Developing the Proactive Speaker Prototype Based on Google Home

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

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

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

ABSTRACT

Smart speakers and conversational interfaces increasingly make it into consumer's homes. Listening to users' commands assigns them a rather passive role. Proactive speakers, on the other hand, have the potential to empower a broad range of applications such as context-aware health interventions and self-tracking systems. To achieve proactivity, however, requires the speaker to become context-aware and be able to detect opportune moments to initiate interactions. In this work, we present a proactive speaker prototype based on Google Home to investigate interaction contexts. We, additionally, built a voice-based Experience Sampling Method (ESM) application to study contextual factors that are correlated to opportune moments for device-triggered interactions. Preliminary results from a three-week field study with 7 participants indicate the proposed prototype is a robust way to achieve proactivity and implement voice-based ESM. We present use cases for future research and applications for proactive smart speakers in the context of digital health.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Voice Assistants, Interruptibility, Voice User Interface, Smart Speaker

ACM Reference Format:
Jing Wei, Tilman Dingler, and Vassilis Kostakos. 2021. Developing the Proactive Speaker Prototype Based on Google Home. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3411763.3451642

1 INTRODUCTION

With improvements in speech recognition, voice user interfaces (VUI) have become more accessible and ubiquitous. Voice assistants, such as Siri and Cortana, are installed on smartphones and computers to assist users in completing tasks with voice input. In recent years, VUIs have commercially taken off in people's homes in the form of smart speakers. They are used to play music, check the weather or read the latest news. Popular speakers, such as Google Home and Amazon Echo, have gradually become the central hub of smart homes as those speakers provide users VUI controls over smart appliances. To control smart speakers, users usually issue voice commands consisting of two parts: an activation keyword (e.g., *Hey Google* or *Alexa*) and the actual command (e.g., *play music on Spotify*). Almost all commercial smart speakers can only react to users' commands when they are activated. In other words, current smart speakers are passive and do not have the ability to proactively initiate conversations with users.

Proactive features for Google Home and Amazon Echo are currently limited to sending voice reminders or setting off alarms. Such interactions merely transfer smartphone notifications to speakers and usually do not require users' voice inputs. Conversations, however, are bi-directional, and in certain circumstances, it is beneficial to have smart speakers initiate the conversation with their users. The inability to proactively talk to humans currently limits the design space and range of applications that can be developed for smart speakers. For example, health-related intervention systems or self-tracking systems can be deployed on smart speakers. Those systems usually involve sending users reminders and even recommendations at pre-specified or appropriate times [13] in order to query information for which proactivity is a prerequisite.

One major factor should be taken into consideration when designing proactive speakers: 1) timing and 2) context. Proactive services, such as push notifications, often have an interruptive element and delivering notifications at opportune moments can minimize that interruption [14]. Compared with smartphones, smart speakers are more interruptive: voice is more intrusive as it can hardly be ignored (unlike a silent alarm or vibration in the user's pocket). It is, therefore, important to investigate under which circumstances smart speakers should initiate a conversation. For example, a proactive speaker should sense the proximity of users.

Another factor is the actual design of proactive dialogues. The way to design dialogues for proactive VUI applications may differ from the way to design passive VUI applications. For passive speaker applications, it is often assumed that the user has an intent when speaking. For proactive speaker applications, however, the speaker, rather than its user, has an intent. Such intent can either be to acquire user data or deliver timely information. Compared to using questionnaires, using speakers to ask questions to collect data, the dialogue might need to be shorter as listening is generally slower than reading [6]. Plus, the proactive speaker might need to ask questions to get confirmation from users after announcing a message.

