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

ANONYMOUS AUTHOR(S)

We envision 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 build a voice-based experience sampling device and conduct a one-week field study with 40 participants living in university dormitories. We collect 3,572 in-situ user experience reports and propose 19 activity categories to study 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 this finding, 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:
Anonymous Author(s). 2019. “Hello There! Is Now a Good Time to Talk?": Opportune Moments for Proactive Interactions with Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 0, 0, Article 0 (November 2019), 25 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Smart speakers feature an intelligent voice-assistant, supporting speech-based interactions for a wide range of tasks, including time checking, delivery tracking, and question-asking. Currently, smart speaker services are mostly reactive to user's commands. Recently, Amazon and Google Home, which have top market shares, started considering proactive services such as reminding a user's schedule [51] and supporting home safety and security [13]. Prior studies already demonstrated the usefulness of various proactive services: context-aware reminders/recommendations [8, 49, 55], and self-tracking/reflection for productivity [27, 60].

These proactive services provide useful information for inspiring and engaging users, but prior studies also warned that timing and relevance are critical for the user experience of proactive conversational services [3, 9]. Delivering services at an inappropriate moment could disrupt users' primary tasks, causing annoyance and resumption lag [21], or even resulting in safety risks, in some contexts such as driving [24, 25, 47]. Previous interruptibility studies examined the opportune moments for an interruption in diverse task contexts and computing environments, ranging from task-switching with desktop computers in offices [2, 22], notification

2474-9567/2019/11-ART0 $15.00
https://doi.org/10.1145/1122445.1122456

delivery via mobile phones [33, 40], conversational interactions in vehicles [24–26] to even the actuation of height adjustments by smart furniture [30].

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 the interruptibility of smart speakers would have different contributing factors due to its distinct characteristics as opposed to other devices (e.g., mobile phones and desktop computers) and locations (e.g., office and vehicular setting). For conversational interactions in the context of stationary speakers at home, users could be both mobile and stationary, performing various activities, and speakers could be shared among family members. Earlier smart home research highlighted the importance of seamlessly supporting domestic routines and key activities [14, 53], and that intelligent services may introduce new challenges and responsibilities in dealing with technology [10] as well as affecting interpersonal relationships [4]. Therefore, we set our goal to understand the opportune moments for smart speakers to initiate proactive services while their users are naturally engaged in different 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 proactive conversational interactions with smart speakers in the home context. In the study, we asked users the question, "Is now a good time to talk?" to answer in yes or no, and further inquired contextual reasoning for the provided answers. We reviewed each response during the exit-interview for a deeper understanding of factors that influence the interruptibility related to home activities. Data were gathered over a week for each of 40 participants (20 dormitories with two people per dorm) living in dormitories where diverse domestic activities naturally occur in a studio setting.

From the collected data, we could identify the key factors relevant to the interruptibility of a proactive audio service from a smart speaker in domestic contexts, specifically: (1) personal contextual factors (e.g., engagement, urgency, psychological/physical states, auditory/verbal channel availability), (2) movement contextual 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 as well as 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 considering spatial relationships (or proxemics) [5] and home routines [53] in parallel. The contributions of our study can be summarized as follows:

- We designed a prototype 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 regarding user interruptibitilly and activity contexts.
- We identified three key factors that affect the interruptibility of proactive services: (1) personal contextual factors, (2) movement contextual factors, and (3) social presence.
- Based on our findings, we proposed and discussed how we can realize context-aware proactive conversation management for smart speakers in domestic contexts.

2 RELATED WORK

In this section, we first review prior studies on proactive services by intelligent agents and overview interruptibility research in Ubicomp's field. 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 mostly reactive and are 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 [49] is a location-based reminder app for mobile phones that, for example, can remind users to bring their belongings when leaving. Context-aware reminders are useful for users with memory deficiency, such as older adults with dementia [23, 37]. Braunhofer et al. [8] proposed a proactive recommendation system in the tourism domain to push relevant items to the target user in the right context; the system can suggest users suitable places to visit based on daily weather and locations they recently visited. Researchers also explored conversational interactions in smart home environments. The Sweet-Home project [55] built a voice-based intelligent controller to help older adults easily control appliances in smart homes and to activate proactive services such as prompting a security reminder to lock doors before leaving. Qin et al. [43] 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).

Recent smart speakers such as LifePod, Google Home, and Amazon Echo are exploring various proactive services as well. LifePod [57], a smart speaker for older adults, proactively reminds users of medication adherence or 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; for example, offering friendly reminders such as suggesting to lock a door when a user says "goodnight," or supporting ambient sensing to alert users of security events (e.g., window breaks, or smoke alarm) [13].

These prior studies well demonstrate the usefulness of proactive services in smart speakers or mobile phones. We envision that such services in mobile phones are also applicable for smart speakers. As shown later, one critical aspect of intelligent agents is knowing when to initiate conversations to deliver proactive services. Therefore, we explore in our work how users' contexts are related to finding opportune moments for conversations.

2.2 Interruptibility and Opportune Moments for Interaction

Interruptibility denotes how interruptible a user is when a system attempts to communicate with the user, for example, to deliver a notification. An opportune moment for user interaction refers to a moment when the disruption of interrupting a current task can be minimized [54]. 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 [21], and even safety risks in driving contexts [24, 25]. Interruptions at opportune moments can facilitate the engagement of mobile users with recommended content (e.g., games, news, and surveys) [40].

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 [33]. Recent studies further considered novel, ubiquitous environments. For example, voice interfaces in vehicles can leverage vehicular sensor data to find opportune moments for conversation [24, 25]. 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 [30].

Despite the recent popularity of proactive services in smart speakers, to our knowledge, interruptibility research on a smart speaker in home contexts is lacking. The smart speaker has several distinct characteristics as opposed to other devices, such as: (1) it mainly supports conversational interactions, (2) it can be shared with multiple users (e.g., family members) while the voice-interaction channel is also a shared medium, (3) it is installed at a fixed location (e.g., living room) while users can be mobile, and (4) home environments are also much more dynamic with multi-user/tasking and mobility factors, and in them various domestic routines and activities happen, compared to other environments such as vehicles or offices. We hypothesize that the above characteristics of smart speakers and home environments together engender unique factors affecting the interruptibility of proactive conversational interactions initiated by smart speakers in home environments.

2.3 Contextual Factors for Interruptibility

A user's personal contextual factors such as current activity [39], engagement [16], busyness [62], and emotion [62] 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 he/she is taking on challenging tasks that require a considerable amount of attention, such as biking or driving. Notification contents in association with the current activity also affect interruptibility, e.g., when in a meeting, a chat notification from a friend is disruptive while an email notification from a project collaborator might be acceptable [39]. As proposed in the work of Siewiorek et al. [48] 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 users' agenda.

