immediate attention of users. If users fail to interact with the primary notification, the secondary voice intervention is deployed to ensure that the health notification is effectively conveyed to the user's attention.

3.1 Health and Well-Being Intervention

The three components–mental break, physical break, and nutritional break–are essential for maintaining health and productivity in the workplace. Based on design recommendations in 'WorkFit'[1], we prioritize these health interventions based on urgency and criticality factors.

3.1.1 Mental Break Intervention. Considering the cognitive demands associated with knowledge work, providing opportunities for mental activities can alleviate stress and mental fatigue, thereby improving concentration and cognitive performance. We classify mental break interventions as urgent and critical. Upon receiving a stress intervention, it is imperative for the knowledge worker to promptly address the stressful situation by opting for a mental break.

3.1.2 Physical Break Intervention. Acknowledging the sedentary nature of office work, we incorporate physical activities to promote frequent movement and posture changes to mitigate the negative impacts of extended periods of sitting. We consider physical interventions as critical but not urgent, as knowledge workers can delay a physical break for some time and act on it accordingly when appropriate. Delaying the physical break may result physical discomfort.

3.1.3 Nutrition Break Intervention. Recognizing the significance of nutrition and hydration in maintaining overall health and energy levels, we integrated nutrition activities to encourage healthy eating habits. Our objective was to support sustained productivity throughout the day by fostering a focus on dietary routine and hydration. A nutritional break can be termed as not urgent and not critical. Users do not incur any loss if the break is postponed for an extended period. However, it is important to act upon it accordingly.

3.2 Modes of Intervention

Ideally, interventions should deliver messages in a non-intrusive manner, avoiding distraction to the user and minimizing attention from surrounding individuals. The five notifications modes examined in this work were selected based on the need for flexibility in how notifications are received.

3.2.1 Graphical. Visual notifications are tailored to users who either prefer or need visual cues, rendering it an inclusive choice suitable for various work environments.

3.2.2 Vibration. Selected for its subtlety, vibration enables users to receive notifications discreetly, making it suitable for environments such as open offices or meetings where minimizing disruptions is important.

3.2.3 Auditory Icon. Employing familiar sounds provides an intuitive method to communicate information without requiring direct screen interaction, seamlessly integrating into the user's surroundings.

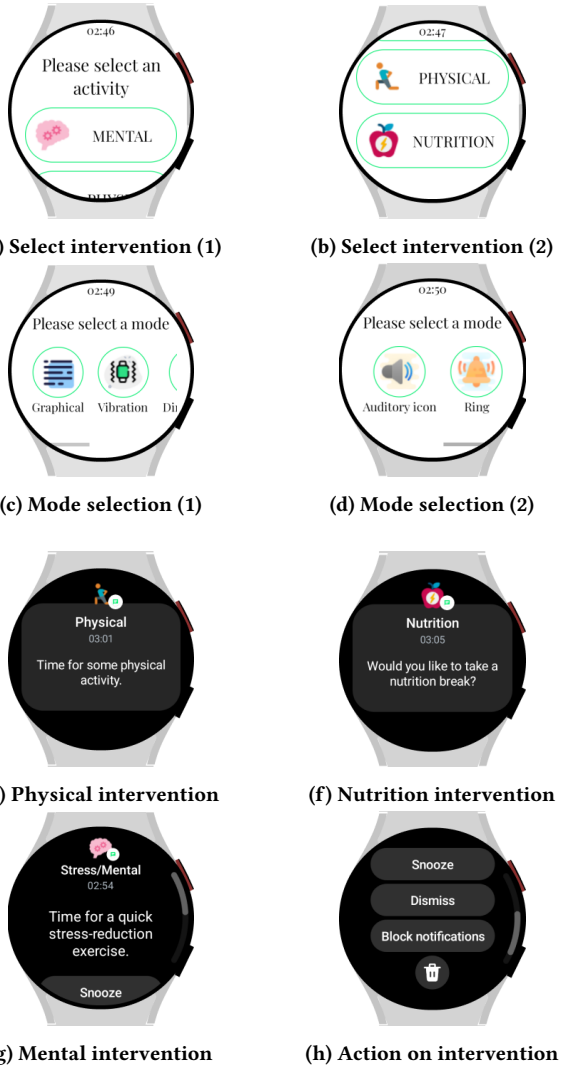
3.2.4 Ring Mode. Utilizing the ring mode can be effective in ensuring that crucial notifications are noticed, particularly in scenarios where other modes, such as graphical, may be overlooked.