CHI '21 Extended Abstracts, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8095-9/21/05…$15.00
https://doi.org/10.1145/3411763.3451642

There is great potential in the developing applications for proactive smart speakers, especially with regard to health apps. Medication adherence, for example, needs to be proactively managed. With passive speakers, users must remember to activate the speaker and then report their medication intake. Proactive speakers, on the other hand, can trigger just-in-time voice reminders based on contextual factors, such as time, user presence, and noise levels, to improve patients' medication adherence. Such applications can also collect information from their users by delivering a series of questions. Henceforth, we aim to find **when** and **how** proactive speakers should talk to their users. Current commercial speakers do not deliver proactive messages. We, therefore, introduce our "workaround" method, in which we externally activate Google Home and turn it into a proactive speaker. A Google action is being developed to employ the Experience-Sampling-Method (ESM) [4] to request contextual information from users. Through the "workaround" method, Google Home can be activated to deliver the voice ESM prompts with an 88.6% activation rate. We present initial data insights from our on-going user study, which contribute to the design of proactive speakers and their application for experience sampling and triggering interactions more generally.

2 BACKGROUND

Smart speakers provide VUI and have many advantages for interaction. For example, having conversations is a natural way for people to communicate, and non-tech-savvy people can easily learn to use VUI. Bentley et al. studied the long-term usage of smart speakers and found that people tended to use specific types of commands at particular times of day [1]. People's patterns of using certain voice commands provide insights for designing the prompting schedules for proactive speakers. For instance, a proactive speaker can tell users about the weather in the morning.

As aforementioned, Google Home and Amazon Echo provide limited proactive features. Google Home has a setting that allows users to turn on notifications on the speaker so that it can deliver proactive reminders for upcoming meetings based on the user's calendar. Amazon Echo will display different light patterns for incoming notifications or messages [17]. LifePod [1], a smart speaker dedicated to older adults, can send proactive reminders to users of medications and doctor appointments. While it is unknown the exact functionality of LifePod, the proactive reminders seem to be pre-scheduled routines and are not prompted based on context and users' interruptibility.

Experience Sampling Method is a method that involves asking participants to report their behaviours or thoughts multiple times throughout the day [4]. ESM is used by many studies to learn about user interruptibility to smartphone notifications. Usually, participants receive notifications for filling out a questionnaire at a scheduled time from smartphones [12, 18]. Fogarty et al. implemented an ESM study that used an audio prompt to ask subjects to report interruptibility in the office [5]. Compared to smartphone notifications, proactive speakers interrupt users through a different modality - *voice*. It is suggested that people have different preferences for the level of proactivity for voice assistants under different situations [10]. Two recent studies [3, 9] investigated participants' interruptibility towards proactive speakers using the ESM approach. Specifically, Komori et al. used a depth camera to detect the activity transition of users and a laptop as the VUI interface. Intermittently, the laptop asked participants, "Do you have a minute?" and participants replied with their availability at that moment. Cha et al. [3] developed a proactive speaker prototype based on a smartphone and a commercial Bluetooth speaker. The application was installed on the smartphone to schedule a voice-based ESM. An audio question, "Is now a good time to talk?" is prompted through the Bluetooth speaker approximately every 20-min to ask about participants' availability. The smartphone application simply recorded participants' responses and provided no feedback. The major drawback of those proactive speaker systems is that such "conversation" is not actually interactive and bi-directional in nature. While their findings can contribute to finding the opportune moments for speakers to talk, they have not fully explored the interaction between users and proactive speakers.

[1]https://lifepod.com/

3 PROTOTYPE DEVELOPMENT

Proactive speakers can be quite useful in implementing various applications ranging from ESM study tools to health management systems, yet programmable proactive speakers do not exist so far. Since proactive speakers should initiate conversations or deliver voice prompts, they need to know about the usage context to determine the right moment to initiate a conversation. To make a passive speaker proactive and collect data using voice ESM prompts to study contextual factors, we propose a proactive speaker prototype based on Google Home.