Engagement is another primary determinant of interruptibility. When performing simple, rote work that is highly engaging but not challenging, users are susceptible to distractions [32]. Yuan et al. [62] found that notifications from a smartphone can be very disruptive when users are busy; depending on user context, busyness levels would vary widely. A user's emotions are also a factor for interruptibility; people are more likely to be interruptible in a pleasant mood than in an unpleasant mood [62]. These studies show that personal contextual factors dictate users' interruptibility in 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. [39] considered four physical activities related to interruptibility, which are 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. [19] 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 indeed signals opportune moments for smart speakers to engage with users. Unlike mobile phones, smart speakers are stationary while users are mobile, and untethered interactions are feasible with speech, which in combination seems to provide unique characteristics to interactions with smart speakers. Thus, our second research question is as follows: *RQ2) What movement contextual factors are relevant to interruptibility concerning proactive interactions with smart speakers?*

As discussed earlier in section 2.2, we also consider social presence factors because a smart speaker is a shared device accessible by any user in its vicinity. Prior studies explored a social relationship in online contexts to understand interruptibility [33]. The social presence of other people in a shared space (e.g., roommates or visitors) affects the interruptibility; Park et al. [38] studied interruptibility of mobile notifications during social activities and found that conversation breakpoints are generally opportune moments for interruptions. Beyond mobile interactions such as 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?*

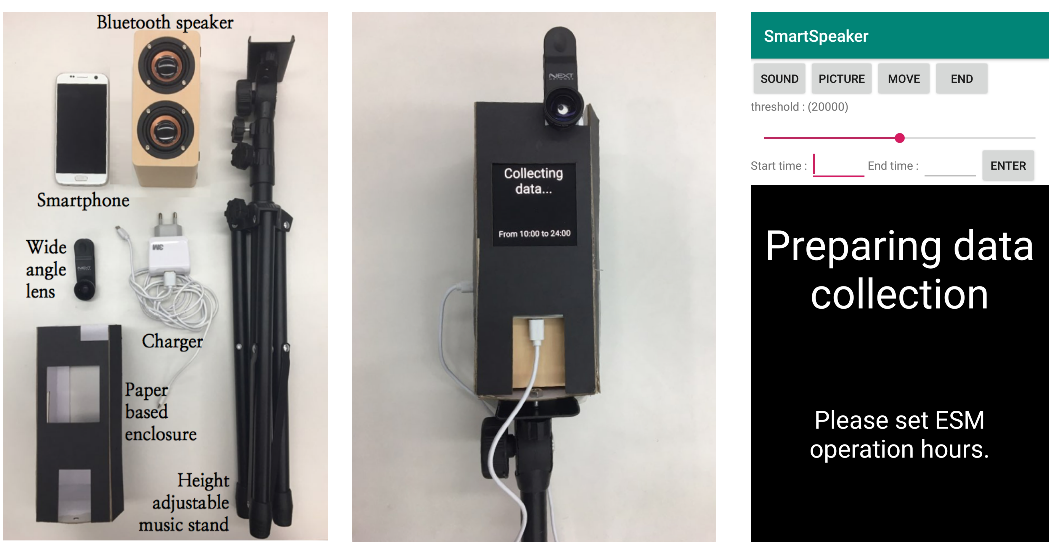
3 METHODS

Our goal is to understand the interruptible moments during natural activities at home. We designed a smart speaker that prompts voice-based ESM messages, which are triggered by two mechanisms: randomly and through movement detection. In this section, we explain how we designed ESM procedures and smart speaker systems, how we recruited participants, and how we collected and analyzed data. Our study was reviewed and approved by the university’s Institutional Review Board (IRB).

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. This study focused on a university dormitory where rooms are shared. This setup helps us to study how the presence of roommates affects the interruptibility of voice interaction. Common dormitories have only one room with two single beds; other amenities such as restrooms, shower rooms, laundry rooms, or kitchens are shared with other students. A dormitory environment has limitations, but it enables 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, which consists of a smartphone that runs an ESM app, a Bluetooth speaker, a wide-angle lens, a height-adjustable music stand, and a paper-based enclosure. A smartphone was used for this purpose. As shown in Figure 1c, we developed an ESM application that supports following features: (1) checking ESM volume, (2) setting operation hours, (3) configuring the movement detection threshold, and (4) starting ESM data collection. The volume level of ESM can be adjusted to prevent unpleasant experiences from loud sounds. Setting operation hours helps us customize ESM working hours depending on each participant’s daily schedule. The threshold controller aims to increase the accuracy of movement detection, which we explain in detail in Section 3.1.4. Throughout data collection, the smartphone was set to display configuration information and to automatically synchronize ESM data files to the cloud storage in real-time.



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

Fig. 1. Smart speaker device for data collection

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 movement using the smartphone's camera, we attached an extra lens on the smartphone's built-in camera lens. This lens supports a 170-degree wide-angle. A black paper-based enclosure was designed to put all components in one place 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 the eye level to accurately detect movements within the whole room.

3.1.3 ESM Method. In the dormitory environment, the smart speaker prompts voice-based ESMs to check a user's interruptibility and collect their contexts. As a smart speaker supports conversational interaction, ESM should be conducted via speech. Prior studies have probed interruptibility via several questions such as whether the timing is good [47], whether it is disruptive [18], or whether a user is available to conduct a requested task [16].

ESM Question: We focused on *timing* for conversational interactions because it can be typically interpreted as an opportune moment for conversations, reflecting a user's preference and self-assessment of availability. Thus, the following question was used: "Is now a good time to talk?" A recent study on proactive audio services in vehicles used a similar voice ESM question, "Is now a good time?" [47] to explore on-road driving interruptibility. We intentionally did not associate a specific service with the ESM question. This enables us to focus on contextual factors of interruptible moments because we can minimize a user's bias on interruptibility due to content preferences [40]

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 turn-takings (bi-directional interactions)". At orientation before data collection began, we instructed this meaning of proactive services and gave the following three example scenarios:

- *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 an ESM message, "Is now a good time to talk?" we first asked participants to answer in "yes" or "no", and further verbally describe what they are currently doing to collect user activity contexts, which provide grounds for the reported interruptibility. To collect data in natural contexts, we asked participants to respond to ESM only when they are at home and that they are not required to be at home intentionally to answer ESM questions. When both participants were at home, we asked them to answer consecutively. This instruction is because each participant may be present in different contexts, and each of their interruptibility may differ in the given situation. Even if they are engaged in the same activity, interruptibility could be different because of personal reasons. Similar to prior studies on smart speakers [6, 41], we recorded voices for 1 minute after an ESM prompt, which is sufficient to fully capture user responses (as confirmed via our pilot study). The recorded audio data was uploaded in real-time, and researchers then examined the recorded ESM audio data by transcribing the ESM answers.