3.2.5 Direct Voice. This mode provides clear, direct communication for users who may benefit from unambiguous auditory statements, especially when the visual or tactile senses are preoccupied.

3.3 Smartwatch

Smartwatches offer multi-modal interactions, encompassing graphical displays, voice input and output, as well as vibration feedback, allowing users to engage through diverse modalities. The constant wearability of smartwatches ensures that interventions and reminders are always immediately accessible, facilitating timely and effective delivery, thereby capturing the user's attention effectively. We selected Samsung Galaxy Watch4[1] for our evaluation purposes.

4 DESIGN AND DEVELOPMENT



(a) Select intervention (1) (b) Select intervention (2)

(c) Mode selection (1) (d) Mode selection (2)

(e) Physical intervention (f) Nutrition intervention

(g) Mental intervention (h) Action on intervention

Figure 1: Screenshots of the DeskCare graphical interface.

To evaluate dual-mode interventions, we developed an Android smartwatch VA named *DeskCare*. The application landing screen shows three different types of health and well-being interventions (as shown in Figure 1a, Figure 1b). The user has to select the modality (Figure 1c, Figure 1d) in which the intervention should be delivered to the participant. After confirming the modality, the intervention is delivered when the set time interval has passed. The time interval can be set by the experimenter.

DeskCare delivered intervention in two stages: first the primary and then the secondary intervention.

Primary Intervention. The primary intervention comprises the graphic, vibration, auditory icon, and ring modality. First, the primary intervention is executed using the chosen modality, along with the corresponding screens (see Figure 1e, Figure 1f, Figure 1g), and the action screen (refer to Figure 1h). Depending on the selected modality: (1) If auditory icon is chosen, the app delivers the associated health and well-being intervention through a sound snippet. The sounds are available online[2]. The sound for the mental intervention is a relaxing spherical sound. The sound for the nutrition intervention is the sound of filling a glass of water. The sound for the physical intervention is the sound of a runner's footsteps. (2) For the ring modality, a continuous ring tone[3] is played for 5 seconds as the primary intervention. (3) In the case of the vibration intervention, a default vibration pattern is triggered as the intervention. (4) Lastly, for the graphical intervention only (refer to Figure 1e, Figure 1f, Figure 1g), screens are displayed followed by the action screen (see Figure 1h).

Secondary Intervention. The secondary intervention is only delivered using voice statements. If the user fails to take any action to snooze or decline via the action screen (Figure 1h), a secondary voice intervention is delivered after a 2-minute interval. In the voice modality, the intervention is as follows:

- Physical: *"It has been 45 minutes. Would you like to engage in a physical activity?"*
- Nutrition: *"It has been 1 hour. Would you like to drink water or eat something healthy?"*
- Mental: *"It seems that you are stressed. Would you like to take a mental break and engage in an exercise to reduce stress?"*

5 STUDY

5.1 Method

We conducted a within-subjects comparative study of dual-mode intervention and direct voice interventions. The study was performed both in a lab and in the real office settings of the participants. Participants who felt uncomfortable conducting the evaluation within their office environments, citing concerns such as privacy, avoidance of disrupting colleagues, and adherence to formalities, were invited to our laboratory instead. During the experiment, participants were requested to bring their laptops and simulate working as they would in their actual office, imagining themselves surrounded by colleagues. The evaluation took place while knowledge workers were seated at work desks. In both settings, the experimenter (second author) guided the participants through the study. To deliver

[1] https://www.samsung.com/global/galaxy/galaxy-watch4/specs/

[2] https://drive.google.com/drive/folders/12lo2yxh-WjCsGNRP7_JASWjUBPIL-mXH?usp=sharing

[3] https://drive.google.com/file/d/1CFgR1d9Vouzyx4t31qH3W9ZaOIPJwF_-/view?usp=sharing

voice interventions, we asked participants to refrain from taking action on the primary intervention(Figure 1h).

