Check for updates

"Hello There! Is Now a Good Time to Talk?": Opportune Moments for Proactive Interactions with Smart Speakers

NARAE CHA, KAIST, South Korea
AUK KIM*, Kangwon National University, South Korea and KAIST, South Korea
CHEUL YOUNG PARK, KAIST, South Korea
SOOWON KANG, KAIST, South Korea
MINGYU PARK, KAIST, South Korea
JAE-GIL LEE, KAIST, South Korea
SANGSU LEE, KAIST, South Korea
UICHIN LEE*, KAIST, South Korea

74

Increasing number of researchers and designers are envisioning a wide range of novel proactive conversational services for smart speakers such as context-aware reminders and restocking household items. When initiating conversational interactions proactively, smart speakers need to consider users' contexts to minimize disruption. In this work, we aim to broaden our understanding of opportune moments for proactive conversational interactions in domestic contexts. Toward this goal, we built a voice-based experience sampling device and conducted a one-week field study with 40 participants living in university dormitories. From 3,572 in-situ user experience reports, we proposed 19 activity categories to investigate contextual factors related to interruptibility. Our data analysis results show that the key determinants for opportune moments are closely related to both personal contextual factors such as busyness, mood, and resource conflicts for dual-tasking, and the other contextual factors associated with the everyday routines at home, including user mobility and social presence. Based on these findings, we discuss the need for designing context-aware proactive conversation management features that dynamically control conversational interactions based on users' contexts and routines.

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

Additional Key Words and Phrases: Smart Speakers, Conversational Interaction, Interruptibility

ACM Reference Format:
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 (September 2020), 28 pages. https://doi.org/10.1145/3411810

*This is the corresponding author.

Authors' addresses: Narae Cha, nr.cha@kaist.ac.kr, KAIST, Daejeon, South Korea; Auk Kim, kimauk@kaist.ac.kr, Kangwon National University, Kangwon, South Korea, KAIST, Daejeon, South Korea; Cheul Young Park, cheulyop@kaist.ac.kr, KAIST, Daejeon, South Korea; Soowon Kang, sw.kang@kaist.ac.kr, KAIST, Daejeon, South Korea; Mingyu Park, mininnov99@kaist.ac.kr, KAIST, Daejeon, South Korea; Jae-Gil Lee, jaegil@kaist.ac.kr, KAIST, Daejeon, South Korea; Sangsu Lee, sangsu.lee@kaist.ac.kr, KAIST, Daejeon, South Korea; Uichin Lee, uclee@kaist.ac.kr, KAIST, Daejeon, South Korea.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
2474-9567/2020/9-ART74 $15.00
https://doi.org/10.1145/3411810

1 INTRODUCTION

Smart speakers come with an intelligent voice assistant feature that supports speech-based interactions for a wide range of tasks, including time checking, delivery tracking, and QnA. Currently, smart speaker services are mostly reactive to user's commands. Recently, Amazon and Google with top shares in the smart speaker market started to consider proactive services such as reminding a user's schedule [59] and supporting home safety and security [52]. Prior studies have already demonstrated the usefulness of various proactive services: context-aware reminders/recommendations [10, 57, 63], and self-tracking/reflection for productivity [30, 68] and mental well-being [14, 35].

While proactive services provide useful information for inspiring and engaging users, prior studies also warned that timing and relevance of proactive services are critical to the user experience [5, 11]. Delivering services at an inappropriate moment could disrupt users' primary tasks and cause annoyance and resumption lag [23], or even result in safety risks in some contexts such as driving [26, 27, 55]. Previous interruptibility studies examined the opportune moments for an interruption in diverse task contexts and computing environments, including task-switching with desktop computers in offices [2, 24], notification delivery via mobile phones [38, 47], conversational interactions in vehicles [26, 27, 29], and the actuation of height adjustments by smart furniture [33].

However, to our knowledge, research on the interruptibility of smart speakers in home contexts remains lacking despite the recent rise in the popularity of proactive services in smart speakers. We hypothesize that user interruptibility to smart speakers would have contributing factors different from other devices (e.g., mobile phones and desktop computers) and locations (e.g., office and vehicular setting) due to their distinct characteristics. For conversational interactions with stationary speakers at home, users can be mobile or stationary and simultaneously perform various activities, and the speakers can be shared among family members. Earlier smart home research highlighted the importance of seamlessly supporting domestic routines and key activities [15, 61], and that intelligent services may introduce new challenges and responsibilities in dealing with technology [12] as well as affecting interpersonal relationships [6]. Therefore, we set our goal to understand the opportune moments for smart speakers to initiate proactive services while their users are naturally engaged in diverse activities at home.

To achieve our goal, we designed a smart speaker that supports a voice-based experience sampling method (ESM) and conducted an in-situ week-long study to understand opportune moments for smart speakers to initiate proactive conversational interactions with smart speakers in the home context. In the study, we asked users to answer the question, "Is now a good time to talk?" in yes or no, and further inquired contextual reasoning for the provided answers. We reviewed each response during the exit interview to gain a deeper understanding of factors that influence interruptibility related to home activities. Data were collected for a week from 40 participants living in dormitories (20 rooms with two people per room) where diverse domestic activities naturally occur in a studio setting.

From the collected data, we identified the following key factors relevant to the interruptibility of users to proactive audio service from a smart speaker in the domestic contexts: (1) personal contextual factors (e.g., engagement, urgency, psychological/physical states, auditory/verbal channel availability), (2) movement-related factors (e.g., entrance/departure behaviors, activity switching), and (3) social presence (e.g., roommate's current activity). These factors are closely related to everyday routines, user mobility, and social contexts in home environments. Our findings highlight that smart speakers should support proactive conversation management for proactive conversational services (i.e., when to start, pause, or resume conversation) by carefully and simultaneously considering spatial relationships (or proxemics) [7] and home routines [61]. The contributions of our study are as follows:

- We designed a prototype of a smart speaker that supports a voice-based experience sampling method via two pilot studies.

- We conducted in-situ studies with 20 pairs of roommates (40 participants) living in dormitories, and collected 3,500 responses related to user interruptibility and activity contexts.
- We identified three key factors that affect the interruptibility of proactive services: (1) personal contextual factors, (2) movement-related factors, and (3) social presence.
- Based on our findings, we propose how to design context-aware proactive conversation management system for smart speakers in domestic contexts.

2 RELATED WORK

In this section, we first review prior studies on proactive services by intelligent agents and provide an overview of interruptibility research in the field of ubiquitous computing. As our work focuses on contextual factors related to proactive conversational interactions with smart speakers, we review prior studies that examine contextual factors of interruptibility.

2.1 Proactive Services and Intelligent Agents

Current smart speakers are predominantly reactive and limited in providing proactive services (e.g., asking users to change their schedules). Beyond simple push notifications, researchers of ubiquitous computing have explored various intelligent proactive services, which apply to smart speakers as well. Place-Its [57] is a location-based reminder application for mobile phones that can remind users to, for example, bring their belongings when leaving. Context-aware reminders are useful for users with memory deficiency, such as older adults with dementia [25, 42]. Braunhofer et al. [10] proposed a proactive recommendation system in the tourism domain to suggest relevant items to the target user in the right context; the system can recommend suitable places to visit based on daily weather and recently visited locations. Researchers also explored conversational interactions in smart home environments. The Sweet-Home project [63] built a voice-based intelligent controller to help older adults control appliances in smart homes more easily and to activate proactive services such as prompting a security reminder to lock doors before leaving. Qin et al. [50] explored active interaction design techniques where intelligent agents nudge users to react to interactions initiated by machines, by providing affordance for a follow-up with a question including contextual information (e.g., proactively welcoming users and checking whether the current room temperature is acceptable).

Recently, smart speakers such as LifePod, Google Home, and Amazon Echo have started exploring various proactive services as well. LifePod [65], a smart speaker for older adults, proactively reminds users of medication adherence and doctors' appointments. Google Home provides proactive notifications about time-sensitive information (e.g., upcoming meetings) based on a user's Google Calendar setting. Amazon announced proactive and ambient services for their intelligent agent, Alexa: offering friendly reminders, for example, suggesting to lock the door when a user says "goodnight," or supporting ambient sensing to alert users of security events (e.g., window break or smoke alarm sound) [52].

These prior studies demonstrate the usefulness of proactive services in smart speakers or mobile phones. Such services in mobile phones are also applicable to smart speakers. Similarly, Google smart speaker allows smartphone applications to send notifications (e.g., voice content and services) and to be controlled via the smart speaker if the applications support voice interfaces [18]. As discussed later, one critical aspect of intelligent agents is knowing when to initiate conversations to deliver proactive services. Therefore, in our work, we explore how user contexts are related to finding opportune moments for conversations.

2.2 Interruptibility and Opportune Moments for Interaction

Interruptibility denotes the extent to which a user is interruptible when a system attempts to communicate with them, for example, to deliver a notification. An opportune moment for user interaction refers to a moment when the disruption of interrupting the user's current task is minimal [62]. Interruptibility is a central topic in ubiquitous computing research as interrupting users at inappropriate moments has considerable adverse effects. Inopportune interruption can create annoyance and resumption lag (context-switching cost) in desktop environments [23], and even safety risks in driving contexts [26, 27] (e.g., deterioration of situation awareness [44]). Interruptions at opportune moments can facilitate the engagement of mobile users with the recommended content (e.g., games, news, and surveys) [47].

Previous studies on interruptibility explored various computing devices and environments, ranging from desktop computers and mobile phones to vehicles and smart furniture. For desktop computers, researchers found that opportune moments are when users switch tasks [2]. For mobile notification delivery, both a user's context and notification content were found critical [38]. Recent studies further considered novel, ubiquitous environments. For example, voice interfaces in vehicles can leverage vehicular sensor data to find opportune moments for conversation [26, 27]. Another example is smart furniture; automated height-adjustable desks are effective in maximizing user comfort with automatic adjustment of heights in between tasks to minimize interruption [33].

Despite the recent popularity of proactive services in smart speakers, to our knowledge, interruptibility research on smart speakers in home contexts is insufficient. The smart speaker has several distinct characteristics compared to other devices: (1) it mainly supports conversational interactions, (2) it can be shared by multiple users (e.g., family members) while the voice-interaction channel is also a shared medium, and (3) it is installed at a fixed location (e.g., living room) while users are mobile. Moreover, home environments are much more dynamic compared to environments such as vehicles or offices, due to multi-user, multi-tasking, and mobility factors, and various domestic routines and activities. Therefore, we hypothesize that the combination of the characteristics of smart speakers and home environments engenders unique factors affecting user interruptibility.

2.3 Contextual Factors for Interruptibility

User's personal contextual factors such as current activity [46], engagement [17], busyness [70], and emotion [70] are closely associated with the interruptibility of the user. The current activity of a user is most frequently the foremost determinant of interruptibility. For example, in a notification delivery situation, a user is less interruptible if they are taking on challenging tasks that require a considerable amount of attention, such as biking or driving. Notification content in association with the current activity also affects interruptibility, e.g., in a meeting, a chat notification from a friend is disruptive while an email notification from a project collaborator may be acceptable [46]. As proposed in the work of Siewiorek et al. [56] and now widely used by many commercial time-management and scheduling products such as Google Calendar, this issue of ongoing activities affecting interruptibility is dealt by inferring interruptibility from the users' agenda.