3.1 Hardware

Google Home requires an activation command to start a conversation. To make Google Home proactive, we created a workaround where the speaker is activated to a "hidden" activation command that is inaudible to users. We built a system where the conversation is initiated by a voice command outputted from earphones. Google Home, in return, responds to the initiation and seemingly triggers a conversation with its users. We use **pre-recorded** voice invocation commands to replace users' speech and activate the speaker in a silent way that nearby users cannot hear.

Figure 1 shows our prototype with two microphone ports on top of Google Home. To enable the silent activation, we designed a 3D-printed earphone holder and attached the earphones to the microphone ports with Blu tack. The earphones can be supported by any device with a 3.5mm jack. This system uses a Raspberry Pi as the controller. **Pre-recorded** voice invocation commands can be issued using a Python script we created. Additionally, our goal is to explore useful environmental contextual information that is correlated with people's interruptibility. While existing literature has suggested that light and noise can indicate people's interruptibility towards push notifications [11, 12], it is unknown whether those factors can indicate people's interruptibility towards voice prompts. Therefore, we further augmented the system with a light sensor to measure the ambient light levels and a USB microphone to detect ambient noise levels. The microphone further allows us to record the real-time interaction between users and the speaker for the purpose of this investigation. Additionally, the Raspberry Pi tracked user

Figure 1: The Proactive Speaker Prototype.

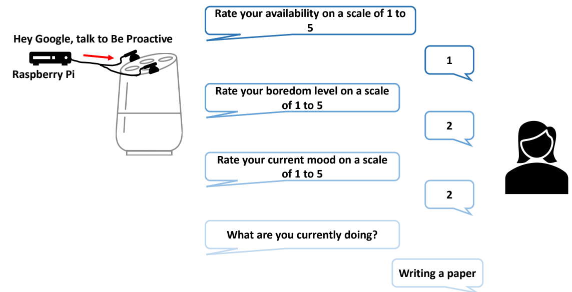

Figure 2: Conversation flow between Raspberry Pi, Google Home, and a user.

proximity by utilizing its Bluetooth scanner to search for the RSSI of participants' smartphones' Bluetooth.

3.2 Google Action

We developed a Google action called *Be Proactive* to implement a proof-of-concept voice-based ESM application. This Google action has three types of invocation: (1) *Hey Google, talk to be proactive*, (2) *Hey Google, talk to be proactive for sound test*, and (3) *Hey Google, talk to be proactive for ping pong test*. Since we were interested in finding the proper way for a proactive speaker to initiate conversations, these three invocations can trigger the ESM with different starters: The first invocation triggers voice prompts directly (baseline starter, as the control condition), the second triggers voice prompts started with a 4-second music clip (Earcon starter [2], simulate a push notification sound), and the third triggers voice prompts started with an utterance - *Hey, are you available?* first (simulate interactions between two humans). For the utterance ESM, the subsequent questions will only be prompted if participants answer "yes", and the speaker will respond *Let's try again next time* if participants answer "no". During the development of the prototype system, we found that the invocation needs to be carefully designed to ensure that it is robust against ambient noise.

Once the action is invoked, the speaker asks users four questions: (1) *Rate your availability on a scale of 1 to 5* (we use availability as the ground truth for opportune moments), (2) *Rate your boredom level on a scale of 1 to 5* (boredom has been shown to be a proxy for opportune content delivery [16]), (3) *Rate your current mood on a scale of 1 to 5* (good mood may suggest good availability [15]), and (4) *What are you currently doing?* These questions are designed to learn contexts that might be indicative of people's interruptibility to voice prompts. Users need to give numerical answers to the first three questions and provide open-ended responses to the last question. An example of the conversation flow with the baseline starter between Raspberry Pi, Google Home, and a user can be seen in Figure 2. If users fail to issue a response or the speech is not correctly recognized, the speaker will prompt: *Please answer number or you can come closer to me.* This fallback prompt is given up to three times until the action ends with the error prompt: *Sorry, I can't help.* Compared to proactive VUI systems developed in [3, 8, 9], our prototype also allows users to engage in turn-taking conversations.