The recorded ESM audio data can be categorized into three types based on their completeness: (1) complete answer, (2) incomplete answer, and (3) unanswered. A complete answer means that users reported an interruptibility decision (yes/no) and detailed contextual information. An incomplete answer happens when users fail to report an interruptibility decision or when their answer is insufficient to determine its context. Unanswered means there is no answer at all. To check the incomplete or unanswered cases every day, researchers set up group chat rooms with participants for each dormitory using a popular messenger app (i.e., KakaoTalk).

For incomplete cases, researchers asked participants to supplement their answers via a messenger by presenting the time of an incomplete ESM answer, and consecutive ESM answers around the incomplete answer (previous and next answers, if any) as cues for recall. Unanswered cases are most likely when there was nobody in the room. A smart speaker cannot discern if there are people in the room despite movement detection. For that reason, we asked 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. If nobody was there, we skipped examining unanswered cases. However, if somebody was there in the room, but no responses were captured, possibly because participants were listening to music using a headset or sleeping, we assumed that it was not an opportune moment. Then, similar to incomplete cases, we asked what they were doing at that time and why they did not answer by presenting cues for recall.

ESM Parameter Selection: ESM parameters, including operating hours, the minimum interval between consecutive prompts [56] were determined to collect a sufficient amount of data. Our goal was to collect 10 ESM answers per day for 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 (circadian rhythms). Similar to our approach, prior studies also limited ESM questions to be prompted during waking hours [34, 35]. Since roommates share the room as well as the smart speaker, we asked them to agree upon the operating hours of ESM mutually. We allowed them to change operating hours during the data collection period, but most participants maintained their initial settings.

Despite over 12 hours of ESM operating time, the actual time of participation in ESM answering was far less because of participants' frequent outdoor schedules (e.g., attending classes, eating out, meetings). Our recruitment criteria included users who stay at home at least three hours a day, which do not need to be in contiguous segments, to collect adequate natural activity contexts. For these reasons, we set their participating hours at three hours a day. To collect 10 ESM answers within 3 hours, time intervals between prompts were determined to be at least 20 minutes to minimize disruption, which we explain later (see Section 3.2).

ESM Trigger Strategies: In our study, ESMs prompts are delivered randomly or when movements are detected. Previous studies [19] reported that users are likely to be interruptible during activity transition moments (e.g. from sitting to walking) when mobile devices deliver messages; these messages are more positively received during activity transitions than when delivered at random times. We want to investigate if the transition period between physical activities is an opportune time to interrupt users. The key difference from prior studies is that unlike mobile phones, smart speakers are immobile and fixed at one place.

We implemented the ESM app so that experience sampling prompts are randomly triggered or by movement detection at approximately the same interval. When a sampling prompt is delivered, the next sampling prompt is reserved in a random interval between 15–25 minutes to maintain the 20 minutes time interval, on an average. When movements are detected (after a few seconds), a sampling prompt is then triggered immediately. However, to guarantee an interval of at least 15 minutes between sampling prompts, movement detection is set to work only after 15 minutes from the triggering of the last prompt. If no movement is detected after the 15 minutes window from the last ESM, a random ESM is triggered at the reserved time; however, if a movement is detected before the reserved time, the reserved ESM is canceled and the detected ESM is triggered.

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

For ground truth verification when a movement was detected, we saved pixel differences (i.e., changed pixels in black and unchanged pixels in white), thereby expunging personally identifiable information—this was approved by IRB. 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 say someone is moving from the bed to the desk, or vice versa (see in Figure 2).

We conducted a validation test to confirm that our movement detection approach was valid and to properly configure the two threshold values (i.e., pixel-level difference and photo-level difference thresholds). 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 include small or momentary movements such as moving only hands or arms, turning while lying on bed, leaning

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

Fig. 2. 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.

left or right while sitting on a chair, tilting head, and changing sitting postures. After setting our test environment simulated the furniture arrangement in dormitories (e.g., a pair of desks, beds, closets located at both sides), we recruited seven volunteers. We provided an activity script that included both slight and large movements. While the participants were moving as instructed in the script, the smart speaker app took pictures every 3 seconds. Consequently, in total, we obtained 203 movement-labeled pictures.

We calculated false-positive rate (FPR) and the 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 are mostly concentrated within this range in our dataset. 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, each room's arrangement could be slightly different. Based on our threshold results, when we deployed our ESM devices, we also tested basic movement scenarios in the 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 inconvenience that could affect data collection. This was conducted with four participants in two dormitories (a pair of females, and another pair of males) for two days, 12 hours per day. This duration was sufficient to find potential inconveniences in ESM-based data collection. We asked the participants to actively report any inconvenience 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 participants' daily routine schedules seems to be helpful; and this feature was added in the ESM software. Negative experiences from proactive announcements (e.g., surprise and sharp voices) 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 revision 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 dormitory in four days. Consequently, one pair answered 156 prompts but the other pair only answered 36 prompts, which is far below our goal. After the interview, we found that this low level of participation was because of the fact that participants stayed out of their homes quite often for a long time. Based on this observation, we added a recruiting requirement: participants typically spending more than 3 hours a day at home.

3.3 Recruitment and Data Collection

We recruited 40 participants from 20 dormitories via an online campus community and Facebook (12 male pairs, and 8 female pairs). As illustrated earlier, we recruited students who live in a dormitory with a roommate (two persons per dormitory), and both roommates should participate in the data collection together. Our restriction was that users should stay at the dormitory at least 3 hours a day. ESM data was collected for a week (including weekend).

The user study began with an orientation. The participants signed an IRB consent form. We illustrated the purpose of data collection, along with how a smart speaker works, what types of conversational tasks in ESM questions are assumed by giving the aforementioned examples, and how they should answer ESM prompts. For each dormitory, 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 (e.g., to supplement incomplete or unanswered ESM if needed). Then, the researchers visited each participant's dormitory to install the smart speaker device, set operating hours as decided by the roommates, and manually adjusted movement detection thresholds.

After data collection, we performed 1:1 exit-interviews with 40 participants. Each interview session lasted for about 20 minutes on an average. 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 the same questions about all ESM data to each participant. Additionally, we asked questions about extreme user experiences, such as the most interrupting/irritating/preferred moment. Probing extreme user experiences about interruptibility helped us understand the opinionated feelings of proactive interactions with smart speakers.

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 [11]. We created a code book by categorizing similar home activities. Focusing on specific contexts (e.g., activity types and movements) helped us in overviewing interruptibility patterns. Exit-interview data supplemented the quantitative responses as we could understand the detailed reasons why they felt interruptible (or not) while doing a certain activity, and uncover other factors affecting to interruptibility.