Engagement is another important determinant of interruptibility. When performing simple rote work that is highly engaging but not challenging, users are susceptible to distractions [37]. Meanwhile, Yuan et al. [70] found that notifications from smartphones can be very disruptive when users are busy, and busyness levels vary widely depending on the user context. User's emotions are also a factor in interruptibility; people are more likely to be interruptible when in a pleasant mood than in an unpleasant mood [70]. These studies show that personal contextual factors dictate user interruptibility to having conversational interactions with a smart speaker. Thus, our first research question is: *RQ1) What personal contextual factors are relevant to interruptibility concerning proactive interactions with smart speakers?*

Researchers also found that users' interruptibility further depends on their physical movements. Pejovic et al. [46] considered four physical activities related to interruptibility (being still, being on foot, being on a bicycle, and being inside a vehicle) and observed that the degree of interruptibility differs for each activity. Ho et al. [21] found that a user is likely to be interruptible when transitioning between two physical activities (e.g., from sitting to walking). Based on these observations, we aim to find whether physical movements signal opportune moments for smart speakers to engage with users. Unlike mobile phones in previous studies, smart speakers are stationary while users are mobile, and untethered interactions are feasible with speech. This combination seems to provide unique characteristics to interactions with smart speakers. Thus, our second research question is as follows: *RQ2) What movement-related factors are relevant to interruptibility concerning proactive interactions with smart speakers?*

As discussed in Section 2.2, smart speakers are shared devices accessible by any user in their vicinity; therefore, we also consider social presence factors in our study. Prior studies explored a social relationship in online contexts to understand interruptibility [38]. The social presence of other people in a shared space (e.g., roommates or visitors) affects the interruptibility; Park et al. [43] studied interruptibility of mobile notifications during social activities and found that conversation breakpoints are generally opportune moments for interruptions. In contrast with mobile interactions such as a notification delivery to a person, the broadcasting nature of smart speaker interactions possibly interrupts other users in a shared space. Thus, our third research question is as follows: *RQ3) What social presence factors are relevant to interruptibility concerning proactive interactions with smart speakers?*

3 METHODS

Our goal was to understand the interruptible moments during natural activities at home. We designed a smart speaker that prompted voice-based ESM messages triggered by two mechanisms: randomly and through movement detection. In this section, we explain the design of the ESM procedures and smart speaker systems, recruitment of participants, and data collection and analysis. Our study was reviewed and approved by the Institutional Review Board (IRB) of the university.

3.1 Data Collection Setup

3.1.1 Data Collection Environment. To collect naturalistic data, we aimed to collect data at participants' homes, where they live their daily life. The selected data collection environment was a university dormitory with shared rooms. This arrangement helped us study the effect of the presence of roommates on the participants' interruptibility for voice interaction. The dormitories had rooms with two single beds; other amenities such as restrooms, shower rooms, laundry rooms, and kitchens were shared with other students. As shown in Figure 1, we placed a smart speaker on the opposite side of an entrance door. The distance between a speaker and a door varied with the length of a room. The sizes of the dormitory rooms were as follows ($length \times width$ in meters): 4.3×3.5, 5.4×3, 6.1×3.5. The dormitory environment has its limitations, but it enabled us to capture diverse home-based activities at a single location, which could be typically observed in a family setting.

3.1.2 Apparatus. We built a smart speaker device that consisted of a smartphone running an ESM application, Bluetooth speaker, wide-angle lens, height-adjustable music stand, and paper-based enclosure. We developed an ESM application (Figure 2c) that supports the following features: (1) controlling ESM volume, (2) setting operation hours, (3) configuring movement detection threshold, and (4) starting ESM data collection. The volume level of ESM could be adjusted to prevent unpleasant experiences caused by loud sounds. Setting operation hours helped us customize ESM working hours depending on each participant's daily schedule. The aim of the threshold controller was to increase the accuracy of movement detection, which we explain in detail in Section 3.1.4. Throughout the data collection, the smartphone was set to display configuration information and to automatically synchronize ESM data files with cloud storage in real time.

We paired the smartphone with a Bluetooth speaker. Adding a Bluetooth speaker not only makes the whole device resemble a smart speaker, but also amplifies the sound volume. To accurately detect a user's movements using the smartphone's camera, we attached an extra lens, which supported a 170-degree wide-angle, on the smartphone's built-in camera lens. This lens supports a 170-degree wide-angle. A black paper-based enclosure was

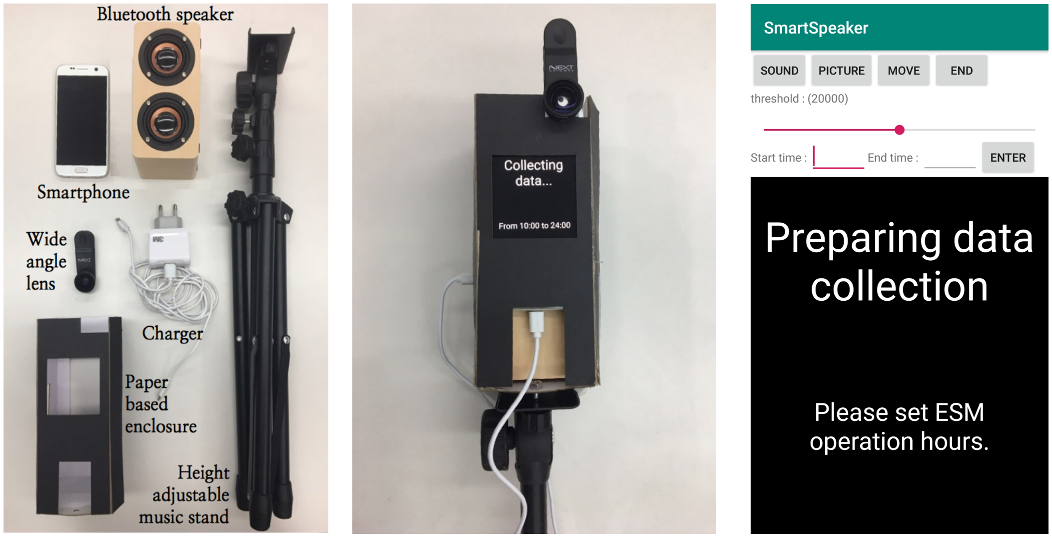
Fig. 1. Configurations of the dormitories with an ESM device installed in the middle

designed to encase the components and to emulate standard smart speakers. This assembled smart speaker was placed on the height-adjustable stand, as the speaker required additional height adjustment for an unobstructed view of the entire room from the camera lens. During data collection, the speaker was located at the center of the room near the inner wall with its height adjusted to eye level to accurately detect movements within the whole room.

3.1.3 ESM Method. The smart speaker placed in the room prompted voice-based ESMs to check user interruptibility and record the context. As the smart speaker supported conversational interaction, ESM was conducted via speech. Prior studies have probed interruptibility via several questions such as whether the timing is good [55], whether it is disruptive [20], or whether a user is available to conduct a requested task [17].

ESM Question: We focused on *timing* for conversational interactions because it can be typically interpreted as an opportune moment for a conversation, reflecting a user's preference and self-assessment of availability. Thus, the following question was used: "Is now a good time to talk?" Similarly, a recent study on proactive audio services in vehicles used a similar voice ESM question, "Is now a good time?" [55] to explore on-road driving interruptibility. We intentionally did not associate a specific service with the ESM question. This enabled us to focus on contextual factors of interruptible moments because we minimized user bias on interruptibility due to content preferences [47].

Proactive Service Scenarios: As a proactive service provided by a smart speaker in home settings, the actual content of a conversation can be diverse. Beyond simple information delivery tasks (e.g., current weather, time), we assume a range of interactive conversational tasks that typically *'require user response (decision-making) and*



(a) Parts of the device (b) Assembled device (c) ESM application

Fig. 2. Smart speaker device for data collection

turn-takings (bi-directional interactions).' At an orientation before data collection began, we explained to the participants this meaning of proactive services and provided them with the following three example scenarios (based on three popular services of Amazon Echo [3, 4]):

- *Scenario #1*: It proactively reminds that you are running low on grocery essentials (e.g., toilet paper), and asks you to select an essential(s) of interest for a recommendation (e.g., most popular products, new products) or automatic online ordering.
- *Scenario #2*: It proactively informs you of daily promotions/sales of interest, and asks you to select a deal(s) of interest for details or automatic online ordering.
- *Scenario #3*: It proactively informs a new movie release and roughly describes the story, and asks you to select a time for automatic online booking after telling you possible showtimes based on your schedule.

We expect that these types of services at home are continuing to increase. For reliable data collection, we also sent a daily reminder to our participants every morning via a text message: *"Please keep in mind that 'Is now a good time to talk?' is asking if it is a good time to begin a conversation to offer a useful service that requires decision-making and several conversational turns."*

Contextual Data Collection: When the smart speaker prompted with the ESM message, "Is now a good time to talk?," we asked participants to first answer in "yes" or "no" and then to verbally describe their current activity. This was done to collect user activity contexts, which provided grounds for the reported interruptibility. To collect data in a natural context, we asked participants to respond to ESM only when they were at home and told them that they were not required to be at home intentionally to answer ESM questions. When both participants were at home, we asked them to answer consecutively. The reason for this instruction was that each participant may be present in different contexts, and their interruptibilities may differ in the given situation. Even if they were engaged in the same activity, their interruptibilities could differ due to personal reasons. Similar to prior studies on smart speakers [8, 48], we recorded voices for 1 minute after the ESM prompt, which was sufficient to fully capture user responses (as confirmed via our pilot study, see Section 3.2). The recorded audio data was uploaded in real time, and researchers then transcribed and examined the recorded ESM audio data.

The recorded ESM audio data was categorized into three types based on their completeness: (1) complete answer, (2) incomplete answer, and (3) unanswered. A complete answer meant that the user reported an interruptibility decision (yes/no) and detailed contextual information. An incomplete answer happened when the user failed to report an interruptibility decision or when their answer was insufficient to determine the context. Unanswered meant no answer at all. To check the context of incomplete and unanswered cases every day, researchers established group chat rooms with participants from each dormitory using a popular messenger application, KakaoTalk. For incomplete cases, researchers asked the participants to supplement their answers via the messenger. The researchers provided the time of the incomplete ESM answer and the ESM answers before and after the incomplete answer, if available, as cues for recall for the participant. Unanswered cases most likely occurred when nobody was in the room. The smart speaker was unable to discern if people were present in the room despite movement detection. For that reason, we asked the participants to report their entrance and departure times via instant messages. By comparing their entrance and departure times, we checked whether anybody was present in the room at the time of the unanswered case. If nobody was present, we did not examine the unanswered case. However, if somebody was in the room but no response was captured, we asked what they were doing at that time and why they did not answer, presenting cues for recall.

ESM Parameter Selection: ESM parameters, including operating hours and the minimum interval between consecutive prompts [64] were determined to collect sufficient amount of data. Our goal was to collect 10 ESM responses per day from each participant. To capture the interruptibility context after waking up and before sleeping, we asked participants to set their waking hours in accordance with their circadian rhythms. The rationale for this approach was that we posited that ESM prompts during sleep time could disturb participants' sleep, negatively influencing their daily activities. Similar to our approach, prior studies also limited ESM question prompts to waking hours [39, 40]. Since two participants shared the room and the smart speaker, we asked them to mutually agree upon the operating hours of ESM. We allowed them to change the operating hours during the data collection period, but most participants maintained their initial settings.