4 FIELD STUDY

To investigate **when** and **how** proactive smart speakers should initiate conversations with users, we conducted a three-week-long study. Participants with previous experiences in smart speakers and proficient in English were recruited. During the study, Raspberry Pi was scheduled to activate Google Home approximately every hour between 9 AM to 11 PM (by default) and allowed participants to customize their prompt schedules. At the same time, Raspberry Pi recorded the ambient noise level using the USB microphone and the ambient light level using the light sensor and measured the RSSI of the Bluetooth signal of participants' smartphone. Additionally, the Raspberry Pi recorded a 90-second audio snippet whenever it was scheduled to play **pre-recorded** commands to Google Home. Those audio recordings allow us to track unsuccessful activation and interaction errors. Three prompt starters were switched in different weeks, and we randomized the starter order among participants. After each week, a questionnaire about user experience with the specific starter was sent to participants. Once the three-week study has completed, we conducted an exit interview with each participant over zoom.

We advertised our study through a university platform. People can register their interests through a Google form. Once recruited, we conducted a zoom orientation session with each participant to teach them how to set up the Raspberry Pi and instruct them how to answer the voice ESM prompts. This study is still running, and we aim to recruit 15 participants in total.

5 PRELIMINARY RESULTS

So far, we recruited 7 participants (6 males) for the study. The mean age of them was 26.8 ± 6.0 years (range 19 to 38). During the study, we encouraged participants to answer as many ESM prompts as they can while maintaining their usual lifestyle.

5.1 Activation Rate

All participants reported their property types, and it was found that they live in households with different sizes and floor plans. Since

Table 1: The Google Home activation rate and the responses rate of three different starters for each participant.

ID	Baseline		Earcon		Utterance	
	Activation Rate	Response Rate	Activation Rate	Response Rate	Activation Rate	Response Rate
P01	1	0.602	0.990	0.5	0.965	0.486
P02	0.897	0.219	0.625	0.5	0.982	0.446
P03	0.915	0.355	0.875	0.366	0.901	0.312
P04	0.680	0.429	0.770	0.418	0.667	0.443
P05	0.828	0.195	0.882	0.289	0.875	0.214
P06	0.942	0.444	0.990	0.378	1	0.165
P07	0.990	0.423	0.991	0.568	0.981	0.452
Average	0.893	0.381	0.875	0.431	0.910	0.360

different noise conditions can happen during the study period, the external earphone-method to activate Google Home may not succeed every time in real-life situations. It is important to have a high activation rate in order to make Google Home proactive. Therefore, we verified the number of Google Home activation using the audio recordings and calculated the activation rate for three invocations. Additionally, we identified potential reasons for unsuccessful activation from the audio recordings, such as people talking in the background or music playing in the background.

The Google Home successful activation rate and the response rate for each participant are shown in Table 1. The average activation rate is 0.893 for the baseline starter, 0.875 for the Earcon starter, and 0.910 for the utterance starter. Among 7 participants, the overall activation rate of P04's Google Home was the lowest: 0.680 for baseline, 0.770 for Earcon, and 0.667 for utterance. It is also noticed that the activation rate for the Earcon starter was only 0.625 for P02. After checking the audio recordings, it was found that unsuccessful activations for P02 were due to unrecognized voice commands, and Google Home prompted the error message: *Sorry, I don't understand* or *Sorry, I didn't understand* in German somehow.

