Poster: Toward Context-aware Proactive Conversation for Smart Speakers

Soowon Kang
sw.kang@kaist.ac.kr
KAIST
South Korea

Heepyung Kim
heepkim@kaist.ac.kr
KAIST
South Korea

Youngtae Noh
ytnoh@inha.ac.kr
Inha University
South Korea

Uichin Lee*
uclee@kaist.ac.kr
KAIST
South Korea

ABSTRACT

Despite the increasing popularity of novel conversational services for smart speakers, the current ones have supported limited proactive interactions. In this work, we focus to understand opportune moments for proactive conversational interactions in domestic contexts. To achieve our goal, we built a speech-based experience sampling device and performed a 1-week field study with 40 students living in university dormitories. From 3,572 in-situ experiences, we identified eleven categories to find contextual features related to opportune moments. We showed that the key determinants for interruptibility are relevant to personal contextual factors, user mobility, and social presence. Through considering the aforementioned factors, we envision that a smart speaker can intelligently manage the timing of conversations.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; **User interface management systems**.

KEYWORDS

Smart Speakers, Conversational Interaction, Interruptibility

ACM Reference Format:
Soowon Kang, Heepyung Kim, Youngtae Noh, and Uichin Lee. 2021. Poster: Toward Context-aware Proactive Conversation for Smart Speakers. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3460418.3479306

1 INTRODUCTION

Recently, top shares in the smart speaker market (e.g., Amazon Echo [9], and Google Home [11]) started to pay attention to the proactive services, while the current smart speakers have generally supported reactive services via voice commands. Although proactive services allow the users to obtain useful information, prior studies also informed that timing and relevance of the services are crucial to the user experience [1, 3]. To the best of our knowledge, the interruptibility studies of smart speakers in home contexts lacks depth in spite of the popularity of proactive services in smart speakers. In this study, we explained the opportune moments for the proactive services of smart speakers in diverse activities at home.

To collect in-situ experiences in home environment, we implemented a smart speaker that supports recording user contexts and responses via a voice-based experience sampling method (ESM). Our smart speaker asked users to answer the question, "IS now a good time to talk?" in yes or no, with detailed contextual information. The 3,572 ESM responses were collected from 20 pairs (n=40) living in dormitories. We examined the responses by conducting the exit interview to understand the core factors that affects interruptibility related to home activities. Thus, we identified the following factors relevant to the user interruptibility of conversational interactions: (i) personal contextual factors (e.g., engagement, urgency, psychological/physical states, auditory/verbal channel availability), (ii) movement-related factors (e.g., entrance/departure behaviors, activity switching), and (iii) social presence (e.g., roommate's current activity). Our findings emphasize that the proactive conversation management is required to the proactive conversational services for smart speakers.

2 BACKGROUND

Traditional smart speakers have mainly provided reactive services. The recent commercial smart speakers, including Amazon and Google, started to look into several proactive services. Amazon Echo supported friendly reminders (e.g., informing to lock the door) and ambient sensing to warn the user of safety issues (e.g., window break) [9]. Google Home can notify the upcoming events that depend on the user's Google Calendar [11].

These prior research on smart speakers supported limited chances to proactively interact with users. One important factor of intelligent services is detecting the appropriate time to open the conversation session. Researchers have been studied the interruptibility in wide contexts (e.g., smartphone [8], in-vehicle [7]). However, home environment is more challenging since it is more dynamic compared to other environments due to multi-user, mobility factors,

*This is the corresponding author.

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA

and numerous domestic routines and activities. In this paper, we focus how user contexts are related to understand interruptibility for proactive services.

3 METHODOLOGY

In this section, we describe the design procedures of our voice-based ESM system, and the corresponding dataset collected by the ESM system. Note that this study was approved by the Institutional Review Board (IRB) of the university.

We built a smart speaker device that is made up of a smartphone, Bluetooth speaker, wide-angle lens, music stand, and enclosure box. The smartphone was utilized to generate the sound signals at the scheduled times, detect the level of movements in the camera angle, and record the voice responses of users. We paired a Bluetooth speaker with the smartphone to support the similar form factor of the smart speaker and the enough range of the sound volume. We also attached an extra wide-angle lens on the smartphone to capture the users' movements accurately. To emulate a common smart speaker, a paper-based enclosure box was used to encase the all aforementioned components. The assembled box with the stand was located at the center of the room so that the camera can detect the user movements within the whole room.

We implemented an ESM application to observe the user contexts of various situations that provides the following features: (i) controlling the sound volume, (ii) adjusting the working hours, (iii) configuring the threshold of movement detection function, (iv) powering on and off the data collection. We also focused to find opportune moments for a conversation as prior studies probed interruptibility by asking several questions [5, 6, 10]. Thus, our ESM app asked a question (i.e., "Is now a good time to talk?") to the study participants. Depend on situations at the ESM prompt, the participants verbally described their status and their activities.

The ESM prompts were triggered as following two rules: (i) random time, (ii) movement detected time. The random time rule reserved the prompts randomly in the waking hours with a random interval between 15-25 minutes. At the movement detected time, the prompts were triggered immediately. To prevent too frequent prompts, our ESM app intentionally disabled the next ESM prompts for 15 minutes if one prompt was triggered.