The study involved two surveys and the evaluation of the DeskCare app. We commenced with a survey gathering demographic information about the participant, followed by an explanation of the purpose of the DeskCare app. Participants completed the second survey form after evaluating each mode. Altogether, participants evaluated fifteen interventions, comprising five modes of intervention across three types of breaks.

The assessment of modalities was counterbalanced in terms of both intervention type and modalities. After evaluating each modality for every intervention, participants were instructed to rate the User Experience Questionnaire (UEQ) score for each modality. This questionnaire focuses on parameters such as enjoyment, understanding, pleasantness, safety, expectations, efficiency, and practicality, understandingly and clear. Participants rated these usability parameters on a scale of 1 to 7. The experiment required approximately 40 to 50 minutes to complete.

5.2 Participants

The study comprised 15 knowledge workers, all of whom were male. Nine knowledge workers participated in a laboratory setting, while six conducted the study in their actual work environment. The participants fell within the age range of 21 to 29 years, with an average age of 24.9 years and a standard deviation of 2.5 years. All participants were either full-time or part-time knowledge workers, with an average weekly working time of 23.86 hours.

6 RESULTS

6.1 Quantitative Findings

The Figure 2 shows the box-plot of UEQ ratings for the scales clear, efficient, enjoyable, friendly, meet expectations, pleasant, practical, and safe. The Direct Voice (DV) intervention was rated lower on all the scales and across all the interventions of dual-mode technique. We checked if there was a significant difference among the the modalities with a Friedman test.

In the case of mental intervention (Table 1), the Friedman test revealed significant different among multiple parameters (Enjoyable, Supportive, Pleasant and Practical). After conducting pairwise comparisons, significant differences were identified between the graphical and direct voice modes, as well as between the vibration and direct voice modes, across all four criteria. Furthermore, significant differences were observed between the direct voice and auditory icon modes on the criteria of enjoyment and pleasantness.

Table 1: Mental activity UEQ ratings. Friedman test of UEQ scales. Tick mark (✓) signifies p < 0.05. GUI: graphical, Vi: vibration, DV: direct voice, AI: auditory icon, R: ring.

	Overall		Pairwise				
	Chi-square	P-value	GUI:DV	Vi:DV	Vi:AI	DV:AI	DV:R
Enjoyable	15.85	0.003	✓	✓		✓	
Supportive	12.09	0.017	✓	✓			
Pleasant	17.16	0.002	✓	✓		✓	
Practical	18.49	0.001	✓	✓	✓		✓

In physical intervention (Table 2), the Friedman test unveiled noteworthy disparities among the criteria (enjoyable, efficient and practical). Upon conducting pairwise comparisons, significant differences were observed between the vibration and direct voice modalities across all three criteria. Furthermore, significant differences were identified between the direct voice and graphical modalities, on the enjoyable criteria.

Table 2: Physical activity UEQ ratings. Friedman test of UEQ scales. Tick mark(✓) signifies p < 0.05. GUI: graphical, Vi: vibration, DV: direct voice.

	Overall		Pairwise	
	Chi-square	P-value	GUI:DV	Vi:DV
Enjoyable	10.19	0.037	✓	✓
Efficient	10.77	0.029		✓
Practical	14.76	0.005		✓

The Friedman test for the nutrition intervention revealed no significant differences among the criteria.

After evaluating each modality for each health intervention, participants were asked to rank each modality for each of the interventions (See Table 3). To determine the ranking of modes across all three activities, we utilized the Borda count method[4].

Table 3: Ranking of modalities for each health intervention.