There are several reasons that can cause an unsuccessful activation of Google Home: 1) Google Home is muted manually, 2) Google Home has no WiFi connection, 3) the speaker is in a noisy environment. If Google Home is muted, it cannot be activated. If Google Home loses network connection, the speaker will prompt a message every time it is spoken to - check connection settings in the google home app. During the study, we found that P04 was manually muting the speaker for some time and the network connection was also not stable, which might be the reason that caused the low activation rate of P04. Lastly, just like using a normal smart speaker, ambient noise and people talking at the same time can affect speech recognition accuracy. During the study, the activation was not successful sometimes due to noise. In this case, participants would usually receive an error prompt from the speaker (*Sorry, I didn't understand*), or the speaker recognized the pre-recorded audio as an unknown command and prompted nothing. On one occasion, a participant's Google Home was activated but invoked the Google action *G Prayer* instead of *Be Proactive*. Occasionally, the wrong starter would be activated. For example, the pre-recorded audio command *Hey Google, talk to Be Proactive for sound test* might be recognized as - *Hey Google, talk to Be Proactive for songs*. Then, the baseline starter would be activated instead of the Earcon starter. This situation only happened less than 10 times during the study.

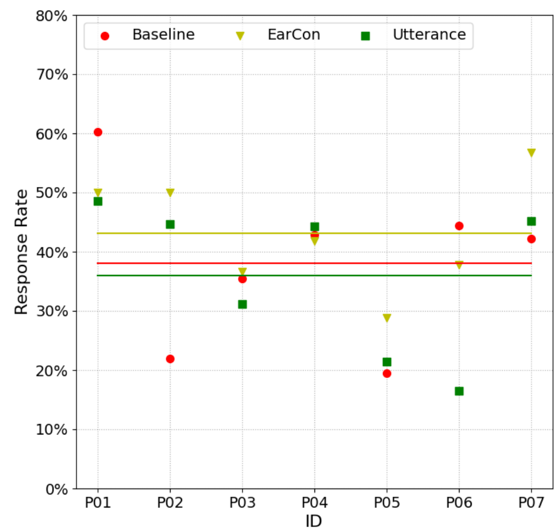

Figure 3: Average response rate per participant for each starter and starter means.

5.2 Response Rate

In total, 1977 ESM prompts were successfully sent and 771 responses were collected from participants; the response rate is 39.0%. As can be seen from Figure 3, different participants have quite different response rates. The average response rate for the Earcon starter is the highest among the three starters. Our study lasted three weeks and the starter for each week was different. The response rate of the last week was the lowest for 5 out of 7 participants, and a decreasing tendency in response rate can be found for 4 out of 7 participants.

5.3 Feedback

After participants finished the three-week study, we conducted an exit-interview to ask about participants' experiences with the proactive speaker and the study. Here, we summarize the findings of the difficulties encountered by participants when they were interacting with proactive speakers. During the interview, when asked

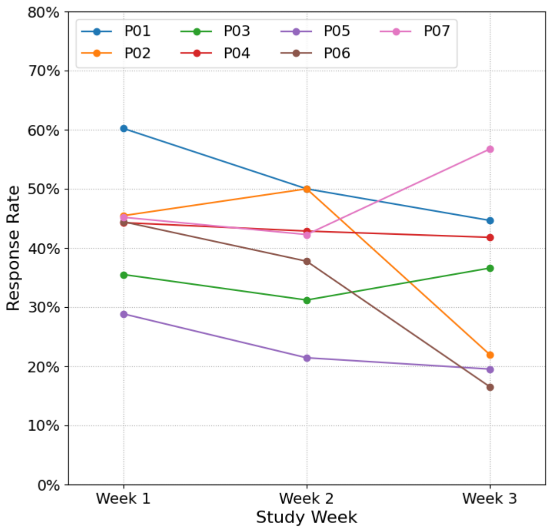

Figure 4: Average response rate over study weeks per participant.

the question *Did you encounter any problems interacting with the proactive smart speaker?*, all participants mentioned that they had had issues with the speech recognition of the speaker. Specifically, four participants mentioned that they needed to get closer to the speaker, and two of them mentioned that they usually needed to increase their volume if the speaker could not recognize their answers. One participant also suggested that background noises can impact speech recognition by Google Home.

