Proactive Smart Speakers for Chronic Disease Management: Challenges and Opportunities

Jing Wei
University of Melbourne
Melbourne, Australia
jwwei2@student.unimelb.edu.au

Tilman Dingler
University of Melbourne
Melbourne, Australia
tilman.dingler@unimelb.edu.au

Enying Gong
University of Melbourne
Melbourne, Australia
egong@student.unimelb.edu.au

Brian Oldenburg
University of Melbourne
Melbourne, Australia
brian.oldenburg@unimelb.edu.au

Vassilis Kostakos
University of Melbourne
Melbourne, Australia
vassilis.kostakos@unimelb.edu.au

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
Smart speakers provide users with a natural modality of interaction - voice. Voice User Interfaces (VUIs) can be natural to use and inclusive to a majority of users. Current smart speakers, however, are mostly passive and only respond to user input once they have explicitly been addressed. Lacking proactivity restrains the range of application scenarios of smart speakers in a number of ways: people with chronic conditions, for example, could benefit from engaging voice-based conversations and reminders. In our research, we investigate opportunities and challenges for proactive, VUI-based chronic disease management systems. We propose a method to enable existing commercial smart speakers to become proactive. We further discuss the technology and interaction challenges in implementing such proactive smart speakers.

Author Keywords
Voice User Interface; Smart Speakers; Conversational Agents; Chronic Disease Management

CCS Concepts
•**Human-centered computing** → **Interaction design;** *Interaction design process and methods;* User interface design;

Introduction

Smart speakers have soared in popularity in recent years. People can use their voice as a natural user interface (NUI) to request information from smart speakers and enable home automation (e.g., turn on/off the lights). Various VUI-based applications, ranging from entertaining games to flight booking systems, for example, are now widely available on commercial smart speakers. Compared to typing or other conventional input devices, voice is intrinsically a more inclusive modality for a broader population to interact with smart devices [5].

Existing smart speakers mostly function as *passive* virtual personal assistants that need to be activated by special voice commands (e.g., 'Hey Google' for activating Google Home) before recognizing and responding to people's commands. Recent smart speakers, such as LifePod and Amazon Alexa, are exploring more proactive services: LifePod [1] claims to be a proactive smart speaker designed for older adults by reminding them about medical adherence and doctors' appointments. Amazon has provided ProactiveEvents API[2] for customized notifications, but the notification messages still need to be explicitly retrieved by users.

Overall, current smart speakers still lack the ability to proactively engage users by initiating conversations, which limits a range of application scenarios. For example, proactivity would benefit the implementation of intervention systems for chronic disease management. Several studies have suggested that conversation-based self-management and coaching programs are effective in strengthening patients' self-efficacy in managing their conditions and improve their quality of life [8]. As such, proactive educational messages and reminders are necessary for effective intervention systems. Proactivity can provide such opportunities to help patients achieve better self-management results. At the same time, potential usability challenges and difficulties also exist when implementing and using proactive speakers. In this position paper, we propose a method to enable proactivity for commercial speakers and identify future opportunities and challenges in prototyping a proactive domain-specific smart speaker system.

Interventions for Chronic Disease Management

Guided by the *Chronic Care Model*, supporting patients' self-management is a prominent component to improving chronic conditions management [4]. The goal of self-management support is to inform and empower patients with knowledge and behavior change skills so that they can be active, confident, and engaged in their self-management process, which will promote shared decision-making and better health outcomes [6, 1].

An optimal self-management support needs to cover a range of functions, such as providing evidence-based information, assisting self-monitoring, coaching patients to acquire skills, providing emotional support, navigating patients within the healthcare system, and supporting their communication with healthcare providers. Previous studies have implemented and evaluated technology-enabled self-management systems by delivering voice- or text-based self-management support to people with chronic conditions through computers, telephones, phone messages, and smartphone applications as interfaces [8]. These studies have demonstrated the effectiveness of these interventions in improving behavior and clinical outcomes, but also identified challenges in user engagement (e.g., users are more engaged when receiving personalized messages) and maintenance.