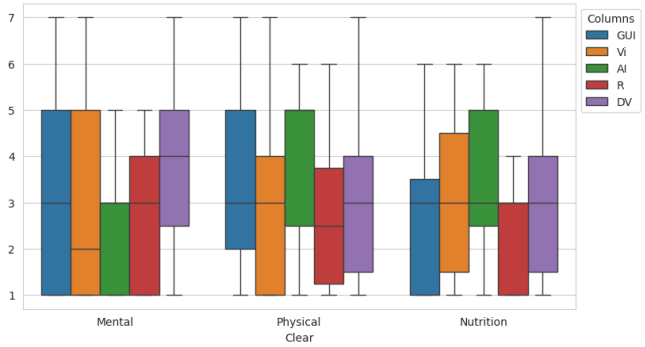
Rank	Points	Mental	Physical	Nutrition
1	5	Vibration	Vibration	Vibration
2	4	Graphical	Auditory icon	Ring
3	3	Auditory icon	Graphical	Graphical
4	2	Ring	Ring	Auditory icon

6.2 Qualitative Findings

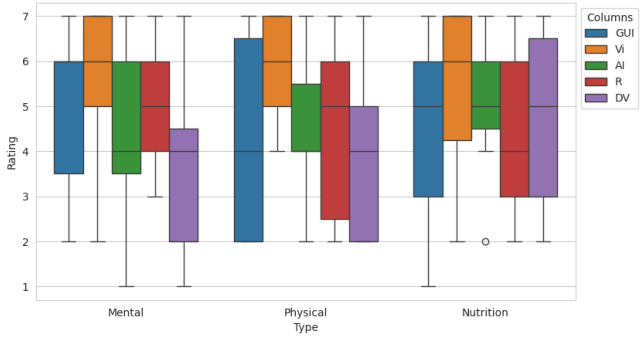
After evaluation, we interviewed participants regarding their modality preference. We found that vibration was appreciated for its unobtrusiveness. One participant explained: *"The discreet nature of vibrations allows me to stay up-to-date without disturbing my colleagues or interrupting my workflow."* [P1] The graphical modality was praised for how visual cues enable quick recognition and minimal distraction. *"I can grasp information with a glance without having to divert from my work, which enhances my productivity."* [P5] The auditory icon was recognized for its effectiveness in situations where visual signals could not be perceived. *"Acoustic signals are helpful when I'm not at my desk; they reach me where visual cues cannot."* [P3]

The ring modality was considered useful by some users for urgent notifications, but with reservations about potential disturbance. *"While ringtones can draw my attention to urgent matters, they often cause interruptions that disrupt my workflow and that of my colleagues."* [P4] For direct voice messages, concerns regarding privacy and the disturbance of the surroundings reflect the limitations

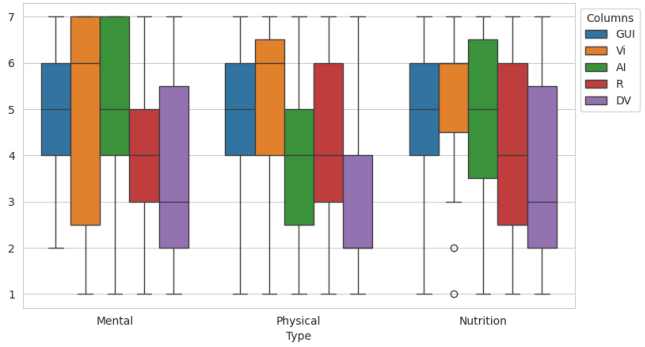
[4]https://en.m.wikipedia.org/wiki/Borda_count



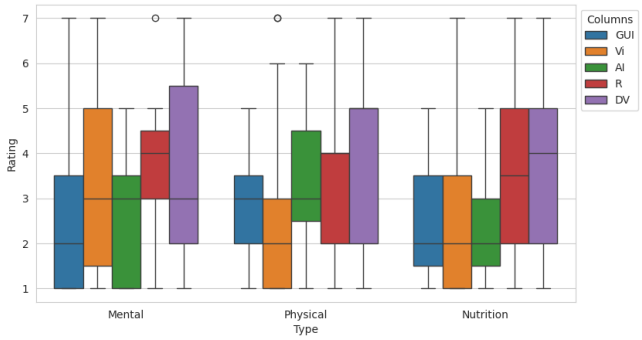
(a) Clear (1= clear, 7=confusing)