4 ACTIVITY CATEGORIZATION

4.1 Data Preparation and Coding Process

A total of 3572 responses were collected; the maximum number of responses collected from a single pair of participants (P33, P34) was 324 while the minimum was 84 (from a pair of P30, P39). Responses were collected over a period of a week. ESM operational hours per day varied from 9-16 hours (M = 13.1, SD = 1.5). We excluded 28 responses, where we could not infer the interruptibility or context from the response despite conducting post-collection interviews to supplement incomplete responses and clarify unanswered ESMs. Participants could not remember the context or were unsure about what happened. In addition, we excluded 44 responses from a participant who misunderstood our ESM's goal. 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 of data with 284 samples from 10 dormitories, selected from respective days with the most amount of data collected for each dormitory. 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 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) between 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 acclimatize to the process and increase their reliability, the coders practiced with the sampled subset prior to coding the entire data. 3288 samples, excluding 284 samples used for developing the preliminary coding scheme, were divided into two sets with 1642 and 1646 samples respectively, with each set being 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.

As shown in 'All (RQ1)' column of Table 2, the most frequently reported home activity was working/studying (n = 824). Whereas, the least frequently reported activity was online interaction (n = 62). As shown in 'Overall' row – the last row, 53% (n = 1859) of the responses were "Yes (interruptible)," implying they were interruptible for conversation, while 47% (n = 1642) of the responses were "No (not interruptible)." 80% (n = 2806) of responses were collected from randomly prompted ESMs, while 20% (n = 694) were prompted via movement detection.

The majority of responses categorized as interruptible were from the following activities: returned from outside (98%, the proportion of answers categorized as "yes" for the activity), Internet/smartphone (92%), returned from other rooms (89%), resting (85%), doing chores (85%), watching videos (72%), after waking up (65%), changing clothes (64%), face or body caring (64%), hair caring (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 interaction (40%), and online interaction (45%).

Home activities tend to periodically repeat across time of the day and day of the week. Table 3 shows the percentage of interruptible cases ("yes") across the day of week and hour of the day. It is notable that participants were more interruptible for the morning (9am-11am) than other times of the day while their interruptibility was not varied across the day of the week. For the morning, the major activities were visiting outside the dorm (16%), hair caring (16%), visiting other rooms (14%), face or body care (10%), sleeping (9%), and after waking up (9%). Of these activities represent 72% of home activities during the morning. The rest activities include resting (7%), working/study (6%), changing clothes (4%), internet/Smartphone (4%), watching videos (2%), before sleeping (2%), returned from other rooms (1%), eating (1%), and online interaction (1%).

The majority of responses categorized as uninterruptible were from the following activities: visiting outside the dorm (0%, the proportion of answers categorized as "yes" for the activity), sleeping (0%), visiting other rooms (0%), before sleeping (0%), hair caring (2%), visiting other rooms, online interaction (33

5 RESULTS

By analyzing the ESM responses and the exit-interview data, we identified RQ1) the key personal contextual factors (i.e., concentration and engagement, urgency and busyness, psychological/physical states, and auditory/verbal channel availability), RQ2) the 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 them the reasons for responding with a "Yes" or "No" in particular situations, and extracted four personal contextual factors affecting interruptibility, which are as follows: 1) concentration and engagement, 2) urgency and busyness, 3) psychological or physical states, and 4) auditory/verbal channel availability. Note that the influence of activity transition and social presence will be discussed in the subsequent sections.

5.1.1 Concentration and Engagement. Whether a person is concentrating on his/her current activity at the moment of interruption is crucial in determining if he/she is available or unavailable for interaction. Depending on their characteristics, we further divided the activities based on their required level of concentration. When participants were focusing on a particular activity requiring a high level of attention, they frequently reported as not being interruptible. This mostly happens while working/studying, video gaming, and face-to-face interaction, 79%, 64%, 60% of which were respectively reported not to be interruptible. However, when they lost their focus or did not need to concentrate for a while, they reported being interruptible as they could temporarily switch to

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

Category	Activity	Description	Example
Working/studying	Working and studying	Working or studying	• *"Just sat down at the desk to study."* • *"I'm focusing on writing an email."*
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."*
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."*
Eating	Eating	Eating a meal or a snack	• *"I'm having some snack while lying."*
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."*
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."*
Just returned	Returned from outside the dorm	Just came back home (e.g., after classes, having meals, or meeting friends)	• *"Just came in after having a dinner."*
	Returned 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."*
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 (Kakao Talk)."* • *"I'm talking to my dad (on the phone)."*
Miscellaneous	Miscellaneous	Any other activity not mentioned above	• *"Just thinking about the dinner menu."* • *"I'm looking for something."*

Table 2. Distribution of interruptibility 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		Movement detected ESMs							
								All		Movement context					
										Ent/Dep		Transition		Others	
				#case	#yes	#case	#yes	#case	#yes	#case	#yes	#case	#yes	#case	#yes
Using media	1124 (32.1%)	776 (69%)	Video gaming	310 (9%)	113 (36%)	288 (10%)	97 (34%)	22 (3%)	16 (73%)			10 (7%)	10 (100%)	12 (5%)	6 (50%)
			Internet / Smartphone	381 (11%)	351 (92%)	359 (13%)	331 (92%)	22 (3%)	20 (91%)			4 (3%)	4 (100%)	18 (7%)	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 (23.5%)	177 (21.5%)	Working and studying	824 (24%)	177 (21%)	743 (26%)	134 (18%)	81 (12%)	43 (53%)			41 (28%)	30 (73%)	40 (15%)	13 (33%)
Resting	433 (12.4%)	369 (85.2%)	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 (9.6%)	115 (34.2%)	Visiting outside	251 (7%)	93 (37%)	101 (4%)	38 (38%)	150 (22%)	55 (37%)	150 (53%)	55 (37%)				
			Visiting other rooms	85 (2%)	22 (26%)	27 (1%)	5 (19%)	58 (8%)	17 (29%)	58 (21%)	17 (29%)				
Sleeping	209 (6.0%)	65 (31.1%)	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 (4.2%)	94 (63.5%)	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%)	18 (6%)	17 (94%)			19 (7%)	11 (58%)
			Face or body Care	66 (2%)	42 (64%)	50 (2%)	34 (68%)	16 (2%)	8 (50%)					16 (6%)	8 (50%)
Social interaction	129 (3.7%)	55 (42.6%)	F2F interaction	67 (2%)	27 (40%)	31 (1%)	8 (26%)	36 (5%)	19 (53%)			2 (1%)	1 (50%)	34 (13%)	18 (53%)
			Online interaction	62 (2%)	28 (45%)	52 (2%)	27 (52%)	10 (1%)	1 (10%)					10 (4%)	1 (10%)
Eating	101 (2.9%)	61 (60.4%)	Eating and drinking	101 (3%)	61 (60%)	74 (3%)	42 (57%)	27 (4%)	19 (70%)			10 (7%)	9 (90%)	17 (6%)	10 (59%)
Just returned	81 (2.3%)	77 (95.1%)	Returned from outside	53 (2%)	52 (98%)	14 (0%)	14 (100%)	39 (6%)	38 (97%)	39 (14%)	38 (97%)				
			Returned from other rooms	28 (1%)	25 (89%)	11 (0%)	9 (82%)	17 (2%)	16 (94%)	17 (6%)	16 (94%)				
Doing chores	66 (1.9%)	56 (84.8%)	Doing laundry or cleaning	66 (2%)	56 (85%)	26 (1%)	22 (85%)	40 (6%)	34 (85%)			3 (2%)	3 (100%)	37 (14%)	31 (84%)
Miscellaneous	49 (1.4%)	14 (28.6%)	Miscellaneous	49 (1%)	14 (29%)	36 (1%)	8 (22%)	13 (2%)	6 (46%)			1 (1%)	1 (100%)	12 (5%)	5 (42%)
Overall	3500 (100%)	1859 (52.8%)	Overall	3500 (100%)	1859 (53%)	2806 (100%)	1440 (51%)	694 (100%)	419 (59%)	282 (100%)	143 (70%)	147 (100%)	120 (81%)	265 (100%)	156 (50%)