Despite more than twelve hours of ESM operating time, the actual time of participation was far shorter because of the participants' frequent outdoor schedules (e.g., attending classes, eating out, meetings). Our recruitment criteria included users who stayed at home for at least three hours a day, not necessarily in contiguous segments, to collect adequate number of natural activity contexts. To collect 10 ESM answers within three hours, time intervals between the prompts were determined to be at least twenty minutes to minimize disruption, which we explain in Section 3.2.

ESM Trigger Strategies: In our study, ESM prompts were delivered randomly or when movements were detected. Previous studies [21] reported that users were likely to be interruptible to mobile devices delivering messages during activity transition moments (e.g. from sitting to walking); these messages were more positively received during activity transitions than when delivered at random times. We wanted to investigate if the transition period between physical activities is an opportune time to interrupt users. The key difference from prior studies was that unlike mobile phones, smart speakers are immobile and fixed in one place.

We implemented the ESM application so that the prompts were triggered randomly or by movement detection at approximately the same time intervals. After a sampling prompt was delivered, the next sampling prompt was reserved for a random interval in 15–25 minutes to maintain an interval of 20 minutes on average. When a movement was detected, a sampling prompt was triggered immediately. However, to guarantee an interval of at least 15 minutes between sampling prompts, movement detection was disabled until 15 minutes had passed from the triggering of the previous prompt. If no movement was detected after 15 minutes from the last ESM, a random ESM was triggered at the reserved time; however, if a movement was detected before the reserved time, the reserved ESM was canceled, and the detected ESM was triggered.

(a) Picture of a room (b) Pixel-difference images (c) Overlapping two images

Fig. 3. Inferring movement context from pixel-difference images. Image (b) shows how pixel differences are saved when someone is moving from the bed to the desk, or vice versa.

3.1.4 Movement Detection. We implemented a simple movement detection method: comparing two consecutive photos taken at a 3-second interval. If there was a significant difference between the two consecutive photos, it was likely that a person had moved in the time interval between the capturing of the two photos. We conducted a validation test to confirm that a 3-second interval is sufficient for detecting movement, and the details of the test are discussed later in this section. For privacy reasons, we applied grayscale and Gaussian blurring filters. As a result, each pixel in the image had values between 0–255 in grayscale. Even in the absence of movement, pixel values could have changed in the time interval because of the fluctuation in ambient brightness. Thus, we considered a pixel to have changed only when the detected difference in a pixel across the two time points was above a threshold (denoted as "pixel-level difference threshold"). After comparing each pair of pixels, we counted the number of pixels where a significant difference was detected. If the aggregated difference across the photos was higher than a threshold (denoted as "photo-level difference threshold"), then it was considered a movement detected.

For ground truth verification, when a movement was detected, we saved the pixel differences (i.e., changed pixels in black and unchanged pixels in white), thereby expunging personally identifiable information. The movement context can be roughly inferred from such pixel-difference binary images (in black and white) by checking the captured shape differences. For example, black masses are captured near the bed and the desk in the pictures, and we can infer that someone is moving from the bed to the desk or vice versa (Figure 3).

We conducted a validation test to confirm that our movement detection approach was valid and to properly configure the pixel-level difference threshold and photo-level difference threshold. We aimed to detect large movements including standing up, sitting down, and walking around, as they signal activity transition and hence, higher interruptibility. Beside large movements, there could be slight movements, which included small or momentary movements such as moving only hands or arms, turning while lying in bed, leaning left or right while sitting on a chair, tilting head, and changing sitting postures. To conduct the test, we created an environment that simulated the furniture arrangement in the dormitory rooms (e.g., a pair of desks, beds, closets located at both sides) and recruited seven volunteers. We provided an activity script to the volunteers that included both slight and large movements. While the participants were moving as instructed in the script, the smart speaker application took pictures every 3 seconds. In total, we obtained 203 movement-labeled pictures from the participants.

We calculated false-positive rate (FPR) and true-positive rate (TPR) while varying pixel-level difference threshold (10%–30%) and photo-level difference threshold (7,500–42,500). The range of photo-level difference threshold was chosen because pixel differences in our dataset were mostly concentrated within this range. The smallest FPR and the largest TPR were achieved when the pixel-level difference threshold was 64 (25%) and photo-level difference threshold was 20,000. In practice, the arrangement of each room was slightly different. Based on our threshold results, when we deployed our ESM devices, we also tested basic movement scenarios in real dormitory environments and manually adjusted the thresholds as needed.

3.2 Pilot Tests

We conducted two pilot tests. The first pilot test focused on verifying the ESM data collection procedure to find any impediment that could affect data collection. It was conducted with four participants in two rooms (a pair of females, and a pair of males) for two days, 12 hours per day. This duration was sufficient to find potential obstacles in the ESM-based data collection. We asked the participants to actively report all inconveniences via text messages. All the participants agreed that the ESM inquiry frequency and operating hours were appropriately set. One participant said "*It was okay. It asked me when I felt like forgetting.*"

One group asked us to change the operating hours from 10:00–22:00 to 11:00–23:00, because they woke up later than expected. Adjusting operating hours depending on the participant's daily routine schedules seemed to have been helpful, and this feature was added into the ESM software. Negative experiences from proactive announcements (e.g., surprise and sharp voice) were also reported: one participant said, "*I felt very startled when it initiated to say something. I think my startling response was recorded.*" In our revision, we added a soft alarm sound ("ding-dong," for about 1-2 seconds) before the question and connected a Bluetooth speaker that could generate softer ESM sounds.

The second pilot test focused on testing the revised version of our ESM software. The second pilot was conducted for four days with the same participants. Participants liked the alarm sound and softer/louder voice from the speaker, as one participant commented, "*I prefer the alarm sound. It is definitely better because of good sound quality, and I do not feel startled.*" We also checked the number of ESM responses per day. Initially, we expected 10 ESM data per day for each participant, or 80 ESM responses per room in four days. However, while one pair answered 156 prompts, the other pair only answered 36 prompts, which was lower than our goal. After the interview, we found that this low level of participation was because the participants did not stay in their rooms frequently and for long periods of time. Based on this observation, we added a recruiting requirement: participants typically spending more than 3 hours a day in their room.

3.3 Recruitment and Data Collection

We recruited 40 participants from 20 rooms via online campus community and Facebook (12 male pairs, and 8 female pairs). As discussed in Section 3.1, we recruited students who lived in a room with one roommate, and both roommates were willing to participate in the data collection together. We requested users to stay in their rooms for at least 3 hours a day. ESM data was collected for a week (including a weekend).

The user study began with an orientation. The participants signed an informed consent form. We explained the purpose of the data collection, how the smart speaker works, what types of conversational tasks are assumed in ESM questions by giving the examples mentioned in Section 3.1, and how they should answer the ESM prompts.

For each room, we created a group chat room with each pair of participants, and asked them (1) to report entrance and departure times and (2) to answer researchers' questions about their ESM responses (to supplement incomplete or unanswered ESMs, if necessary). Then, the researchers visited each participant's room to install the smart speaker device, set operating hours as decided by the roommates, and manually adjust movement detection thresholds.

After completing the ESM data collection, we performed one-on-one exit interviews with all 40 participants. Each interview session lasted approximately 20 minutes. We mainly asked why they answered "yes" or "no" in specific situations after reviewing their responses. To understand the factors affecting activity-related interruptibility, we asked each participant the same questions about all ESM data.

3.4 Data Analysis

Two types of data were collected: ESM data and interview data. ESM data was mainly used for quantitative analysis, with columns for participant ID, ESM day and time, interruptibility, activity context, detected or randomly triggered, and labeled category. For categorization, we performed inductive coding with affinity diagramming [13]. We created a code book by categorizing similar home activities. Focusing on specific contexts (e.g., activity types and movements) helped us in creating an overview of interruptibility patterns. Exit interview data supplemented the quantitative responses as we could understand the detailed reasons of the participant interruptibility while doing a certain activity and uncover other factors affecting it.

4 ACTIVITY CATEGORIZATION

4.1 Data Preparation and Coding Process

A total of 3,572 responses were collected; the maximum number of responses collected from a single pair of participants (P33, P34) was 324 while the minimum was 84 (P30, P39). Responses were collected over a period of a week. ESM operational hours per day varied from 9 to 16 hours (M = 13.1, SD = 1.5). We excluded 28 responses, in which we could not infer the interruptibility or context from the response even after the post-collection interview; the participants could not remember the context or were unsure about what happened. In addition, we excluded 44 responses from a participant who misunderstood the goal of ESM. Finally, we analyzed the remaining 3,500 valid responses.

Preliminary categories for home activities to find contextual factors for interruptibility were initially created from a subset with 284 samples from 10 rooms, selected from respective days with the most amount of data collected for each room. The sampled subset was representative of the entire set and sufficient to create the preliminary categories of home activities.

One researcher manually examined responses and created an affinity diagram to group responses with similar themes together and develop an initial coding scheme for categorizing the home activities. Three researchers then coded the sample subset using the initial coding scheme and iteratively refined the coding scheme. The resulting scheme consisted of 19 unique categories (excluding miscellaneous category) and the final level of agreement (Krippendorff's alpha) among the three coders was 0.907.

After reaching an adequate level of agreement on the coding scheme, one researcher and three volunteers proceeded to label the entire set of responses. To become accustomed to the process and increase their reliability, the coders practiced with the sampled subset prior to coding the entire data. After excluding the 284 samples used for developing the preliminary coding scheme, the remaining 3288 samples were divided into two sets with 1642 and 1646 samples respectively. Each set was coded by two different pairs of coders. The respective levels of agreement for the first set and the second set were 0.925 and 0.892, both measured with Krippendorff's alpha. Table 1 shows the coding scheme in detail with definitions and sample responses for each category.