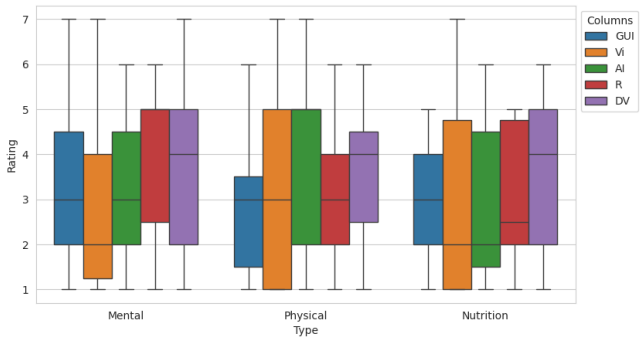
(b) Efficient (1=inefficient, 7=efficient)



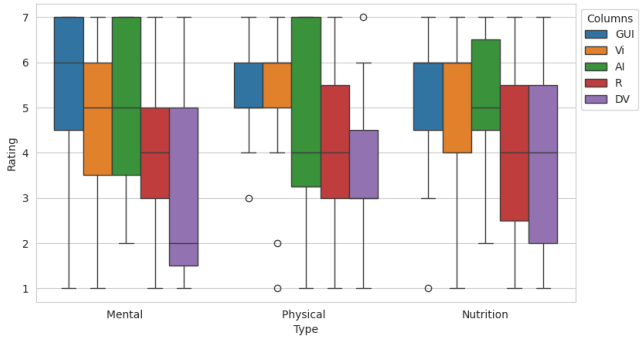
(c) Enjoyable (1=annoying, 7=enjoyable)



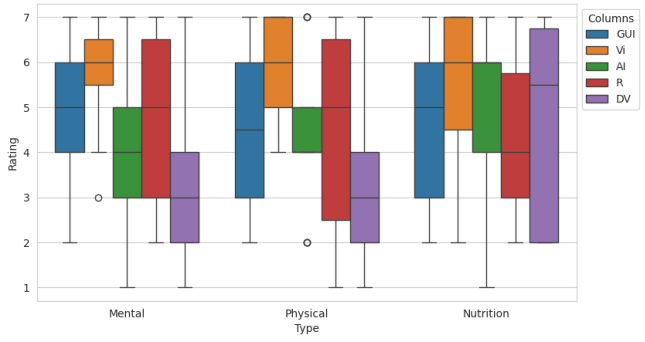
(d) Friendly (1=friendly, 7=unfriendly)



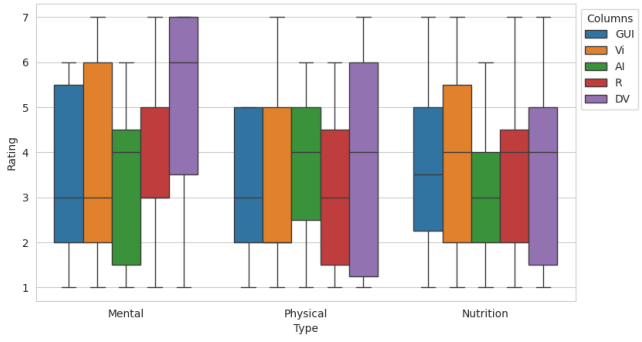
(e) Meet Expectations (1= meets expectations, 7= does not meet expectations)



(f) Pleasant (1=unpleasant, 7=pleasant)



(g) Practical (1=impractical, 7=practical)



(h) Safe (1=secure, 7=not secure)

Figure 2: For each of the three interventions (mental, physical, nutrition) and each of the five modalities (graphical, vibration, auditory icon, ring, direct voice) the UEQ ratings are given. GUI = Graphical user interface, Vi = Vibration, AI = Auditory Icon, R = Ring, DV = Direct Voice.

associated with this mode in a professional context. *"Direct voice messages provide an unambiguous way of communication, but they compromise too much privacy and disturb others in an open office setting."* [P7]

7 DISCUSSION

Having control over proactive voice interventions is crucial, especially in sensitive settings like the office. In the dual-mode intervention, we provided a secondary intervention if the user did not respond to the primary one. The time interval between the primary and secondary interventions could vary based on the urgency and criticality of the situation. The more urgent and critical the event, the shorter the time interval.