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	0.0	20.0	25.0	0.0	50.0	40.0	0.0	19.3
10	26.7	41.7	39.1	52.9	40.0	35.0	0.0	33.6
11	67.6	77.8	64.7	50.0	52.9	47.8	85.7	63.8
12	64.1	67.9	59.6	67.5	70.3	30.8	53.8	59.1
13	57.5	45.5	28.6	71.4	33.3	56.7	56.5	49.9
14	63.3	69.6	52.4	45.0	41.4	57.1	58.8	55.4
15	71.8	60.0	50.0	58.6	50.0	66.7	37.8	56.4
16	54.3	56.4	71.9	47.4	65.9	46.2	48.8	55.8
17	48.3	50.9	50.9	58.0	42.9	60.9	50.0	51.7
18	53.2	62.0	54.2	59.1	62.9	61.9	46.2	57.1
19	60.5	58.2	57.1	63.8	48.3	42.1	62.1	56.0
20	62.2	51.6	38.0	35.0	57.1	38.5	63.4	49.4
21	50.0	49.2	53.8	51.9	37.5	57.6	54.3	50.6
22	51.0	60.0	49.0	49.0	36.1	30.0	52.8	46.8
23	54.2	51.1	53.7	59.0	48.7	50.0	38.2	50.7
Overall	52.3	54.8	49.9	51.2	49.2	48.1	47.2	50.4

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 am available when I was taking a short break from work."* Likewise, P31 commented, *"There are some games where you must complete a series of stages to finish. It is difficult to pause in the middle."* Some activities like changing clothes can be completed in a short time and may even take a shorter 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 the smart speaker. For example, P14 said, *"I can finish it quickly [changing clothes], but if it is interrupted in the middle, I felt that it would take longer."* In contrast, while performing activities requiring low concentration or engagement, our participants were generally interruptible. Such activities include 1) using the Internet/smartphone (92% was reported to be interruptible), 2) resting (85%), 3) doing chores (85%), 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 converse."* Another participant (P13) commented, *"If it was a really funny or important scene from YouTube, I said no, but if it was a trivial scene, I said yes."*

5.1.2 Urgency and Busyness. Urgency and busyness significantly influenced the interruptibility. When participants were facing immediate deadlines (e.g., assignments due, class schedules), they tended to report being unavailable at the moment. For example, P25 said, *"I was in a hurry to finish my homework."* Participants needed to quickly leave the room because of their upcoming schedules. P22 said, *"When I am late for my class, I need to prepare for leaving, and I do not have time to converse."* Similar patterns occurred 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 bothering to converse 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. These states are 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 the speaker. P5 said, *"[before sleeping] If I talk, then I become awake."* In this situation, 11% was reported to be interruptible. While participants were sleeping, the sound of the ESM prompt also woke them up, which made them respond to the prompt while not being fully awake. 4% was reported to be interruptible. Likewise, after waking up, awaken state affects their interruptibility. P12 said, *"[after waking up] I was not ready for a conversation, and I still felt really tired."* However, participants reported being more interruptible (65%) after waking up compared to when preparing to sleep (11%) or while sleeping (4%). In addition, participants were unwilling to interact with smart speakers when they were not in a good mood for talking (n = 16), were tired (n = 22), or were sick (n = 8); for example, *"[while studying] I was not in a good mood to converse. I felt a bit bothered to talk to and I was tired."* [P22]. This was also true when users were resting; *"Because of a headache, I lied 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 to engage with the speaker; *"I just came back from the store. I bought some food and I felt good to eat it. And I said yes."* [P21].

5.1.4 Auditory / Verbal Channel Availability. If a user's ability to hear or to speak is obstructed, he/she is incapable of interacting with a smart speaker. Some activities such as blowing hair with a hand dryer produced loud noises and interfered with hearing, such as when watching videos: for example, *"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 may obstruct their speech: *"It was awkward to talk while I was brushing my teeth."* [P37]. When participants were having face-to-face interaction (#yes = 40%) or online conversations (#yes = 45%), 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."*, *"I was talking with my friend, and it is difficult to talk to it simultaneously."* [P6]. Interestingly, users' availability was also dependent on their level of engagement. *"It was not an important conversation and it was trivial. (interruptible)."* [P14].

5.2 RQ2: Movement Contextual Factors

We then examined how interruptibility was related to movement related factors. Among movement detected ESMs, 59% of responses were marked as interruptible, whereas among randomly prompted ESMs, 51% of responses were interruptible (see 'Overall' row – the last 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 brief movements during sedentary activities. Overall, we found distinguishable differences in interruptibility depending on the movement subcontexts. Participants were generally interruptible for entrance (96%) and activity transition (82%), but they were not interruptible for departure (39%). Unlike these subcontexts, personal contextual factors (discussed in RQ1) were considered important in other dynamic activities.

5.2.1 Entrance and Departure: We found that interruptibility during entrance and departure in the vicinity of the smart speaker was closely related to the communication range, that is, being close enough to hear the sound or speak with the smart speaker. Inbound movements toward the communication range of a speaker (i.e., entrance) are more interruptible than outbound movements (i.e., departure).