[1] https://lifepod.com/

[2] https://developer.amazon.com/en-US/docs/alexa/smapi/proactive-events-api.html

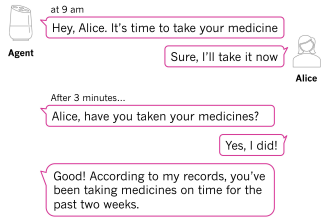

Figure 1: Conversation example 1.

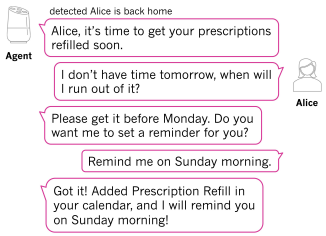

Figure 2: Conversation example 2.

Overall, existing intervention systems for chronic disease management aim to improve patients' awareness of disease prevention and help them maintain a healthy lifestyle. Proactively providing patients with necessary and evidence-based advice and interventions through either text-, voice-, or app-based applications is a critical component of intervention systems. Henceforth, proactive smart speaker systems that provide more engaging and optimal supports may further improve on existing intervention systems.

Opportunities for VUI-based Systems

We argue that several advantages of VUI make them more than just an alternative to traditional GUI-based mobile application intervention systems or interactive voice response systems (IVRS).

Compared to text, voice is more versatile and engaging for interaction [3]. Based on users' preferences, the voice of the smart speaker can be customized to be both genders and different ages for various applications, and even the language used by the speaker can set as local dialects. It is easier to convey emotions in voice messages, and users can benefit from the emotion of the speech they get from smart speakers [7]. Smart speakers can give patients options to choose their favorite voice to hear messages so that they can interact more enjoyably. Furthermore, voice communication is intrinsically hands-free and even eyes-free, which makes VUI a suitable modality for certain populations, such as people with physical disabilities.

Compared to existing IVRS (which are centrally controlled and limited to interaction through telephone keypads), smart speakers are designed to support richer range of conversations. The conversations could be of different forms and have more diverse content, which could range from simply reminding patients (e.g., see Figure 1 and 2) to administering self-assessment (e.g., asking patients *how are you feeling today?*). The improved engagement fulfills the various and changing needs of patients and encourages them to be more proactive in their disease management journey.

Compared to smartphones, smart speakers are not mobile and mostly located at user homes. The location of smart speakers implicitly suggests that it might be more appropriate to discuss potentially sensitive information of users. When being outside of the home, people may be reluctant to respond to phone calls and text messages for chronic condition management purposes in public due to privacy concerns. Smart speakers located at home can serve as a central interface for a wide range of devices and monitors, such as smartwatches and smart blood pressure monitors. Key lifestyle factors, such as sleep patterns, diet, and blood pressure/glucose measurement, can be monitored by those smart-home devices, and feedback can be generated from tracked data to promote patients' behavior changes. Additionally, if those smart devices can directly transmit data to speakers and skip the smartphone as the intermedium, patients who find using smartphones difficult, especially illiterate older adults in developing countries, can easily access their everyday data.

A stark limitation of current commercial smart speakers, which can potentially become a research opportunity, is that today smart speakers are not designed to be proactive, and they can only respond to users' commands. As such, people are likely to perceive non-proactive smart speakers as servants, and will not form relationships with them [2]. Once proactivity is enabled for smart speakers, voice messages cannot easily be ignored as phone notifications. While this obtrusiveness can bring inconvenience, it may also increase the likelihood of messages being acknowledged and responded to. Proactivity is likely to spark new research opportunities

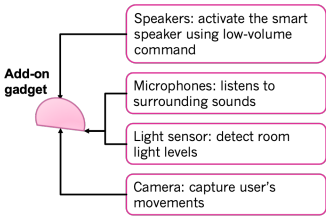

Figure 4: The add-on gadget prototype.