Initially our hypothesis was that on certain factors of the UEQ scale (such as friendliness, clear, and pleasant) knowledge workers could have been biased towards voice due to its anthropomorphic characteristic [9]. However, knowledge worker's responses considering the office environment were entirely different. Furthermore on certain factors voice had significant difference in-comparison to graphical, vibration and auditory icon (as shown in Table 1 and Table 2)

In our research, vibration emerged as the most favored primary intervention across all intervention types. Its non-intrusive and private nature made it a popular choice among participants. While auditory icons and ring tones also offer privacy, they were deemed potentially intrusive to participant's colleagues. On the other hand, graphical outputs were prone to being overlooked. Based on our findings, we recommend incorporating both the vibration and graphical modality as primary interventions and voice as a secondary intervention for promoting health and well-being in office settings. Our results align with previous literature where authors proposed implementing an additional tactile modality with voice for VUI in cars. They found that combining voice and tactile modalities enhances CUI efficiency and user experience, while minimizing distraction of the driver, as demonstrated in driving scenarios [6].

It is important to note that our study was conducted specifically at office desks, so these recommendations are tailored to that context. However, modal preferences may vary in other office locations such as a kitchen or conference room. Therefore, further investigations into modality preferences across different locations is warranted.

8 LIMITATION

The study was exclusively conducted with male participants. Additionally, the study's participants were notably younger in age.

9 CONCLUSION

In conclusion, our study introduces a novel approach known as *dual-mode intervention*, to overcome the lack of user agency in proactive voice interventions. The dual-mode technique initially employs non-intrusive primary interventions through various modalities (including graphical, vibration, auditory icon, and ring) and delivers voice interventions only if the user fails to interact with the primary intervention. We evaluated the effectiveness of the dual-mode intervention technique in office environments for health and well-being concerns. We performed a study to compare dual-mode interventions and direct voice interventions. The evaluation encompassed 15 knowledge workers in both laboratory-controlled and real office settings. The results show that the dual-mode technique is strongly preferred over direct voice output in proactive conversational interfaces. To enhance user acceptance, we strongly advocate for utilizing an alternative modality prior to resorting to voice output. The UEQ ratings show significant differences on scales such as enjoyable, supportive, efficient, pleasant, and practical. Additionally, we identified knowledge worker modality preferences for health interventions. Our findings reveal that users exhibit diverse preferences for different modalities depending on the type of intervention. Vibration emerged as the preferred modality, followed by graphical, auditory icon, and ring interventions.

REFERENCES

[1] Shashank Ahire, Simon Benjamin, 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) *(CUI '24)*. Association for Computing Machinery, New York, NY, USA, 14 pages. https://doi.org/10.1145/3640794.3665561

[2] Shashank Ahire, Aaron Priegnitz, Oguz Önbas, Michael Rohs, and Wolfgang Nejdl. 2021. How Compatible is Alexa with Dual Tasking? — Towards Intelligent Personal Assistants for Dual-Task Situations. In *Proceedings of the 9th International Conference on Human-Agent Interaction* (Virtual Event, Japan) *(HAI '21)*. Association for Computing Machinery, New York, NY, USA, 103–111. https://doi.org/10.1145/3472307.3484165

[3] Shashank Ahire, Michael Rohs, and Simon Benjamin. 2022. Ubiquitous Work Assistant: Synchronizing a Stationary and a Wearable Conversational Agent to Assist Knowledge Work. In *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work* (Durham, NH, USA) *(CHIWORK '22)*. Association for Computing Machinery, New York, NY, USA, Article 3, 9 pages. https://doi.org/10.1145/3533406.3533420

[4] 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

[5] 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

[6] Jingun Jung, Sangyoon Lee, Jiwoo Hong, Eunhye Youn, and Geehyuk Lee. 2020. Voice+Tactile: Augmenting In-vehicle Voice User Interface with Tactile Touchpad Interaction. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, Hi, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376863

[7] 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

[8] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 70 (jul 2018), 26 pages. https://doi.org/10.1145/3214273

[9] Seul Chan Lee, Harsh Sanghavi, Sangjin Ko, and Myounghoon Jeon. 2019. Autonomous driving with an agent: speech style and embodiment. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (Utrecht, Netherlands) *(AutomotiveUI '19)*. Association for Computing Machinery, New York, NY, USA, 209–214. https://doi.org/10.1145/3349263.3351515

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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

[15] 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