Entrance contexts include the event of a 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 the building and other rooms were interruptible, respectively. An inbound movement into the communication range enables a user to talk with a smart speaker, especially because they just came back and did not start any other activities yet. 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 a dormitory, and visiting other rooms in a dorm, 37% and 29% of which were respectively reported to be interruptible. Departure activity is an outbound movement from the communication range, which often makes it hard to start an interaction with a smart speaker. When leaving, our participants often mentioned that they were in a hurry. P25 said, *" It is a bad moment when I am in a hurry to prepare to go outside, such as when wearing clothes or applying makeup. If I do the same preparation activities but have free time, I would be interruptible, though."* When leaving, our participants often exhibited routine behaviors which would be automatically done without deep thoughts (e.g., going to a restroom, or leaving the room after taking a shower basket). These routine behaviors were more often observed in the instances of visiting other rooms in the dormitory. They said that it is often disturbing to stop and physically rollback such routine behaviors in the middle. P25 said, *"It was not an urgent moment, but if I would talk with a smart speaker, I have to come back and get out again, which is not preferred."* P31 said, *"I was not interruptible because I decided to do laundry and pick up my clothes."*

5.2.2 Physical Activity Transition. Physical activity transitions included the following contexts: about to do something or moving to do something (ESM response examples: *"I am about to study," "I am going from a chair to a bed for rest"*). As shown in 'Overall' row – the last row of Table 2, interruptibility ratios of movement-based ESMs related to physical activity transitions (81%) were higher than those of randomly prompted ESMs (53%). Especially, for the working/studying activity, there was a notable difference in the interruptibility ratio; 73% of responses was yes (or interruptible) in activity transition contexts, whereas it was 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 got up from my chair. I guess this would be detected as a movement."*, *"When I moved from a bed to a chair to sit, I think, I could just talk while walking."* [P39], *"This was the moment after I brought delivered food and was about to set it up on a table. I could do that while talking because it was before I started to eat."* [P34]

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 (e.g., hanging laundry), which require physical movements, excluding entrance/departure activities or activity transitions discussed earlier. Besides, short movements could be detected during sedentary activities. Our manual examination of pixel differences in images revealed that some sedentary activities involved brief movements such as bringing something, or bending down to find something while studying, or moving around while conversing. In some cases, small user movements were wrongly detected (e.g., moving back and forth sitting on a chair, changing sitting postures). In these cases, interruptibility was largely dependent on users' personal contextual factors; multitasking was likely to be feasible when doing light chores (e.g., hanging laundry), whereas it was less feasible when doing cognitively demanding or urgent tasks.

5.3 RQ3: Social Presence

As shown in Table 4, we classified entire ESM responses (3500 responses) depending on social presence (i.e., whether a roommate was present). Among all the responses, 1574 (45%) responses were collected when both roommates were present. These responses can be further categorized into two contexts: 1) doing activities together (e.g, chatting together) and 2) separately doing their own activities. Overall, influence of social presence on interruptibility was relatively small with the yes ratios; i.e., Alone: 1066/1926 (55.35

Our analysis on interview data and ESM responses revealed that when roommates were doing the same activity together, interruptibility of this joint activity was largely dependent on their current level of engagement. When they were separately doing their own activities, their interruptibilities were dependent on their own activity contexts. However, when their voice interaction with a smart speaker could interrupt the activity of the other party, or another roommate (e.g., sleeping or studying), they reported that they were not interruptible to minimize possible interpersonal conflicts, but this pattern also differed across different users (e.g., roommates' sensitivity).

5.3.1 Doing Activities Together. Roommates can do the same activity together; e.g., talking to each other, playing a game together, studying and discussing some topic, eating meals together, and watching Youtube by sharing a device. In these contexts, 98 responses (47 pair cases) were collected and interruptibility was dependent on their current engagement. For example, P37 said, *"I was having in 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 conversation topics and YouTube contents were not important, so I was able to answer enough."*

Although participants were engaging in the same activity together, their responses for interruptibility sometimes did not match with one another (14 responses; 7 pair cases). Those cases were mainly when they were performing different sub-tasks or being in different stages. For example, when playing a computer game together, one person lost a life in a game, waiting for the other who was still alive. In this case, their responses were yes (a person who was waiting) and no (a person who was still playing), respectively. Their responses were also varied when one of them was performing additional activities at the same time (i.e., concurrent multitasking). For example, when two roommates were talking together, one roommate was heading to a restroom, which was inside the room. In this case, the response of the one roommate was no while another was yes.

5.3.2 Separately Doing Their Own Activities. When roommates were separately doing their own activities, 1476 responses (738 pair cases) were collected. When performing separate activities, participants considered the other party's activity in the room (e.g., sleeping). When the sound of voice interaction interrupted the other party, they 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 moment."*

During the exit-interviews, we further investigated the extent of influence of the other party's contexts (i.e., sleeping and studying) on their decisions. We asked the following questions: Will you change your answer, even though you were interruptible, your roommate was asleep? Likewise, we also asked the same question when their roommate was concentrating on studying. In our interview, we found that 29 participants (72.5%) preferred not to talk when their roommate was asleep, whereas it did not matter to 11 participants (27.5%) because they claimed that their roommate was not sensitive. For example, P18 said *"I know that he would sleep very well, even if*

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	Overall		1574	793	781
	Activity context	Doing activities together	98	43	55
		Separately doing own activities	1476	750	726
Overall			3500	1859	1641

I talk with a smart speaker." Studying was less influenced than sleeping. There were 11 participants (27.5%) who preferred not to talk when their roommate was concentrating on studying; this number is far smaller than that of sleeping. P25 said "*Usually, we freely initiate a conversation while each other is studying. So I wasn't sensitive [interruptible] when she was studying. But when sleeping, she should not be awaken [not interruptible].*"

6 DISCUSSION

We review our findings in relation to prior studies, present design insights for context-aware proactive conversation management 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, and social factors — which show different characteristics as opposed 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 context 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 interruptibility studies; users engaging with challenging tasks are less interruptible [39] than those participating in rote tasks [32]: i.e., participants were interruptible for 21% of cases while working or studying versus 85% while doing chores. Users' current states such as busyness and mood may negatively affect interruptibility [62]: e.g., our interview data revealed uninterruptible moments 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 [59], our users could perform two concurrent tasks (or dual-tasking [45]) by utilizing different types of mental/motor resources as long as an 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 auditory and verbal channels of a user and limited the user's availability for speech-based interaction.

Movement contextual factors are also closely related to interruptibility. Users are generally more interruptible after the entrance and physical activity transition events, but less interruptible during departure; participants were interruptible for 98% of cases upon returning to their rooms (100% if a movement detection triggered an ESM), but only of 37% of departures were reported interruptible. While high interruptibility during activity transition is unsurprising as it was previously shown [19, 39], that people exhibit different levels of availability for interaction during entrance and departure events suggests that a communication range and a user's movement respect to that range are critical to the interruptibility of 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 since smart speakers are often located in a shared space. Participants were interruptible for only about half of ESMs triggered while they were engaging in face-to-face (52%) or remote/online (48%) interactions. In prior interruptibility studies [33], 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 like a living room where social interactions frequently occur [58], and loud voice interactions in such shared spaces may disturb others. Therefore, the interruptibility of smart speaker users may be contingent upon the interruptibility of co-located users. If co-located users engage in 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 (e.g., sleeping or studying) to minimize the chance of discomforting the other. Our interview data indicated that the decision process was also influenced by cohabitants’ traits such as sensitivity to ambient environments. As shown in an earlier study [9], a shared space brings up new issues of timing, relevance, and privacy in smart home environments.