Table 1. Home activity categories with corresponding descriptions and sample responses

Category	Activity	Description	Example
Using media	Video gaming	Playing a video game on a computer or a mobile phone	• *"Just got back and started a game."* • *"I'm playing a game at the moment."*
	Internet / smartphone	Internet surfing with a mobile phone or a laptop	• *"Using Facebook sitting at the desk."*
	Watching videos	Watching videos on Youtube, Netflix, etc.	• *"Watching Youtube sitting on a chair."*
Working/studying	Working and studying	Working or studying	• *"Just sat down at the desk to study."* • *"I'm focusing on writing an email."*
Resting	Resting and relaxing	Resting for a while to relax and refresh or doing nothing at the moment	• *"Was sitting at the desk and just laid down to take a break."* • *"Not doing anything at the moment."* • *"I'm listening to music sitting down."*
About to leave	Visiting outside the dorm	About to leave to visit places outside the dorm (e.g., to take classes, to eat out, to meet friends)	• *"I'm packing my bag preparing to leave."* • *"I'm about to go to a class."*
	Visiting other rooms in the dorm	About to leave to visit nearby rooms in a dorm (e.g., going to the restroom, emptying trash)	• *"I'm about to leave to take a shower."*
Sleeping	Preparing to sleep	Preparing to sleep or about to fall into a sleep	• *"I'm about to take a nap."*
	Sleeping	Being asleep	• *"I'm sleeping on my bed."*
	After waking up	Just had woken up from sleep but still in a bed	• *"I just woke up."*
Self caring	Hair caring	Drying or combing hair	• *"I'm drying/combing my hair."*
	Changing clothes	Changing or wearing clothes	• *"I'm changing clothes."* • *"I'm changing before going out."*
	Face or body caring	Putting cosmetics on or cleaning a face or a body	• *"I'm wearing makeup before going out."* • *"I'm brushing my teeth."*
Social interaction	Face-to-face (F2F) interaction	Having face-to-face interaction with someone present in the same room	• *"I'm chatting with my roommate."*
	Online interaction	Having an online interaction with someone not present in the same room	• *"I was chatting with my friend on a messenger (KakaoTalk)."* • *"I'm talking to my dad (on the phone)."*
Eating	Eating	Eating a meal or a snack	• *"I'm having some snack while lying."*
Just returned	Returning from outside the dorm	Just came back home (e.g., after classes, having meals, or meeting friends)	• *"Just came in after having a dinner."*
	Returning from other rooms in the dorm	Just came back from nearby rooms in the dorm (e.g., restrooms, shower rooms)	• *"I just had a shower."*
Doing chores	Doing laundry, cleaning, or fixing	Doing chores related to laundry, cleaning, or fixing	• *"I'm doing/folding laundry."* • *"I'm cleaning my room/desk."*
Miscellaneous	Miscellaneous	Any other activity not mentioned above	• *"Just thinking about the dinner menu."* • *"I'm looking for something."*

Table 2. Frequency of response cases and interruptible cases ("Yes") across different activities. #case = number of cases (percentage in entire cases), #yes = number of interruptible cases (percentage within a given category/activity), Ent/Dep = Entrance/Departure, Transition = physical activity transition. Others = other dynamic activities.

Category-level			Activity-level												
Category type	#case	#yes	Activity type	All		Types of ESM prompt									
						Randomly prompted ESMs (RQ1)		Movement detected ESMs (RQ2)							
								All		Movement context					
										Ent/Dep		Transition		Others	
				#case	#yes	#case	#yes	#case	#yes	#case	#yes	#case	#yes	#case	#yes
Using media	1124 (32%)	776 (69%)	Video gaming	310 (9%)	113 (36%)	288 (10%)	97 (34%)	22 (3%)	16 (73%)			10 (7%)	10 (100%)	12 (4%)	6 (50%)
			Internet / Smartphone	381 (11%)	351 (92%)	359 (13%)	331 (92%)	22 (3%)	20 (91%)			4 (3%)	4 (100%)	18 (6%)	16 (89%)
			Watching videos	433 (12%)	312 (72%)	414 (15%)	300 (72%)	19 (3%)	12 (63%)			2 (1%)	2 (100%)	17 (6%)	10 (59%)
Working / Studying	824 (24%)	177 (21%)	Working and studying	824 (24%)	177 (21%)	743 (26%)	134 (18%)	81 (12%)	43 (53%)			41 (28%)	30 (73%)	40 (14%)	13 (33%)
Resting	433 (12%)	369 (85%)	Resting and relaxing	433 (12%)	369 (85%)	361 (13%)	300 (83%)	72 (10%)	69 (96%)			45 (31%)	44 (98%)	27 (10%)	25 (93%)
About to leave	336 (10%)	115 (34%)	Visiting outside	251 (7%)	93 (37%)	101 (4%)	38 (38%)	150 (22%)	55 (37%)	150 (57%)	55 (37%)				
			Visiting other rooms	85 (2%)	22 (26%)	27 (1%)	5 (19%)	58 (8%)	17 (29%)	58 (22%)	17 (29%)				
Sleeping	209 (6%)	65 (31%)	Preparing to sleep	45 (1%)	5 (11%)	36 (1%)	5 (14%)	9 (1%)	0 (0%)			8 (5%)	0 (0%)	1 (0%)	0 (0%)
			Sleeping	76 (2%)	3 (4%)	74 (3%)	3 (4%)	2 (0%)	0 (0%)					2 (1%)	0 (0%)
			After waking up	88 (3%)	57 (65%)	75 (3%)	48 (64%)	13 (2%)	9 (69%)			13 (9%)	9 (69%)		
Self caring	148 (4%)	94 (64%)	Hair caring	26 (1%)	16 (62%)	15 (1%)	7 (47%)	11 (2%)	9 (82%)			8 (5%)	7 (88%)	3 (1%)	2 (67%)
			Changing clothes	56 (2%)	36 (64%)	19 (1%)	8 (42%)	37 (5%)	28 (76%)					37 (13%)	28 (76%)
			Face or body Caring	66 (2%)	42 (64%)	50 (2%)	34 (68%)	16 (2%)	8 (50%)					16 (6%)	8 (50%)
Social interaction	129 (4%)	55 (43%)	F2F interaction	67 (2%)	27 (40%)	31 (1%)	8 (26%)	36 (5%)	19 (53%)			2 (1%)	1 (50%)	34 (12%)	18 (53%)
			Online interaction	62 (2%)	28 (45%)	52 (2%)	27 (52%)	10 (1%)	1 (10%)					10 (4%)	1 (10%)
Eating	101 (3%)	61 (60%)	Eating	101 (3%)	61 (60%)	74 (3%)	42 (57%)	27 (4%)	19 (70%)			10 (7%)	9 (90%)	17 (6%)	10 (59%)
Just returned	81 (2%)	77 (95%)	Returning from outside	53 (2%)	52 (98%)	14 (0%)	14 (100%)	39 (6%)	38 (97%)	39 (15%)	38 (97%)				
			Returning from other rooms	28 (1%)	25 (89%)	11 (0%)	9 (82%)	17 (2%)	16 (94%)	17 (6%)	16 (94%)				
Doing chores	66 (2%)	56 (85%)	Doing laundry or cleaning	66 (2%)	56 (85%)	26 (1%)	22 (85%)	40 (6%)	34 (85%)			3 (2%)	3 (100%)	37 (13%)	31 (84%)
Miscellaneous	49 (1%)	14 (29%)	Miscellaneous	49 (1%)	14 (29%)	36 (1%)	8 (22%)	13 (2%)	6 (46%)			1 (1%)	1 (100%)	12 (4%)	5 (42%)
Overall	3500 (100%)	1859 (53%)	Overall	3500 (100%)	1859 (53%)	2806 (100%)	1440 (51%)	694 (100%)	419 (60%)	264 (100%)	126 (48%)	147 (100%)	120 (82%)	283 (100%)	173 (61%)

4.2 Summary of Categorization Results

Table 2 shows the frequency of response cases and interruptible cases (i.e., user responded “Yes” to the question) across different activities at the time of interruption. For the activity transition contexts (‘Transition’ column; e.g., from sleeping to studying), we further categorized interruption contexts based on the next physical activity (e.g., studying) the participants were transitioning into.

As shown in ‘All’ column of Table 2, the most frequently reported home activity was working/studying (n = 824), whereas the least frequently reported activity was hair caring (n = 26). From ’Overall’ row − the last row of Table 2, it can also be seen that 53% (n = 1859) of the responses were “Yes”, indicating they were interruptible for conversation, while 47% (n = 1641) of the responses were “No.” Randomly prompted ESMs account for 80% (n = 2806) of the responses, while ESMs triggered by movement detection account for 20% (n = 694).

The majority of responses categorized as interruptible were from the following activities: returning from outside (98% reported as interruptible), Internet/smartphone use (92%), returning from other rooms (89%), resting (85%), doing chores (85%), watching videos (72%), waking up (65%), changing clothes (64%), face or body care (64%), hair care (62%), and eating (60%). On the other hand, the majority of responses for the following activities were uninterruptible: sleeping (4%), preparing to sleep (11%), working/studying (21%), visiting other rooms (26%), video gaming (36%), visiting outside the dorms (37%), f2f (face-to-face) interaction (40%), and online interaction (45%).

Home activities tended to periodically repeat across time of the day. Table 3 shows the percentage of interruptible cases for each day of the week and hour of the day. It is notable that participants were less interruptible in the mornings (9am–11am) than other times of the day, and their interruptibility did not vary across the day of the week. In the morning, the major activities were visiting outside the dorm (16%), hair care (16%), visiting other rooms (14%), face or body care (10%), sleeping (9%), and waking up (9%). These activities represented 72% of home

Table 3. Percentage of interruptible cases (“Yes”) across the day of week and hour of the day

Hour	**Day of Week**							**Overall**
	Sun	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	
9	0.00	20.02	25.01	0.00	50.00	40.02	0.00	19.29
10	26.67	41.67	39.13	52.94	40.00	35.00	0.00	33.63
11	67.57	77.78	64.71	50.00	52.94	47.83	85.71	63.79
12	64.10	67.86	59.57	67.50	70.27	30.77	53.85	59.13
13	57.50	45.45	28.57	71.43	33.33	56.67	56.52	49.92
14	63.33	69.57	52.38	45.00	41.38	57.14	58.82	55.37
15	71.79	60.00	50.00	58.62	50.00	66.67	37.84	56.42
16	54.29	56.36	71.88	47.37	65.85	46.15	48.84	55.82
17	48.28	50.88	50.94	58.00	42.86	60.87	50.00	51.69
18	53.19	62.00	54.17	59.09	62.86	61.90	46.15	57.05
19	60.47	58.21	57.14	63.83	48.28	42.11	62.07	56.02
20	62.22	51.56	38.00	35.00	57.14	38.46	63.41	49.40
21	50.00	49.15	53.85	51.92	37.50	57.58	54.35	50.62
22	51.02	60.00	49.02	49.02	36.11	30.00	52.83	46.86
23	54.24	51.11	53.66	58.97	48.72	50.00	38.24	50.71
Overall	52.31	54.77	49.87	51.25	49.15	48.08	47.24	50.38

activities during the morning. The rest of the activities included resting (7%), working/studying (6%), changing clothes (4%), Internet/smartphone use (4%), watching videos (2%), activities before sleeping (2%), returning from other rooms (1%), eating (1%), and online interaction (1%).

For the morning, the majority of responses categorized as uninterruptible were from the following activities: visiting outside the dorm (0% reported as interruptible), sleeping (0%), visiting other rooms (0%), activities before sleeping (0%), hair care (2%), visiting other rooms, online interaction (33%), and changing clothes (44%). Whereas the majority of responses for the following activities were interruptible: eating (100%), watching videos (75%), internet/smartphone (67%), and resting (61%). Note that among major activities, the highest proportion of interruptible cases was from after waking up (52%).

5 RESULTS

By analyzing the ESM responses and the exit interview data, we identified three key factors related to our research questions: RQ1) personal contextual factors (i.e., concentration and engagement, urgency and busyness, psychological/physical states, and auditory/verbal channel availability), RQ2) movement-related factors (i.e., entrance/departure and physical activity transition), and RQ3) the social presence factors (i.e., roommate's activities).

5.1 RQ1: Personal Contextual Factors

We conducted interviews with each participant after data collection and asked for reasons they responded “Yes” or “No” in a particular situation. Then we extracted four personal contextual factors affecting interruptibility: 1) concentration and engagement, 2) urgency and busyness, 3) psychological or physical states, and 4) auditory/verbal channel availability. In the following section, we discuss how these factors are related to the activities users were performing at the time of interruption. The influence of activity transition and social presence is discussed respectively in Sections 5.2 and 5.3.