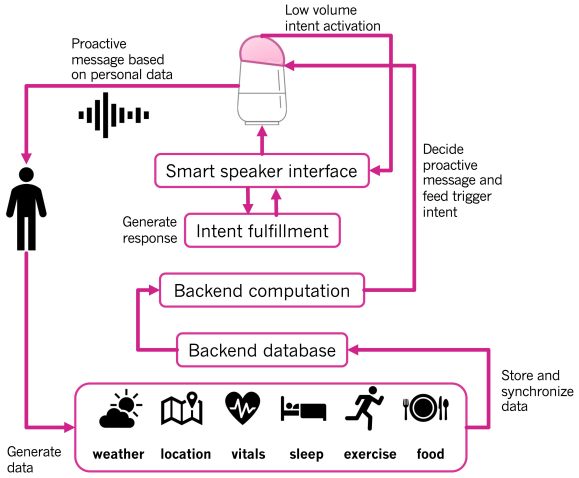

Figure 3: Architectural overview. The pink middleware add-on gadget connects to the backend system and can activate the speaker.

for developing novel applications (i.e., intervention systems in our case). We also hypothesize that an additional implication of proactivity is that it raises the expectation that the speakers try to understand users continuously, and may thus be perceived as more "equal" to humans and lead to new types of *bonds* between users and smart speakers.

Towards Proactive Smart Speakers

While companies like Google and Amazon provide a plethora of tools and easy access to develop VUI-based applications, both of them offer very limited ways to proactively interact with users through Google Assistant or Alexa Voice Service. To build a proactive prototype based on current commercial smart speakers, a *workaround* is to add a middleware add-on gadget to activate the speaker.

Enable the Proactivity

The architectural overview of the proactive speaker is shown in Figure 3, and the add-on gadget is depicted in Figure 4. The add-on gadget provides the speaker proactivity by receiving the message from the backend system, sensing the environment to determine when to deliver voice messages, and activating the speaker using a low-volume voice. The gadget can contain sensors including light sensor, microphone, and camera. Data from those sensors can then be used to detect users' context (e.g., come back home can be a good time to deliver messages).

Behind the Proactivity

The proactivity needs to be supported by a backend database that stores the end user's data and a backend system that computes what messages to be sent to users and what intents to be triggered. Once the message and intent are decided, the backend system notifies the middleware add-on gadget to activate the speaker.

Challenges for Proactive Speakers

The proactivity does not merely imply that speakers are capable of delivering proactive prompts; in fact, there is a series of challenges in different stages of implementing proactive smart speakers depicted in Figure 5. Challenges lie in not only the message generation and delivery parts, but also in the interaction process and user experiences.

The first challenge is to determine **what** messages and information are suitable to be delivered via VUIs. The VUI message could simply be a reminder of taking medications on time, and the notification time can be set by users or physicians. Or the message could contain information such as users' daily exercise records and be used to encourage

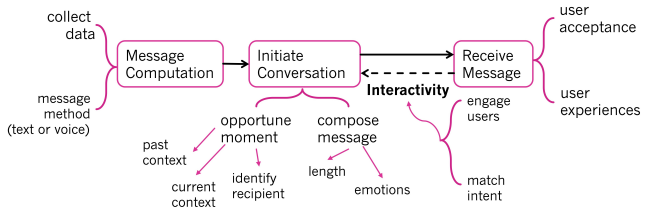

Figure 5: Challenges in different stages.

users at a certain point. We hypothesize that plenty of data collected from wearable devices, smart home applications, and the user's personal information and individual preferences can be stored in the backend database. What data are relevant and how they will be transmitted to the backend system and be used to produce the proper message for specific users are of great importance. Additionally, there can be mixed ways of delivering information proactively, and determining what type of messages are suitably delivered through voice is critical.

Once the information is determined in the backend system, the second major challenge is to identify the **opportune moment (when)** for message delivery. Determining the opportune moment typically requires timely context-awareness. However, current smart speakers lack enough sensors to detect users' context at home. Thus further work is needed to discover what sensors are useful for smart speakers to identify the correct message recipient and find out whether they can be approached. Cameras might be helpful for person identification, yet the sensitivity of image data needs to be taken into account as well. Also, regarding many people stay outside of the home for work and running errands during the daytime, non-timely contexts such as their moods, physical fatigue level should be used in computing the opportune moment.