6.2 Implications on Privacy Concerns

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 the 3-second interval when a significant movement was detected, and (2) recorded surrounding sounds including voices for two minutes, starting from 1 minute before the beginning of an ESM. As our prototype needs to collect and process privacy sensitive information in real time, we asked participants during exit interviews if they had any privacy concerns.

All of our participants were well aware of our approach, which was instructed at orientation before data collection started. To our surprise, the majority did not express any concerns. One of the main reasons was that they estimated potential privacy risks to be 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 trust in researchers that we will 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 assured against the privacy violation, similar to P39: “Because you mentioned the school’s policies on the treatment of collected data, I didn’t worry.”

Several participants reported mild concerns that they felt uncomfortable for being photographed, recorded, or both, but this lasted only for one or two days in the beginning. P24 commented “Maybe I had a bit more [concern] in the beginning, but I became used to it more I went on [with the experiment].” However, 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 for audio recordings?’) No, I didn’t think so.”

The participants’ overall sentiment on the issue of privacy threats was at most 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 violation is likely not too detrimental, in comparison to more severe instances like data leakage by companies. Our participants’ such behaviors are representative of findings from a recent study, which reported a users’ lack of privacy concerns on the data practice 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 the privacy risks [29, 50].

The exit interview results also revealed that participants may accept, even willingly, to compromise their privacy in exchange for a smart speaker that offers personalized care. P7’s remark is as follows: “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.” P7 further noted, nevertheless, 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.”

6.3 Towards Context-aware Proactive Conversation Management

Proactive services with smart speakers could bring benefits (e.g., providing useful information, inspiring users) as well as challenges (e.g., timing, privacy/surveillance concerns) [9]. In social contexts, they may result in interpersonal conflicts by disrupting other's work as well as interpersonal facilitation by promoting playful interaction and more efficient household task handling [4]. Our findings clearly demonstrated that it is important to carefully consider everyday activities in home contexts when we determine timing of proactive service delivery. This concurs with the claim of Tolmie et al. [53] that ubiquitous computing should seamlessly augment and support everyday routines at home, including awareness of people's 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 [44] discussed 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 [10]. Thus, it is critical to carefully consider not only spatial relationships (or proxemics) [5] but also domestic routines [53] for technology integration and evolution [10, 44].

Toward this goal, we envision designing a *proactive conversation management* feature for proactive smart speakers; e.g., determining when to start, pause, or resume conversation by analyzing users' contexts. Supporting proxemic interactions for smart speakers basically assumes that devices can acquire fine-grained knowledge about nearby users and their devices, including identity and position/movement/orientation [5]. This context information can be used to control/trigger user interactions as well as coordinate multi-user interactions. In our work, since a user's mobility is an important factor for interruptibility (e.g., less interruptible when leaving), a user's mobility context 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. For fine-grained control of triggering conditions, we can leverage Hall's interpersonal distance (e.g., personal vs. social vs. public spaces) [17], by discretizing space around the speaker.

Existing proxemics 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 as well as their personal contextual factors. There are various ways of supporting context-awareness in home environments. For example, we can leverage existing tools to recognize routine behaviors based on acoustic fingerprinting [28]. Computer vision or wearable sensing techniques can be used to recognize current activities as well users' contextual information including engagement, mood, and stress [20, 40]. As in prior Ubicomp studies [24, 39, 40], 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. Also, it is possible to routinize proactive conversational interactions based on simplified user feedback or end-user programming methods [15, 61], or to help users to handle the deviations and exceptions from routines via human-in-the-loop system design [36]. Conversational management can leverage prior studies on identifying conversation breakdowns and devising repair strategies in diverse domestic setting [6, 41], or analyzing conversational log data to understand interaction patterns for context adaptation and personalization [7, 46].

6.4 Emotional Attachment and Interruptibility

Our exit-interview points to another compelling design implication for smart speakers, that emotions invoked by the device may impact the interruptibility of proactive services. We had several participants during the exit interview, referring to speakers as their ’friends’. Notably, P19, who named the speaker “Jarvis,” said it will be “too bad to stop the experiment” because they together have been like a family for one week. Similarly P13 said “it was good that a speaker talks 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 their speakers, despite the short duration of the experiment, and the limited interaction capacity of the speaker. That attachment may cause people to find it uneasy about rejecting smart speakers coldly. Another comment from P9 pointing out that she did not feel obliged to respond to the speaker because she had no reason to return, as the speaker is “something like a robot,” which “throws questions without emotions” hints that there indeed is 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 the human-likeness of an interacting device is a critical determinant of a user’s interruptibility. This is also related to personification of smart speakers. Prior studies [31, 42] hinted that device personification is related with sociability, politeness, and customer loyalty. There is less chance that a message is neglected if it feels natural, and a flesh-and-blood person or an 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 attempts to create a system that interacts with us as humans do, thereby promoting emotional interactions. In that regard, a device’s ability 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 opportune moments in which a smart speaker can proactively interact with their users: (1) just returned (95%, the proportion of answers categorized as “yes” for the activity), (2) internet/smartphone (92%), (3) doing chores (84.8%). Upon their return, people were willing to interact with the smart speaker. Users are unlikely to be occupied with any task when they just returned to their rooms and have a higher chance of being open for an interaction with a smart speaker. Smart speakers could utilize a set of internal sensors to detect particular domestic routine contexts, including returning home, signaling opportune moments for the speakers. A smart speaker generally supports Bluetooth or/and WiFi connections. Such connection signals of their users’ smartphones can be used to detect their home return. Multiple users may exist in home environments. Depending on which user’s smartphone signal appears, the speaker could provide personalized services to the particular user. Similarly, the speaker could monitor traffic usages of the home WiFi network to detect home return as well as internet/smartphone usages. In our study, we used a camera to detect movement contexts (see Section 3.1.4). Smart speakers could utilize an internal microphone sensor and analyze surrounding sounds (e.g., human voice and ambient noise - doorstep and door open sound). Similarly, Sensay [48], a context-aware mobile phone, analyzed voice sound (e.g., speaking or not) and ambient noises (e.g., low, medium, high noise levels) to infer the current context or environment (e.g., having a meeting, sitting alone at a computer). [1]. In addition to an internal microphone sensor, smart speakers could also utilize external devices and their embedded sensors. For example, IoT home devices (e.g., turning on or off a television, open or close a fridge door, turning a light on or off) or smartphones carried at home (e.g., indoor localization) can be used to sense a user’s movement in a home environment. Furthermore, entrance detecting sensors (e.g., smart door locks, motion sensors for security) can be used to infer interruptibility related to domestic routines. Instead of a normal camera, a thermal camera could be equipped with a smart speaker 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. The thermal images from the thermal camera could also benefit to detect such activities [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. Dormitory environments preclude several in-home activities such as cooking and restrict the range of inhabitants' movements due to their small size. Family members, compared to roommates, might have a more shared context for a particular notification as they likely share a more intimate bond. However, 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, and allow us to collect a rich dataset in single-rooms conveniently.The living room (45%) was the most common location for smart speakers [51] where social interactions between family members occur and may affect interruptibility of each other (e.g., watching TV together, having a conversation, eating together, one is sleeping while the other is watching TV). These kinds of social interactions were observed in dorms as well (e.g, watching a video together on a phone/computer, having a conversation, sharing a meal, one sleeping while the other is watching a video). 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 residency, and friends sharing an apartment, could further generalize our findings.