5.1.1 Concentration and Engagement. Whether a person is concentrating on their current activity at the moment of interruption is crucial in determining if they are available or unavailable for an interaction. Depending on their characteristics, we further divided activities based on how much concentration they require. When participants were focusing on a particular activity that demanded a high level of attention, they frequently reported not being interruptible. This mostly happened while participants were working/studying, having face-to-face interactions, and playing video games, 82%, 74%, and 66% of which, respectively, were reported as uninterruptible. However, when the participants lost their focus or did not need to concentrate for a while, they reported being interruptible as they could temporarily switch to other tasks. P27 said, *“It depended on how much I was concentrating. I tended to not respond if I had made up my mind to focus on studying. But I said that I was available when I was taking a short break from work.”* Likewise, P31 commented, *“There are some games, in which you must complete a series of stages to finish. It is difficult to pause in the middle.”* Some activities, such as changing clothes, can be completed in a short time period and may even take less time than having a conversation with a smart speaker. When participants were already engaged in such activities, they generally preferred to finish them first before interacting with a smart speaker. For example, P14 said, *“I can finish it quickly [changing clothes], but if I were to be interrupted in the middle, I felt that it would take longer [to stop changing and respond to the speaker].”* In contrast, while performing activities requiring low concentration or engagement, the participants were generally interruptible. Such activities included 1) using the Internet/smartphone (92% reported as interruptible), 2) doing chores (85%), 3) resting (83%), and 4) watching videos (72%). For example, P22 said, *“You know, while using smartphones I do not need to focus too much. While browsing news feeds or checking KakaoTalk messages, I can talk.”* Another participant (P13) commented, *“If it was a really funny or important scene from YouTube, I said no, but if it was a not so important scene, I said yes.”*

5.1.2 Urgency and Busyness. The urgency of an activity and the busyness of a participant were highly related with the interruptibility. When participants were facing immediate deadlines (e.g., assignments due, class schedules), they tended to report being unavailable. For example, P25 said, *"I was in a hurry to finish my homework."* At times, the participants needed to leave the room quickly because of their upcoming schedules. P22 said, *"When I am late for my class, I need to prepare for leaving, and I don't have time to talk."* A similar pattern was observed when participants were busy with multiple activities. For example, P20 said, *"I was preparing to go out. I bathed and was drying my hair. I also needed to take care of my laundry. I needed to take care of multiple things at the same time, and it was a bit bothersome to chat with her [the speaker]. I did not have enough time. I also needed to fold up the bedding and brush my teeth."*

5.1.3 Psychological or Physical States. Participants' psychological and physical states were predominantly related to the quality of being awake/conscious and the activity of sleeping. When participants were preparing to sleep (or about to sleep), they were often disturbed by a randomly prompted ESM message from a speaker. P5 said, *"[before sleeping] If I talk, then I become awake."* In this situation, only 14% of the responses indicated being interruptible. While participants were sleeping, the sound of the ESM prompt often woke them up, which made them respond to the prompt while not being fully awake; in this case, participants reported that they were interruptible in a mere 4% of responses. Likewise, participants' wakefulness immediately after waking up affected their interruptibility. P12 said, *"[after waking up] I was not ready for a conversation, and I still felt really tired."* However, the participants reported being more interruptible (64%) after waking up than when preparing to sleep (14%) or while sleeping (4%). In addition, participants were unwilling to interact with the smart speakers when they were not in a mood for talking (n = 16), were tired (n = 22), or were sick (n = 8). P22 reported: *"[while studying] I was not in a good mood to talk. I felt a bit bothered to talk, and I was tired."* This was also true when users were resting: *"Because of a headache, I laid down to get some rest. I did not want to be disturbed."* [P17]. In contrast, some activities made participants feel good (e.g., eating something), which helped them engage with the speaker; *"I just came back from the store. I bought some food and I felt good while eating it. And I said yes."* [P21].

5.1.4 Auditory / Verbal Channel Availability. If the user's ability to hear or speak was obstructed, they were incapable of interacting with the smart speaker, hence their interruptibility was low. Some activities, such as drying one's hair with a hair dryer and watching videos, produced loud noises and interfered with hearing: *"Because of my dryer, I could not hear the prompt."* [P35] and *"I was watching a video with a headset."* [P21]. Certain activities like changing clothes or brushing teeth prevented from interacting with the speaker: *"It was awkward to talk while I was brushing my teeth."* [P37]. When participants were having face-to-face interactions (26% reported as interruptible) or online conversations (52%), they often said that it was hard to engage with the speaker. P22 said, *"I was on my phone. I was talking with someone, but she wanted to talk to me, and I could not say yes."*, and [P6] reported: *"I was talking with my friend, and it was difficult to talk to both simultaneously."* Nonetheless, users' availability depended on their level of engagement in the ongoing conversation: *"It was not an important conversation and it was trivial. (interruptible)."* [P14].

5.2 RQ2: Movement-related Factors

We further examined the relationship of interruptibility and movement-related factors. Among movement detected ESMs, 60% of responses were marked interruptible, whereas 51% of responses among randomly prompted ESMs were interruptible (see 'Overall' row of Table 2). Detected activities were further categorized into three movement sub-contexts: (1) entrance and departure, (2) physical activity transition, and (3) other dynamic activities. Here, other dynamic activities include changing clothes, doing chores, and small and brief movements during sedentary activities (such as stretching, tottering, and bending). Overall, we found a noticeable relationship between the participants' interruptibility and the movement sub-contexts. Participants were generally interruptible during entrance (96%) and activity transition (82%), but they were less interruptible during departures (35%). Unlike in these three sub-contexts, personal contextual factors (discussed in Section 5.1) were considered important for other dynamic activities.

5.2.1 Entrance and Departure. We found that interruptibility during entrance and departure was closely related to the communication range: being close enough to hear the sound or speak with the smart speaker increased the chance that participants were interruptible. Inbound movements toward the communication range of the speaker (entrance) were more interruptible than outbound movements (departure).

Entrance contexts included the user returning from outside the building and from other rooms. As shown in 'Ent/Dep' column (see '#case' sub-column) of Table 2, 97% and 94% of entrance contexts from outside of buildings and from other rooms, respectively, were interruptible. An inbound movement into the communication range enabled users to talk with a smart speaker because they just came back and were yet to engage in other activities. In the interview, P19 said, *"I was just entering my room and doing nothing special,"* and P25 said, *"It was okay because I just came back after taking a shower and would stay in my room."*

Departure contexts cover visiting other places outside the dormitory, and visiting other rooms in the dormitory, 37% and 29% of which, respectively, were reported to be interruptible. Departure activity is an outbound movement relative to the communication range of a smart speaker, which often makes it difficult for smart speaker to start an interaction. When leaving, some participants also mentioned that they were in a hurry. P25 said, *"It is a bad moment [to start a conversation] when I am in a hurry preparing to go outside, such as when putting on clothes or applying makeup. If I were to do the same preparation activities but had more time, I would be interruptible, though."* When leaving, the participants often exhibited routine behaviors that often occurs automatically (e.g., going to a restroom, or leaving the room after taking a shower basket). These routine behaviors were more frequently observed in the instances where participants were visiting other rooms in the dormitory. They responded that it is often disruptive to stop such routine behaviors in the middle. P25 said, *"It was not so urgent, but if I talked with the smart speaker, I would have had to come back and get out again, which I don't prefer."* P31 said, *"I was not interruptible because I decided to do laundry and (already) picked up my clothes."*

5.2.2 Physical Activity Transition. We defined physical activity transitions as instances in which users were transitioning to another (or next) physical activity, and included the following contexts: about to do something or moving to do something, which are represented by ESM responses: *"I am about to study,"* and *"I am going from a chair to a bed to rest"*. 'Transition' column in Table 2 shows the interruptibility ratios in relation to the next physical activities that users were transitioning into at the time of interruption. As shown in 'Overall' row – the last row of Table 2, interruptibility ratios of movement-based ESMs related to physical activity transitions (82%) were higher than those of randomly prompted ESMs (51%). Especially, for the working/studying activity, there was a notable difference in the interruptibility ratio; 73% of responses were yes (or interruptible) in activity transition contexts, whereas it was only 18% for randomly prompted ESMs. P23 said, *"[while studying] It was okay because at the moment of activity transition, my level of concentration was low. At the moments of activity transition, I turned around to check my phone or get up from my chair. I guess this was detected as a movement."* Other participants also responded similarly: *"When I moved from a bed to a chair to sit, I think, I could just talk while walking."* [P39], and *"This was the moment after I brought delivery food and was about to set it up on the table. I could do that while talking because it was before I started to eat."* [P34] These responses indicate that users are more likely to be interruptible when they are transitioning from one activity to the other compared to when they are interrupted in the middle of their activities.

5.2.3 Other Dynamic Activities. Movements were also detected while users were performing other types of dynamic activities including changing clothes and doing chores that required physical movements (e.g., hanging laundry). In addition to entrance/departure activities or activity transitions discussed in previous sections, short movements could be detected during activities. Our manual examination of pixel differences in images revealed that some sedentary activities involved brief movements such as bringing something, bending down to find something while studying, or moving around while conversing. In some cases, small user movements, such as moving back and forth sitting on a chair or changing sitting postures, were wrongly detected as activity transitions. ‘Others’ column (see ‘#yes’ sub-column) – the last column of Table 2 shows the interruptibility ratio across users’ activities when their movements were wrongly detected as activity transitions. In these cases, interruptibility was largely dependent on users’ personal contextual factors (or which activities the users were engaging in); multitasking was likely to be feasible when doing light chores, whereas it was less feasible when doing cognitively demanding or urgent tasks.

5.3 RQ3: Social Presence

As shown in Table 4, we categorized all ESM responses (3,500 responses) depending on social presence (i.e., whether a roommate was present at the moment of interruption). Among all the responses, 1,574 (45%) responses were collected when both roommates were present. These responses can be further categorized into two contexts: 1) doing activities together (e.g., talking together) and 2) separately doing their own activities. Overall, the influence of social presence on interruptibility was relatively small: Alone: 1,066/1,926 (55.35%) vs. Together: 793/1574 (50.38%).

Our analysis of the interview data and ESM responses revealed that when roommates were doing the same activity together, the interruptibility was largely dependent on their current level of engagement in that joint activity. Conversely, when roommates were separately doing their own activities, their interruptibilities were dependent on individual activity contexts. However, if a conversation with the smart speaker could interrupt the activity of their roommates, participants reported that they were not interruptible as they wanted to avoid disturbing their roommates and minimize possible interpersonal conflicts. Nonetheless, this pattern differed across individuals.

5.3.1 Doing Activities Together. Examples of roommates doing an activity together include talking to each other, playing a game together, studying and discussing some topic, eating meals together, and watching videos on the same device. In these contexts, 98 responses (49 pair cases) were collected and interruptibility was dependent on their level of engagement in the activity. For example, P37 said, *“I was having a heated debate with my roommate, so it was hard to interrupt our conversation.”* In contrast, P34 said *“I was watching YouTube together with my roommate while talking. Both, the conversation topic and YouTube content were not important, so I was able to answer.”*

Table 4. Number of responses across social presence contexts (i.e., whether a roommate was present)

Contexts			**Number of responses**		
			Overall	**Response type**	
				Yes	**No**
Staying alone			1926	1066	860
Staying together (RQ3)	Overall		1574	793	781
	Activity context	Doing activities together	98	43	55
		Separately doing own activities	1476	750	726
Overall			3500	1859	1641

In other instances, although roommates were engaged in an activity together, their responses for interruptibility sometimes did not match (14 responses; 7 pair cases). Those cases were mainly when they were performing different sub-tasks or were in different stages of an activity. For example, when playing a computer game together, one person lost a life in the game and was waiting for the other who is still alive to finish playing. In this case, the person who was waiting was interruptible, while the person who was still playing was not. Participants’ responses also differed in a situation when one of roommates was performing additional activities at the same time (i.e., concurrent multitasking). For example, when two roommates were having a conversation, one roommate decided to go to a restroom, which was inside the room. In this case, the person who was going to a restroom responded no when a smart speaker asked if he/she was available for a conversation, while the other person responded yes.