The third challenge lies in the **interaction process** between proactive smart speakers and users. As speakers take the responsibility to initiate a conversation, it is critical to compose the message at a suitable length, and the content should be concise and easy for users to follow. The voice can contain emotion and be more "human-like" so that the conversation is natural. Besides, when users respond to the proactive speakers and the interaction continues, the intent identification might be difficult. For current non-proactive speakers, users usually activate the speaker with a goal in mind, and it is easy for users to tell whether their intents are identified correctly. However, the dialog model for a conversation initiated by speakers is different. For proactive smart speakers, there are more possibilities and variations in users' responses. It is worthwhile to consider what strategies can be used to build up users' trust towards such voice interaction, and increase engagement, and potentially reduce the chance of having mismatched intent errors when conversation progresses and deepens.

Regarding the user experience, proactively initiating conversations and delivering voice messages can be a big leap for current non-proactive smart speakers, and the **user acceptance** of such a way of interaction might not be high. How to guide users into using proactive speakers and having expectations that the speakers can talk without being activated is unknown. During a speaker-initiated conversation, which can mostly be task-oriented, it is possible that users feel confused in the beginning of the conversation. People's views towards smart speakers might change, and how people posit smart speakers in their daily life is shaping future VUI applications.

Lastly, **privacy issues** occur in every part of proactive smart speaker conversations. Personalization and personal data storage require security checks. Voice fingerprint might also

need to be used to identify each user so that sensitive data will not be hacked. When implementing sensors on speakers to find opportune moments, cameras and microphones need to be set at a low sampling rate, and all algorithms should be run locally. The location of smart speakers also matters considering the constant check of opportune moments.

Conclusion

In summary, we propose that proactivity in smart speakers can be used to develop potentially optimal intervention systems for chronic disease management. Patients can obtain self-management prevention knowledge through having engaging voice interactions with proactive smart speakers. As existing smart speakers are mostly not practically proactive, we propose a possible way to add the proactivity to commercial models. Also, we discuss future opportunities and prominent challenges brought by proactive smart speakers in detail.

REFERENCES

[1] Thomas Bodenheimer, Kate Lorig, Halsted Holman, and Kevin Grumbach. 2002. Patient self-management of chronic disease in primary care. *Jama* 288, 19 (2002), 2469–2475.

[2] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, and others. 2019. What makes a good conversation? challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[3] Michael H Cohen, Michael Harris Cohen, James P Giangola, and Jennifer Balogh. 2004. *Voice user interface design*. Addison-Wesley Professional.

[4] Katie Coleman, Brian T Austin, Cindy Brach, and Edward H Wagner. 2009. Evidence on the chronic care model in the new millennium. *Health affairs* 28, 1 (2009), 75–85.

[5] Eric Corbett and Astrid Weber. 2016. What can I say? addressing user experience challenges of a mobile voice user interface for accessibility. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 72–82.

[6] Russell E Glasgow, Martha M Funnell, Amy E Bonomi, Connie Davis, Valerie Beckham, and Edward H Wagner. 2002. Self-management aspects of the improving chronic illness care breakthrough series: implementation with diabetes and heart failure teams. *Annals of Behavioral Medicine* 24, 2 (2002), 80–87.

[7] Clifford Nass, Ulla Foehr, Scott Brave, and Michael Somoza. 2001. The effects of emotion of voice in synthesized and recorded speech. In *Proceedings of the AAAI symposium emotional and intelligent II: The tangled knot of social cognition*. AAAI North Falmouth, MA.

[8] Brian Oldenburg, C Barr Taylor, Adrienne O'Neil, Fiona Cocker, and Linda D Cameron. 2015. Using new technologies to improve the prevention and management of chronic conditions in populations. *Annual review of public health* 36 (2015), 483–505.