We recruited 20 student pairs (N=40, M:F=24:16) who lived in the dormitories. We requested them to participate the 1-week data collection for three hours per a day, so that each participant could record at least 10 ESM responses per a day. After the data collection week, we conducted the 20-minute exit interviews with all participants to clearly understand several misleading ESM responses.

4 RESULTS

A total of 3,572 ESM responses were collected from 1-week data collection, but we excluded 72 responses that did not clearly meet our goal. To find contextual factors for interruptibility in home activities, three researchers created 19 unique activities by performing the affinity diagramming [4]. Then we categorized the valid 3,500 responses into 11 categories (e.g., about to leave, doing chores, eating, just returned, resting, self-caring, sleeping, social interaction, studying/working, using media, miscellaneous). Except for the miscellaneous (n=49, 1%), the participants reported that the most frequently captured category was *using media* (n=1124, 32%), whereas the least frequently captured category was *doing chores* (n=66, 2%). They also reported that 53% of responses were interruptible. The most interruptible case was *just returned* (95%), whereas the least interruptible case was *studying/working* (21%).

We revealed that three key contextual factors were related to the interruptibility of smart speaker-based proactive services in domestic environment: (i) personal contextual factors, (ii) the movement-related factors, and (iii) the social presence factors. Regarding personal factors, we found concentration and engagement, urgency and busyness, psychological/physical states, and auditory/verbal channel availability. Regarding movement factors, we found departure, entrance, and physical activity transitions. Interestingly, the communication range is an important aspect of movement factors. Only 18% of the outbound movements from the speaker (departure) were interruptible, whereas 97% of the inbound movements (entrance) were interruptible. We found social presence also influenced interruptibility since roommates wanted to minimize the possible interpersonal conflict such as disturbing their roommate during sleeping or studying.

5 CONCLUSION

To observe the opportune moments for smart speaker-based proactive services, we conducted 1-week field study with 40 participants and collected 3,572 in-situ experience responses. As a result, several unique factors related to daily routines at home were identified. We concluded that proactive services in the smart speakers should carefully consider the temporal and spatial patterns to support context-aware proactive conversational interactions. We call for further studies on conducting follow-up research and resolving technical issues for managing proactive interactions.

ACKNOWLEDGMENTS

This paper originated from the author's previous study [2]. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean government (MSIT) (2020R1A4A10187747).

REFERENCES

[1] S. Amershi, D. Weld, M Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. N. Bennett, K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems* (Glasgow, Scotland Uk). ACM, Article 3, 13 pages.

[2] N. Cha, A. Kim, C. Y. Park, S. Kang, M. Park, J.-G. Lee, S. Lee, and U. Lee. 2020. Hello There! Is Now a Good Time to Talk? Opportune Moments for Proactive Interactions with Smart Speakers. *Proc. of the ACM on IMWUT* 4, 3 (2020), 1–28.

[3] A. P. Chaves and M. A. Gerosa. 2020. How Should My Chatbot Interact? A Survey on Social Characteristics in Human–Chatbot Interaction Design. *Int. J. of Human–Computer Interaction* (2020), 1–30.

[4] J. Y. Cho and E.-H. Lee. 2014. Reducing Confusion about Grounded Theory and Qualitative Content Analysis: Similarities and Differences. *The Qualitative Report* 19, 32 (2014), 1–20.

[5] J Fogarty, A. J. Ko, H. H. Aung, E. Golden, K. P. Tang, and S. E. Hudson. 2005. Examining Task Engagement in Sensor-based Statistical Models of Human Interruptibility. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems* (Portland, Oregon, USA). ACM, 331–340.

[6] J. Hernandez, D. McDuff, C. Infante, P. Maes, K. Quigley, and R. Picard. 2016. Wearable ESM: Differences in the Experience Sampling Method Across Wearable Devices. In *Proc. of the 18th Int. Conf. on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy). ACM, 195–205.

[7] A. Kim, W. Choi, J. Park, K. Kim, and U. Lee. 2018. Interrupting Drivers for Interactions: Predicting Opportune Moments for In-vehicle Proactive Auditory-verbal Tasks. *Proc. of the ACM on IMWUT* 2, 4, Article 175 (Dec. 2018), 28 pages.

[8] M. Pielot, B. Cardoso, K. Katevas, J. Serrà, A. Matic, and N. Oliver. 2017. Beyond interruptibility: Predicting opportune moments to engage mobile phone users. *Proc. of the ACM on IMWUT* 1, 3 (2017), 1–25.

[9] B. F. Rubin and Ry Crist. 2019. *What Amazon's Alexa will tell us in 2019.* https://www.cnet.com/news/what-amazon-alexa-will-tell-us-in-2019/

[10] R. Semmens, N. Martelaro, P. Kaveti, S Stent, and W. Ju. 2019. Is Now A Good Time?: An Empirical Study of Vehicle-Driver Communication Timing. In *Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems* (Glasgow, Scotland Uk). ACM, Article 637, 12 pages.

[11] TechRepublic. 2019. *Google Home: Cheat sheet.* https://www.techrepublic.com/article/google-home-the-smart-persons-guide/