5.3.2 Separately Doing Their Own Activities. 1,476 responses (738 pair cases) were collected when roommates were separately doing their own activities. When performing separate activities, participants took their roommate’s activity into the account (e.g., sleeping). When the sound of the voice interaction could interrupt their roommates, participants said that they were not interruptible. P3 said, *“I think it’s better not to talk when my conversation can distract my roommate’s study.”* Also, there were 19 ESM responses that explicitly mentioned that their interruptibility was influenced because their roommate was asleep. P39 said, *“I am using my cell phone in front of the fridge, but my roommate is sleeping. Since I could wake her up, it is not a good time [to talk].”*

During the exit interviews, we further investigated to what extent the other party’s contexts influenced the interruptibility participants with the following question: *“Did you change your answer, even though you were interruptible, because your roommate was (asleep/studying/doing something)?”* From the interview, we found that 29 participants (72.5%) preferred not to talk when their roommates were asleep, whereas it did not matter to 11 participants (27.5%) as they claimed that their roommates are not so sensitive. For example, P18 said *“I know that he would sleep very well, even if I talked with the smart speaker.”* Roommate’s studying was less influential on the interruptibility of the participants than sleeping. There were 11 participants (27.5%) who preferred not to talk when their roommates were concentrating on studying which is a far smaller proportion than that of sleeping. P25 said *“Usually, we freely initiate a conversation while the other is studying. So, I was interruptible when she was studying. But when sleeping, she should not be awakened [so I wasn’t interruptible].”*

6 DISCUSSION

We review our findings in relation to prior studies, present design insights for designing context-aware proactive conversation management system for smart speakers, and discuss limitations of our work.

6.1 Summary of Major Findings

Our study empirically examined three contextual factors relevant to the interruptibility of smart speaker-based proactive services in domestic contexts: personal, movement-related, and social factors. These factors show different characteristics compared to the factors affecting the interruptibility of services provided in conventional computing contexts, such as computers and mobile devices in offices and vehicles.

The personal contextual factors include concentration and engagement, urgency and busyness, psychological/physical states (e.g., wakefulness, mood, and fatigue), and auditory/verbal channel availability. Our results are consistent with prior studies on interruptibility; users engaged in challenging tasks are less interruptible [46] than those participating in rote tasks [37]. Participants were interruptible in 18% of the cases while working or studying versus 85% while doing chores. Users’ current states such as busyness and mood may negatively affect interruptibility [70]; our interview data revealed that participants tended to be uninterruptible when they were hurrying to leave or busy to finish an urgent task and also when they were not in a good mood. As suggested in multiple resource theory [67], our participants could perform two concurrent tasks (or dual-tasking [53]) by utilizing different types of mental/motor resources as long as the overall resource requirement did not exceed their upper limit of mental resources. We also observed that behaviors like hair drying and talking with other people preoccupied the users' auditory and verbal channels and limited their availability for a speech-based interaction.

Movement-related factors were also closely related to interruptibility. Users were generally more interruptible after entering a room and were transitioning between physical activities, but were less interruptible during departure; participants were interruptible in 96% of cases upon returning to their rooms, but only 35% of departures were reported as interruptible. High interruptibility during activity transition is consistent with previous research [21, 46]. During activity transitions, as the users were not occupied with any activities, they could easily shift their attention to the smart speakers [67]. Furthermore, different levels of availability for interaction during entering and departing suggest that the communication range and a user's movement with respect to that range are critical to user interruptibility to smart speaker services. If users are moving away from the device and out of the communication range (or have already engaged in short routines such as visiting a bathroom), it becomes difficult to initiate new conversational interactions. On the other hand, if users are entering the communication range, that signals an opportune moment for a smart speaker to start an interaction.

Finally, the presence of other people also influences interruptibility. In prior interruptibility studies [38], this factor was not generally considered important because these studies mostly considered (physically) solitary scenarios (e.g., exchanging messages with remote users using mobile phones). However, smart speakers are typically located in shared spaces, such as living rooms, where social interactions frequently occur [66], and voice interactions in such spaces may disturb others. Therefore, the interruptibility of smart speaker users may be contingent upon the interruptibility of co-located users. If the co-located users are occupied by the same activity, the interruptibility is mainly dependent on both of their personal contextual factors, such as engagement and urgency. However, when co-located users are engaged in different activities, one user may consider the other user's status or activity to minimize the chance of discomforting them. Our interview data indicated that the decision process was also influenced by the cohabitant's traits such as sensitivity to ambient environments. As shown in an earlier study [11], shared spaces introduce new issues of timing, relevance, and privacy into smart home environments.

6.2 Privacy Concerns in Vision-based ESM Triggering

We used the following approach to capture contextual information related to interruptions: (1) saved pixel differences between two consecutive photos (see Section 3.1.4) at a 3-second interval when a significant movement was detected, and (2) recorded surrounding sounds including voices for 1 minute after an ESM prompt. As it was necessary for our study to collect and process privacy sensitive information in real time, we asked the participants during exit interviews about any privacy concerns.

All of our participants were informed about our approach, which was explained at the orientation before the data collection started. Surprisingly, the majority did not express any concerns. One of the main reasons was that they estimated the potential privacy risks as marginal. For example, P32 commented: *"I knew that I was being recorded from the beginning. In fact, I have nothing in the [voice] recordings that shouldn't be heard by others, so I didn't feel so uncomfortable saying anything. [...] The pictures were all blacked-out and had no distinguishable figures as they were only used to see the difference, so it didn't matter."* Participants also justified their unconcerned attitude with their belief that researchers would not misuse the collected information. For example, P32 said *"The experimenters explained (about how the data will be collected and used), so I went on trusting them and didn't worry much."* Some participants even directly mentioned IRB regulations and said that they were not concerned as the IRB approval assured against any privacy violation, similar to P39: *"Because you mentioned the school's policies on the treatment of the collected data, I didn't worry."*

The participants' overall sentiments on the issue of privacy threats were moderate. This result is possibly due to the fact that the data was collected in the context of academic research where potential damage from privacy violations is likely not too detrimental, in comparison to more severe instances like data leakage by companies. The participants' behaviors and views are consistent with findings from a recent study that reported that users' lack of privacy concerns about data collection of smart home devices (e.g., Nest cam indoor, Amazon Echo) can be attributed to users' trust in data collectors and their tendency to underestimate privacy risks [32, 58].

Despite the fact that the majority did not express any concerns, some participants reported concerns about being photographed. P24 commented *"Maybe I had a bit more [concern] in the beginning, but I became used to it the more I went on [with the experiment]."* Furthermore, some privacy concerning moments raised a user's awareness of being monitored as P30 said, *"... when I was changing clothes, I remembered that I was being photographed. I knew that pictures wouldn't have so much detail, but I still was aware, so maybe I was a bit concerned."* (Interviewer: *"Did you feel the same about audio recordings?"*) *"No, I don't think so."*

Interestingly, the exit interview results revealed that participants may accept, even willingly, to compromise their privacy in exchange for a smart speaker that offers personalized care. P7 remarked: *"I thought I wouldn't be so lonely but be in a good vibe if she could ask me about more private stuff, like, feel cared for."* (Interviewer: *"Sometimes people are creeped out when an AI speaker asks you about private stuff, but you think that is okay?"*) *"Wouldn't you feel more attached if it could ask you [private stuff]? Not just saying the same thing to everyone, but better if it could say something special just for me, only if you could get rid of those [privacy] problems."* Nevertheless, P7 further noted that she would stop responding if she would get *"the slightest feeling that the speaker is following a routine"* because she would think: *"It is not real."*

While offering personalized care indicates one way to mitigate user privacy concerns about being photographed, the concerns can be further mitigated by utilizing different technology, such as an internal microphone sensor and/or existing IoT (Internet of Things) home devices and sensors, instead of a camera. We further discuss these technical mitigation strategies in Section 6.5.

6.3 Towards Context-aware Proactive Conversation Management

Proactive services with smart speakers could bring benefits (e.g., providing useful information, inspiring users) and challenges (e.g., timing, privacy/surveillance concerns) [11]. In social contexts, they may result in interpersonal conflicts by disrupting the work of others, but also provide interpersonal facilitation by promoting playful interaction and more efficient household task handling [6]. Our findings clearly demonstrate that it is important to carefully consider everyday activities in home contexts when determining the timing of proactive service delivery. This concurs with the claim of Tolmie et al. [61] that ubiquitous computing should seamlessly augment and support everyday routines at home, which starts with awareness of those routines.

Another important aspect is a user's mobility in home contexts, which creates a unique issue for (wireless) smart speaker interactions. Rodden and Benford [51] recommended that to integrate ubiquitous computing technologies into a smart building, designers should consider multiple dimensions such as sites and space planning. For example, when home wireless networking was reviewed from this perspective, prior work showed that users' perception on service boundaries (related to sites) are malleable, and wireless networking requires careful space planning with proper access control [12]. Thus, it is critical to carefully consider not only spatial relationships (proxemics) [7] but also domestic routines [61] for technology integration and development [12, 51].

Given this goal, we envision designing a *proactive conversation management* feature for proactive smart speakers, which would determine when to start, pause, or resume conversation by analyzing user contexts. Supporting proxemic interactions for smart speakers assumes that devices can acquire fine-grained knowledge about nearby users and their devices, including identity and position/movement/orientation [7]. This context information can be used to control/trigger user interactions as well as coordinate multi-device and -user interactions (e.g., passing messages across multiple devices) [34]. Since user mobility is an important factor for interruptibility, the context of user mobility, such as their movement and orientation, can be used to infer whether a user is currently departing or arriving. For example, if a user is leaving out of the communication range, then the system automatically pauses the current interaction and asks to resume the conversation later when the user returns. To achieve fine-grained control of triggering conditions, we can leverage the interpersonal distance (e.g., personal vs. social vs. public spaces as proposed by Hall [19]) by discretizing the space around the speaker.

Existing proxemic interaction models can be further extended by considering everyday routines. As shown earlier, some activities like hair drying or phone conversations make it difficult for users to dual-task. Smart speakers can make informed decisions about whether to initiate proactive interactions or to coordinate ongoing interactions, by having a deeper understanding of users' domestic routines and their personal contextual factors. There are various ways of supporting context-awareness in home environments. For example, existing tools based on acoustic fingerprinting can be leveraged to recognize routine behaviors [31]. Computer vision or wearable sensing techniques can be used to recognize current activities as well users' contextual information including engagement, mood, and stress [22, 47].