Two participants reported that the speaker randomly ended the conversation a few times without giving any prompts even after they gave an answer. Additionally, two participants mentioned that sometimes the speaker did not ask the pre-defined questions but gave random prompts. One participant gave a detailed example and said that the speaker recognized her pronunciation of 'two' as 'tattoo'. Instead of prompting *Please answer numbers*, the speaker started to announce nearby tattoo shops to the participant. Besides the speech recognition problem, two participants also mentioned the difficulty of setting up the prototype. As the Raspberry Pi also takes up an electrical outlet, one participant mentioned that he needed to get an extra outlet to power the whole prototype. Another participant mentioned that she needed to move the location of the prototype during the study several times.

6 DISCUSSION

Activating Google Home using pre-recorded audio commands through a pair of earphones is one way to turn current passive speakers into proactive ones. In this work, we introduced our prototype and evaluated the robustness of this method based on preliminary results from an ongoing field user study. Our results presented in Table 1 show that our workaround is feasible and it opens up opportunities for proactive speaker-related research. In the future, different Google actions can be developed and deployed on this prototype to implement different VUI applications.

However, there are also several limitations of our prototype. Firstly, the physical installation (*i.e., attaching earphones to the microphone ports on Google Home*) is not very robust if users tend to move the speaker prototypes often. We learned that some misalignment in the positioning of earphones could cause a low activation rate; therefore, users need to be careful when sticking earphones to the speaker. As the microphone ports are covered by earphones, it was reported by participants that the speaker seemed to have a shorter range of recognizing speech from users. Also, it is not guaranteed that the speaker and the designed Google action can be correctly activated all of the time. If there are people talking close to the speaker when the Raspberry Pi is just playing the pre-recorded audio command, it is likely that the activation will not be successful. Failed activation can sometimes result in FIN error messages from Google Home, and it may confuse users when they hear those messages at random times. Also, this method will fail when participants mute the speaker manually. Nevertheless, this "workaround" method can work well if the physical set-up is done correctly and users do not manually mute the Google Home.

The response rates of voice ESM prompts in this study tend to be lower than those of smartphone-based ESM studies [19]. Compared to smartphones, which are mobile, current smart speakers are located in a user's home and not mobile. People usually are rarely home during the daytime and therefore are unable to respond to the voice ESM. Additionally, Google Home only can reach a certain range of areas when speaking. If users live in a big house, they sometimes cannot hear the speaker speaking if they are in different rooms. The nature of fixed smart speakers might explain that the slightly higher response rate for the Earcon starter. Compared to the other two starter methods, Earcon consists of 4-second music, which can be more alarming. Also, participants tend to respond less as the study proceeded. However, we will need more participants to confirm those findings.

Lastly, we found a limitation in the interaction with the proactive speaker. As the prototype is essentially based on Google Home, the dialogue flow is designed for passive speakers. Most participants experienced that the speaker stopped talking due to errors in speech recognition resulting in data loss. For passive speakers, it is reasonable that the speaker stops talking if it hits the fallback response more than three times, as users usually will reorganize their speech and try another time [7]. For proactive speakers, however, it should not stop due to poor speech recognition, especially if the speaker is designed to collect data from users. Unfortunately, this limitation is hard to overcome in our prototype design.

7 FUTURE WORK

In this paper, we focused on introducing a proactive speaker prototype and a voice-based ESM user study. We presented preliminary results of the robustness of the proposed prototype and the response rate for 7 participants. A further and more in-depth analysis will be conducted once more participants have completed the study, and all data have been collected. Specifically, we will aim to find contextual factors that are correlated to users' interruptibility towards the proactive speaker and the suitable starter method for participants. The ESM data will serve as ground truth to detect opportune moments for triggering user interactions. We aim to train machine-learning models based on sensor and ESM data and provide insights into both environmental and personal contextual factors. Additionally, with a more systematic analysis of audio recordings capturing the conversation between participants and the proactive speaker, we also aim to uncover interaction errors and user behaviours when handling those errors.