Another limitation is that we did not consider real conversational tasks to judge interaction timing, but we asked users to envision hypothetical productivity scenarios (e.g., restocking detergents or changing schedules). Prior research frequently used similar approaches to judge users' interruptiblity [24, 47] to minimize personal bias in usefulness by contents. 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 [26], or when a user accepts or misses a phone call [52]. As alluded in a recent work on multi-stage receptivity model [12], actual engagement with proactive services (availability vs. adherence) requires further investigation by conducting user studies by designing realistic and contextualized proactive services; e.g., information delivery, entertainment, decision-making, and IoT home device control tasks.

7 CONCLUSION

A field study with forty participants for one week 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 the everyday routines at homes such as habituated activities, resource conflicts for dual-tasking, user mobility types (e.g., entrance/departure), and interpersonal conflicts. We discussed 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.

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 Article 33 (Sept. 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 (CHI '04)*. ACM, New York, NY, USA, 271–278.

[3] 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 (CHI '19)*. ACM, New York, NY, USA, Article 3, 13 pages.

[4] 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 (CSCW '19)*. ACM, New York, NY, USA, 142–145.

[5] 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 (ITS '10)*. ACM, New York, NY, USA, 121–130.

[6] 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 (CHI '19)*. ACM, New York, NY, USA, Article 243, 13 pages.

[7] 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.

[8] 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 (MobileHCI '15)*. ACM, New York, NY, USA, 1070–1075.

[9] 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

[10] 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 (UbiComp '07)*. Springer-Verlag, Berlin, Heidelberg, 127–144.

[11] 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.

[12] 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.

[13] cnet. 2019. *What Amazon's Alexa will tell us in 2019.* https://www.cnet.com/news/what-amazon-alexa-will-tell-us-in-2019/

[14] Andy Crabtree and Tom Rodden. 2004. Domestic routines and design for the home. *Computer Supported Cooperative Work (CSCW)* 13, 2 (2004), 191–220.

[15] 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 (UbiComp'06)*. Springer-Verlag, Berlin, Heidelberg, 19–34.

[16] 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 (CHI '05)*. ACM, New York, NY, USA, 331–340.

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

[18] 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 (MobileHCI '16)*. ACM, New York, NY, USA, 195–205.

[19] 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 (CHI '05)*. ACM, New York, NY, USA, 909–918.

[20] 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 (UbiComp '15)*. ACM, New York, NY, USA, 493–504.

[21] 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 (CHI EA '05)*. ACM, New York, NY, USA, 1489–1492.

[22] 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 (CHI '08)*. ACM, New York, NY, USA, 93–102.

[23] 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.

[24] 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.

[25] 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 Article 14 (March 2020), 22 pages. https://doi.org/10.1145/3381009

[26] 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 (CHI '15)*. ACM, New York, NY, USA, 487–496.

[27] 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 (DIS '18)*. ACM, New York, NY, USA, 881–894.

[28] 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 (UIST '18)*. ACM, New York, NY, USA, 213–224.

[29] 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 Article 102 (Nov. 2018), 31 pages. https://doi.org/10.1145/3274371

[30] 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 (CHI '19)*. ACM, New York, NY, USA, Article 328, 10 pages.

[31] 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 (CHIIR '18)*. Association for Computing Machinery, New York, NY, USA, 265–268. https://doi.org/10.1145/3176349.3176868

[32] 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 (CSCW '15)*. ACM, New York, NY, USA, 903–916.

[33] 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 (UbiComp '15)*. ACM, New York, NY, USA, 813–824.

[34] 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 (UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 813–824. https://doi.org/10.1145/2750858.2807544

[35] 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 (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1021–1032. https://doi.org/10.1145/2858036.2858566

[36] 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 (UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 105–115. https://doi.org/10.1145/2632048.2636076

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

[38] 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 (CSCW '17)*. ACM, New York, NY, USA, 541–554.

[39] 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 (UbiComp '14)*. ACM, New York, NY, USA, 897–908.

[40] 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.

[41] 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 (CHI '18)*. ACM, New York, NY, USA, Article 640, 12 pages.

[42] 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 (CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 2853–2859. https://doi.org/10.1145/3027063.3053246

[43] 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.

[44] 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 (CHI '03)*. ACM, New York, NY, USA, 9–16.

[45] 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 (CHI '09)*. ACM, New York, NY, USA, 1819–1828.

[46] 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 (DIS '18)*. ACM, New York, NY, USA, 857–868.

[47] 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 (CHI '19)*.

ACM, New York, NY, USA, Article 637, 12 pages.

[48] 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 *ISWC*, Vol. 3. 248.

[49] 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 (UbiComp'05)*. Springer-Verlag, Berlin, Heidelberg, 232–250.

[50] 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 (SOUPS'19)*. USENIX Association, USA, 435–450.

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

[52] 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 (MobileHCI '07)*. ACM, New York, NY, USA, 178–181.

[53] 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 (CHI '02)*. ACM, New York, NY, USA, 399–406.

[54] 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 (UbiComp '15)*. ACM, New York, NY, USA, 801–812.

[55] 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. https://hal.archives-ouvertes.fr/hal-00953006

[56] Niels van Berkel, Jorge Goncalves, Simo Hosio, Zhanna Sarsenbayeva, Eduardo Velloso, and Vassilis Kostakos. 2019. Overcoming compliance bias in self-report studies: A cross-study analysis. *International Journal of Human-Computer Studies* (2019).

[57] 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/

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

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

[60] 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 (CHI '18)*. ACM, New York, NY, USA, Article 88, 13 pages.

[61] 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 (UbiComp '15)*. ACM, New York, NY, USA, 779–790.

[62] 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 (CHI '17)*. ACM, New York, NY, USA, 5346–5360.