As in prior ubiquitous computing studies [26, 46, 47], an interesting future research direction is to consider multimodal sensory data to automatically identify opportune moments for proactive conversational interactions, which is an important part of context-aware proactive conversation management. Moreover, it is possible to routinize proactive conversational interactions based on simplified user feedback or end-user programming methods [16, 69], or to help users handle the deviations and exceptions from routines via human-in-the-loop system design [41]. In some cases, it would be quite important to offer further control to users by allowing them to explicitly set "interaction restraints" [28, 45] on proactive conversational interactions (e.g., silent mode during important conversations). Conversational management can leverage prior studies on identifying conversation breakdowns and devising repair strategies in diverse domestic settings [8, 48], or analyzing conversational log data to understand interaction patterns for context adaptation and personalization [9, 54].

6.4 Emotional Attachment and Interruptibility

Our exit interview indicates another compelling design implication for smart speakers: emotions invoked by the device may impact the interruptibility to proactive services. Several participants referred to the speakers as their *"friends"* during the exit interview. Notably, P19, who named the speaker *"Jarvis,"* said it will be *"too bad to stop the experiment"* because they have been together like a family for one week. Similarly P13 said *"it was good that the speaker speaks to you first because people these days are lonely, so I think it is psychologically comforting to have something that talks to you, like a pet. I think my room felt a bit empty once the speaker was gone."* Altogether, people felt an emotional attachment towards the speakers, despite the short duration of the experiment and the limited interaction capacity of the speaker. That attachment may cause people to feel uneasy about rejecting smart speakers. Conversely, a comment from P9 revealed that she did not feel obliged to respond to the speaker as it is *"something like a robot,"* which *"throws questions without emotions"*. This suggests that there may be a relationship between an emotion the user feels towards the speaker and their interruptibility.

A design implication that encapsulates the discussion so far would be that human likeness of the interacting device is a critical determinant of user interruptibility. This is related to personification of smart speakers. Prior studies [36, 49] hinted that device personification is associated with sociability, politeness, and customer loyalty. There is a lower chance that a message will be neglected if it feels natural, and a real person or emotionally attached bot is behind the message. While the goal of this work is to identify and understand opportune moments for smart speakers to initiate proactive services, the interruptibility research at its essence aims to create a system that interacts with humans as humans do, thereby promoting emotional interactions. In that regard, the ability of a device to seamlessly interrupt a user at a given moment as a social actor (e.g., visual appearance, social/emotional attachment) needs further studies in the future.

6.5 Utilizing Opportunities in Domestic Routine Contexts

Our results showed three notable activity contexts in which a smart speaker can proactively interact with their users: (1) just returned (95% reported as interruptible), (2) Internet/smartphone use (92%), and (3) doing chores (85%). Upon their return, the participants were willing to interact with the smart speaker as they were unlikely to be occupied with any task. Smart speakers could utilize a set of internal sensors to detect particular domestic routine contexts, such as returning home, and recognize opportune moments for proactive interactions. Smart speakers generally support Bluetooth or/and Wi-Fi connections, which, in combination with the users' smartphones, could be used to detect their home return. Furthermore, multiple users may exist in one home environment. Depending on which user's smartphone signal appears, the speaker could provide personalized services to that particular user. Similarly, the speaker could monitor traffic usage of the home Wi-Fi network to detect home return and Internet/smartphone usage.

In our study, we used a camera to detect movement contexts (see Section 3.1.4), which caused privacy concerns and discomfort for some participants. Instead of a camera, smart speakers could utilize an internal microphone sensor and analyze surrounding sounds, for example, human voice and ambient noise—doorstep and door open sound. Similarly, Sensay [56], a context-aware mobile phone, analyzed voice sound (speaking or not) and ambient noises (low, medium, or high noise levels) to infer the current context or environment (e.g., having a meeting, sitting alone at a computer). In addition to an internal microphone sensor, smart speakers could also utilize external devices and their embedded sensors. For example, activity of IoT home devices (e.g., turning a television on or off, opening or closing a refrigerator door, turning a light on or off) or smartphones carried at home (e.g., indoor localization) can be used to sense the user's movement in the home environment. Moreover, entrance detecting sensors (e.g., smart door locks, motion sensors) can be used to infer interruptibility related to domestic routines. Instead of a normal camera, for privacy reasons, a thermal camera could be used to detect the appearance of users in a doorway. People were generally interruptible while performing activities requiring low concentration or engagement, such as Internet/smartphone and doing chores [43]. Thermal images from a thermal camera could contribute to the detection of cognitive activities (e.g., cognitive heats [1]).

6.6 Limitations

We conducted data collection in college dormitories with two roommates per room. Despite the limited setting, dormitory environments could provide useful insights that can extend to a family home environment. However, dormitory environments preclude several in-home activities such as cooking and restrict the range of inhabitants' movements due to their small size. Furthermore, family members may have more shared context for a particular notification as they likely share a more intimate bond than roommates. Nevertheless, dormitories still bear similarity to the family home to some extent as roommates socialize with each other while sharing many activities necessary for daily living, which allowed us to collect a rich dataset conveniently. The living room, where social interactions between family members occur that may affect their interruptibilities (e.g., watching TV together, having a conversation), is the most common location for smart speakers (45%) [66], and these types of social interactions were observed in the dormitory rooms as well. Further studies with people living in more diverse locations and social settings, such as a person living alone in a studio, family living together in a residence, and friends sharing an apartment, would further generalize our findings.

Another limitation is that we did not consider real conversational tasks to judge the interaction timing, but we asked users to envision hypothetical productivity scenarios (e.g., restocking products or changing schedules). Prior research frequently used similar approaches to judge users' interruptibility [26, 55] while minimizing personal bias resulting from the content of the conversation. Alternatively, other behavioral markers can be used as proxy measures to indirectly infer opportune moments, e.g., when a user naturally engages in secondary tasks in driving contexts [29], or when a user accepts or misses a phone call [60]. As alluded to in recent work on multi-stage receptivity model [14], actual engagement with proactive services (availability vs. adherence) requires further investigation. This could be achieved by conducting user studies using realistic and contextualized proactive services: information delivery, entertainment, decision-making, and IoT home device control tasks.

7 CONCLUSION

A week-long field study with forty participants was conducted to collect 3,572 in-situ experience reports about opportune moments for proactive conversational interactions with smart speakers. We identified several unique factors related to everyday routines at home, such as habituated activities, resource conflicts for dual-tasking, user mobility types (e.g., entrance/departure), and interpersonal conflicts. We concluded that proactive smart speakers should carefully consider domestic routines and spatial relationships to support context-aware proactive conversation management. We believe that our findings made the first step towards exploring novel proactive conversational interactions in home environments, and we call for further studies on conducting follow-up user experience research and developing technical solutions for controlling proactive interactions.

ACKNOWLEDGMENTS

This research was partly supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2017M3C4A7083534), and by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the the Korea government (MSIT) (2020R1A4A10187747).

REFERENCES

[1] Yomna Abdelrahman, Eduardo Velloso, Tilman Dingler, Albrecht Schmidt, and Frank Vetere. 2017. Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 33 (sep 2017), 20 pages. https://doi.org/10.1145/3130898

[2] Piotr D. Adamczyk and Brian P. Bailey. 2004. If Not Now, when?: The Effects of Interruption at Different Moments Within Task Execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria) *(CHI '04)*. ACM, New York, NY, USA, 271–278. https://doi.org/10.1145/985692.985727

[3] Amazon. 2020. Alexa Skills. https://www.amazon.com/alexa-skills/b/?node=13727921011&ref_=topnav_storetab_a2s

[4] Amazon. 2020. Alexa Voice Shopping. https://www.amazon.com/alexa-voice-shopping/b?node=14552177011

[5] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 3, 13 pages. https://doi.org/10.1145/3290605.3300233

[6] Noah Apthorpe, Arunesh Mathur, Pardis Emami-Naeini, Marshini Chetty, and Nick Feamster. 2019. You, Me, and IoT: How Internet-Connected Home Devices Affect Interpersonal Relationships. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19)*. ACM, New York, NY, USA, 142–145. https://doi.org/10.1145/3311957.3359489

[7] Till Ballendat, Nicolai Marquardt, and Saul Greenberg. 2010. Proxemic Interaction: Designing for a Proximity and Orientation-Aware Environment. In *ACM International Conference on Interactive Tabletops and Surfaces* (Saarbrücken, Germany) *(ITS '10)*. Association for Computing Machinery, New York, NY, USA, 121–130. https://doi.org/10.1145/1936652.1936676

[8] Erin Beneteau, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication Breakdowns Between Families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 243, 13 pages. https://doi.org/10.1145/3290605.3300473

[9] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 91 (Sept. 2018), 24 pages. https://doi.org/10.1145/3264901

[10] Matthias Braunhofer, Francesco Ricci, Béatrice Lamche, and Wolfgang Wörndl. 2015. A Context-Aware Model for Proactive Recommender Systems in the Tourism Domain. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Copenhagen, Denmark) *(MobileHCI '15)*. ACM, New York, NY, USA, 1070–1075. https://doi.org/10.1145/2786567.2794332

[11] Ana Paula Chaves and Marco Aurélio Gerosa. 2019. How should my chatbot interact? A survey on human-chatbot interaction design. *CoRR* abs/1904.02743 (2019). arXiv:1904.02743

[12] Marshini Chetty, Ja-Young Sung, and Rebecca E. Grinter. 2007. How Smart Homes Learn: The Evolution of the Networked Home and Household. In *Proceedings of the 9th International Conference on Ubiquitous Computing* (Innsbruck, Austria) *(UbiComp '07)*. Springer-Verlag, Berlin, Heidelberg, 127–144.

[13] Ji Young Cho and Eun-Hee Lee. 2014. Reducing Confusion about Grounded Theory and Qualitative Content Analysis: Similarities and Differences. *The Qualitative Report* 19, 32 (2014), 1–20.

[14] Woohyeok Choi, Sangkeun Park, Duyeon Kim, Youn-kyung Lim, and Uichin Lee. 2019. Multi-Stage Receptivity Model for Mobile Just-In-Time Health Intervention. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 2, Article 39 (June 2019), 26 pages. https://doi.org/10.1145/3328910

[15] Andy Crabtree and Tom Rodden. 2004. Domestic routines and design for the home. *Computer Supported Cooperative Work (CSCW)* 13, 2 (2004), 191–220. https://doi.org/10.1023/B:COSU.0000045712.26840.a4

[16] Scott Davidoff, Min Kyung Lee, Charles Yiu, John Zimmerman, and Anind K. Dey. 2006. Principles of Smart Home Control. In *Proceedings of the 8th International Conference on Ubiquitous Computing* (Orange County, CA) *(UbiComp'06)*. Springer-Verlag, Berlin, Heidelberg, 19–34.

[17] James Fogarty, Andrew J. Ko, Htet Htet Aung, Elspeth Golden, Karen P. Tang, and Scott E. Hudson. 2005. Examining Task Engagement in Sensor-based Statistical Models of Human Interruptibility. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 331–340. https://doi.org/10.1145/1054972.1055018

[18] Google. 2020. Google Assistant: Google Developers. https://developers.google.com/assistant

[19] Edward Twitchell Hall. 1966. *The hidden dimension.* Vol. 609. Garden City, NY: Doubleday.