The prototype we developed poses as a workaround for researchers to explore aspects of proactive voice assistants. This enables future work on various sensor and user contexts as well as privacy implications and user concerns. Our ESM study will pave the way towards a blueprint to conduct experience sampling studies for ground truth collection. Potential applications, such as health intervention systems or self-tracking systems, can subsequently be developed as Google actions and be deployed on our prototype. We expect our prototype to spark more research and application opportunities to make smart speakers more functional.

REFERENCES

[1] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the long-term use of smart speaker assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–24.

[2] Meera M. Blattner, Denise A. Sumikawa, and Robert M. Greenberg. 1989. Earcons and Icons: Their Structure and Common Design Principles. *Human–Computer Interaction* 4, 1 (1989), 11–44. https://doi.org/10.1207/s15327051hci0401_1

[3] 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. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–28.

[4] Sunny Consolvo and Miriam Walker. 2003. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing* 2, 2 (2003), 24–31.

[5] James Fogarty, Scott E. Hudson, and Jennifer Lai. 2004. Examining the Robustness of Sensor-Based Statistical Models of Human Interruptibility. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 207–214. https://doi.org/10.1145/985692.985719

[6] Jonathan Grudin. 1988. Why CSCW Applications Fail: Problems in the Design and Evaluationof Organizational Interfaces. In *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work* (Portland, Oregon, USA) *(CSCW '88)*. Association for Computing Machinery, New York, NY, USA, 85–93. https://doi.org/10.1145/62266.62273

[7] Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How Do Users Respond to Voice Input Errors? Lexical and Phonetic Query Reformulation in Voice Search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Dublin, Ireland) *(SIGIR '13)*. Association for Computing Machinery, New York, NY, USA, 143–152. https://doi.org/10.1145/2484028.2484092

[8] Auk Kim, Jung-Mi Park, and Uichin Lee. 2020. Interruptibility for in-vehicle multitasking: influence of voice task demands and adaptive behaviors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–22.

[9] Mitsuki Komori, Yuichiro Fujimoto, Jianfeng Xu, Kazuyuki Tasaka, Hiromasa Yanagihara, and Kinya 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*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 527–539.

[10] Michal Luria, Rebecca Zheng, Bennett Huffman, Shuangni Huang, John Zimmerman, and Jodi Forlizzi. 2020. Social Boundaries for Personal Agents in the Interpersonal Space of the Home. 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/10.1145/3313831.3376311

[11] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing Content-Driven Intelligent Notification Mechanisms for Mobile Applications. 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, 813–824. https://doi.org/10.1145/2750858.2807544

[12] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding People's Receptivity to Mobile Notifications. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1021–1032. https://doi.org/10.1145/2858036.2858566

[13] Inbal Nahum-Shani, Shawna N Smith, Bonnie J Spring, Linda M Collins, Katie Witkiewitz, Ambuj Tewari, and Susan A Murphy. 2018. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine* 52, 6 (2018), 446–462.

[14] Chunjong Park, Junsung Lim, Juho Kim, Sung-Ju Lee, and Dongman Lee. 2017. Don't Bother Me. I'm Socializing! A Breakpoint-Based Smartphone Notification System. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 541–554. https://doi.org/10.1145/2998181.2998189

[15] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond interruptibility: Predicting opportune moments to engage mobile phone users. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–25.

[16] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When Attention is Not Scarce - Detecting Boredom from Mobile Phone Usage. 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, 825–836. https://doi.org/10.1145/2750858.2804252

[17] David Priest. 2020. What all those lights on your Amazon Echo mean. https://www.cnet.com/how-to/what-do-the-light-ring-colors-on-your-amazon-echo-mean/

[18] 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

[19] Niels van Berkel, Chu Luo, Theodoros Anagnostopoulos, Denzil Ferreira, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2016. A Systematic Assessment of Smartphone Usage Gaps. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4711–4721. https://doi.org/10.1145/2858036.2858348