[20] Javier Hernandez, Daniel McDuff, Christian Infante, Pattie Maes, Karen Quigley, and Rosalind Picard. 2016. Wearable ESM: Differences in the Experience Sampling Method Across Wearable Devices. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. ACM, New York, NY, USA, 195–205. https://doi.org/10.1145/2935334.2935340

[21] Joyce Ho and Stephen S. Intille. 2005. Using Context-aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Portland, Oregon, USA) *(CHI '05)*. ACM, New York, NY, USA, 909–918. https://doi.org/10.1145/1054972.1055100

[22] Karen Hovsepian, Mustafa al'Absi, Emre Ertin, Thomas Kamarck, Motohiro Nakajima, and Santosh Kumar. 2015. cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 493–504. https://doi.org/10.1145/2750858.2807526

[23] Shamsi T. Iqbal and Brian P. Bailey. 2005. Investigating the Effectiveness of Mental Workload As a Predictor of Opportune Moments for Interruption. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. ACM, New York, NY, USA, 1489–1492. https://doi.org/10.1145/1056808.1056948

[24] Shamsi T. Iqbal and Brian P. Bailey. 2008. Effects of Intelligent Notification Management on Users and Their Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) *(CHI '08)*. ACM, New York, NY, USA, 93–102. https://doi.org/10.1145/1357054.1357070

[25] Tomoko Kamimura, Rina Ishiwata, and Takenobu Inoue. 2012. Medication Reminder Device for the Elderly Patients With Mild Cognitive Impairment. *American Journal of Alzheimer's Disease & Other Dementiasr* 27, 4 (2012), 238–242. https://doi.org/10.1177/1533317512450066

[26] Auk Kim, Woohyeok Choi, Jungmi Park, Kyeyoon Kim, and Uichin Lee. 2018. Interrupting Drivers for Interactions: Predicting Opportune Moments for In-vehicle Proactive Auditory-verbal Tasks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 175 (Dec. 2018), 28 pages. https://doi.org/10.1145/3287053

[27] Auk Kim, Jung-Mi Park, and Uichin Lee. 2020. Interruptibility for In-Vehicle Multitasking: Influence of Voice Task Demands and Adaptive Behaviors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 1, Article 14 (March 2020), 22 pages. https://doi.org/10.1145/3381009

[28] Jaejeung Kim, Joonyoung Park, Hyunsoo Lee, Minsam Ko, and Uichin Lee. 2019. LocknType: Lockout Task Intervention for Discouraging Smartphone App Use. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300927

[29] SeungJun Kim, Jaemin Chun, and Anind K. Dey. 2015. Sensors Know When to Interrupt You in the Car: Detecting Driver Interruptibility Through Monitoring of Peripheral Interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 487–496. https://doi.org/10.1145/2702123.2702409

[30] 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)*.

ACM, New York, NY, USA, 881–894. https://doi.org/10.1145/3196709.3196784

[31] Gierad Laput, Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-Play Acoustic Activity Recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. ACM, New York, NY, USA, 213–224. https://doi.org/10.1145/3242587.3242609

[32] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (Nov. 2018), 31 pages. https://doi.org/10.1145/3274371

[33] Bokyung Lee, Sindy Wu, Maria Jose Reyes, and Daniel Saakes. 2019. The Effects of Interruption Timings on Autonomous Height-Adjustable Desks That Respond to Task Changes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. ACM, New York, NY, USA, Article 328, 10 pages. https://doi.org/10.1145/3290605.3300558

[34] Jemin Lee, Uichin Lee, and Hyungshin Kim. 2019. PASS: Reducing Redundant Interactions between a Smartphone and a Smartwatch for Energy Saving. *IEEE Transactions on Mobile Computing* (2019), 1–1.

[35] Uichin Lee, Kyungsik Han, Hyunsung Cho, Kyong-Mee Chung, Hwajung Hong, Sung-Ju Lee, Youngtae Noh, Sooyoung Park, and John M Carroll. 2019. Intelligent positive computing with mobile, wearable, and IoT devices: Literature review and research directions. *Ad Hoc Networks* 83 (2019), 8–24.

[36] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a Mindless Companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (New Brunswick, NJ, USA) *(CHIIR '18)*. Association for Computing Machinery, New York, NY, USA, 265–268. https://doi.org/10.1145/3176349.3176868

[37] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. 2015. Focused, Aroused, but So Distractible: Temporal Perspectives on Multitasking and Communications. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. ACM, New York, NY, USA, 903–916. https://doi.org/10.1145/2675133.2675221

[38] 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)*. ACM, New York, NY, USA, 813–824. https://doi.org/10.1145/2750858.2807544

[39] 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

[40] 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

[41] Sarah Mennicken, Jo Vermeulen, and Elaine M. Huang. 2014. From Today's Augmented Houses to Tomorrow's Smart Homes: New Directions for Home Automation Research. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 105–115. https://doi.org/10.1145/2632048.2636076

[42] Alex Mihailidis and Geoffrey R Fernie. 2002. Context-aware assistive devices for older adults with dementia. *Gerontechnology* 2, 2 (2002), 173–188.

[43] 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)*. ACM, New York, NY, USA, 541–554. https://doi.org/10.1145/2998181.2998189

[44] Daehee Park, Wan C. Yoon, and Uichin Lee. 2020. Cognitive States Matter: Design Guidelines for Driving Situation Awareness in Smart Vehicles. *Sensors* 10 (2020), 2978.

[45] Joonyoung Park, Jin Yong Sim, Jaejeung Kim, Mun Yong Yi, and Uichin Lee. 2018. Interaction Restraint: Enforcing Adaptive Cognitive Tasks to Restrain Problematic User Interaction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3170427.3188613

[46] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. ACM, New York, NY, USA, 897–908. https://doi.org/10.1145/2632048.2632062

[47] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 91 (Sept. 2017), 25 pages. https://doi.org/10.1145/3130956

[48] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 640, 12 pages. https://doi.org/10.1145/3173574.3174214

[49] Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. “Alexa is My New BFF”: Social Roles, User Satisfaction, and Personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA ’17)*. Association for Computing Machinery, New York, NY, USA, 2853–2859. https://doi.org/10.1145/3027063.3053246

[50] Jingyan Qin, Zhibo Chen, Wenhao Zhang, Daisong Guan, Zhun Wu, and Min Zhao. 2019. Research on Active Interaction Design for Smart Speakers Agent of Home Service Robot. In *International Conference on Human-Computer Interaction*. Springer, 253–263.

[51] Tom Rodden and Steve Benford. 2003. The Evolution of Buildings and Implications for the Design of Ubiquitous Domestic Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI ’03)*. ACM, New York, NY, USA, 9–16. https://doi.org/10.1145/642611.642615

[52] Ben 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/

[53] Dario D. Salvucci, Niels A. Taatgen, and Jelmer P. Borst. 2009. Toward a Unified Theory of the Multitasking Continuum: From Concurrent Performance to Task Switching, Interruption, and Resumption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI ’09)*. ACM, New York, NY, USA, 1819–1828. https://doi.org/10.1145/1518701.1518981

[54] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "Hey Alexa, What’s Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) *(DIS ’18)*. ACM, New York, NY, USA, 857–868. https://doi.org/10.1145/3196709.3196772

[55] Rob Semmens, Nikolas Martelaro, Pushyami Kaveti, Simon Stent, and Wendy Ju. 2019. Is Now A Good Time?: An Empirical Study of Vehicle-Driver Communication Timing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI ’19)*. ACM, New York, NY, USA, Article 637, 12 pages. https://doi.org/10.1145/3290605.3300867

[56] Daniel P Siewiorek, Asim Smailagic, Junichi Furukawa, Andreas Krause, Neema Moraveji, Kathryn Reiger, Jeremy Shaffer, and Fei Lung Wong. 2003. SenSay: A Context-Aware Mobile Phone. In *Seventh IEEE International Symposium on Wearable Computers*, Vol. 3. 248–249. https://doi.org/10.1109/ISWC.2003.1241422

[57] Timothy Sohn, Kevin A. Li, Gunny Lee, Ian Smith, James Scott, and William G. Griswold. 2005. Place-Its: A Study of Location-based Reminders on Mobile Phones. In *Proceedings of the 7th International Conference on Ubiquitous Computing* (Tokyo, Japan) *(UbiComp’05)*. Springer-Verlag, Berlin, Heidelberg, 232–250. https://doi.org/10.1007/11551201_14

[58] Madiha Tabassum, Tomasz Kosiundefinedski, and Heather Richter Lipford. 2019. “I Don’t Own the Data”: End User Perceptions of Smart Home Device Data Practices and Risks. In *Proceedings of the Fifteenth USENIX Conference on Usable Privacy and Security* (Santa Clara, CA, USA) *(SOUPS’19)*. USENIX Association, USA, 435–450.

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

[60] G. H. (Henri) ter Hofte. 2007. Xensible Interruptions from Your Mobile Phone. In *Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services* (Singapore) *(MobileHCI ’07)*. ACM, New York, NY, USA, 178–181. https://doi.org/10.1145/1377999.1378003

[61] Peter Tolmie, James Pycock, Tim Diggins, Allan MacLean, and Alain Karsenty. 2002. Unremarkable Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI ’02)*. ACM, New York, NY, USA, 399–406. https://doi.org/10.1145/503376.503448

[62] 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)*. ACM, New York, NY, USA, 801–812. https://doi.org/10.1145/2750858.2807514

[63] Michel Vacher, Benjamin Lecouteux, Pedro Chahuara, François Portet, Brigitte Meillon, and Nicolas Bonnefond. 2014. The Sweet-Home speech and multimodal corpus for home automation interaction. In *The 9th edition of the Language Resources and Evaluation Conference (LREC)*. Reykjavik, Iceland, 4499–4506.

[64] Niels van Berkel, Jorge Goncalves, Simo Hosio, Zhanna Sarsenbayeva, Eduardo Velloso, and Vassilis Kostakos. 2020. Overcoming compliance bias in self-report studies: A cross-study analysis. *International Journal of Human-Computer Studies* (2020), 1–12. https://doi.org/10.1016/j.ijhcs.2019.10.003

[65] Voicebot.ai. 2019. *LifePod is Taking Reservations for ‘Proactive’ Voice Assistant and Smart Speaker for the Elderly.* https://voicebot.ai/2019/06/27/lifepod-is-taking-reservations-for-proactive-voice-assistant-and-smart-speaker-for-the-elderly/

[66] Voicebot.ai. 2019. *Smart speaker Consumer Adoption report.* https://voicebot.ai/wp-content/uploads/2019/03/smart_speaker_consumer_adoption_report_2019.pdf

[67] Christopher D. Wickens. 2008. Multiple Resources and Mental Workload. *Human Factors* 50, 3 (2008), 449–455.

[68] Alex C. Williams, Harmanpreet Kaur, Gloria Mark, Anne Loomis Thompson, Shamsi T. Iqbal, and Jaime Teevan. 2018. Supporting Workplace Detachment and Reattachment with Conversational Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI ’18)*. ACM, New York, NY, USA, Article 88, 13 pages. https://doi.org/10.1145/3173574.3173662

[69] Jong-bum Woo and Youn-kyung Lim. 2015. User Experience in Do-it-yourself-style Smart Homes. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 779–790. https://doi.org/10.1145/2750858.2806063

[70] Fengpeng Yuan, Xianyi Gao, and Janne Lindqvist. 2017. How Busy Are You?: Predicting the Interruptibility Intensity of Mobile Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. ACM, New York, NY, USA, 5346–5360. https://doi.org/10.1145/3025453.3025946