Experimental Study on Estimation of Opportune Moments for Proactive Voice Information Service Based on Activity Transition for People Living Alone

Mitsuki Komori[1], Yuichiro Fujimoto[1], Jianfeng Xu[2], Kazuyuki Tasaka[2], Hiromasa Yanagihara[2], and Kinya Fujita[1(✉)]

[1] Graduate School, Tokyo University of Agriculture and Technology, 2-24-16 Nakacho, Koganei, Tokyo 184-8588, Japan
Sl766l5w@st.go.tuat.ac.jp,
{y_fuji,kfujita}@cc.tuat.ac.jp
[2] KDDI Research Inc., 2-1-15 Ohara, Fujimino, Saitama 356-0003, Japan
{ji-xu,ka-tasaka,yanap}@kddi-research.jp

Abstract. Smart speakers that listen to a user's commands and respond vocally are being used in homes across the world. Making smart speakers proactive by delivering information without an explicit command from the user might extend their applications and benefit users. However, such improvements also pose a risk for disturbing users. Therefore, this study aims at developing technology for estimating the opportune moments for information delivery without disturbing the user's daily activity. To analyze the subjective acceptability of users at home, we prototyped an experimental system that detects the activity transitions of participants based on his/her location and body motion using a depth camera and vocally asks his/her acceptability for information delivery at that moment. We conducted an experiment with three participants that lived alone. The results suggested that the acceptability of users relates to the activity patterns both before and after activity transition.

Keywords: Smart speaker · Acceptability · Proactive service · Depth camera · Notification

1 Introduction

Smart speakers that communicate with users in a conversational voice, such as Google Home [1] and Amazon Echo [2], are gaining attention and are being used in to homes around the world. Current smart speakers are manufactured to focus on passive information service. In other words, they wait for the users' vocal request and provide the necessary information. Although they are currently used for passive information delivery, proactive delivery of certain information may benefit users. For instance, a forecast on rainfall just before leaving or a reminder of immediate schedules will be

M. Kurosu (Ed.): HCII 2019, LNCS 11566, pp. 527–539, 2019.
https://doi.org/10.1007/978-3-030-22646-6_39

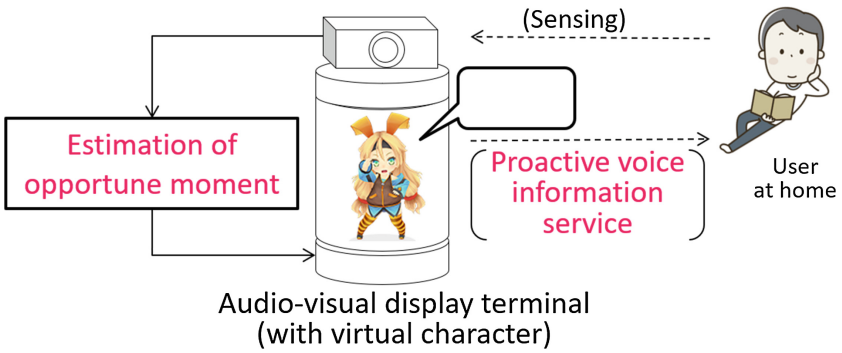


Fig. 1. Concept of proactive voice information service.

appreciated even if they are delivered without an explicit request. As represented by these examples, the proactive service, which satisfies the user's implicit needs, has the potential to make daily-life better.

Another promising scenario of the proactive information service is advertisement delivery. Delivery of an advertisement, which might attract the consumer's interest, may be positively accepted and facilitate a willingness for purchase. However, even if the delivered advertisement is one potentially of interest, a delivery at an inadequate time will be declined by the users and may lead the users to turn off the proactive information delivery function. For avoiding such situations, proactive information delivery needs to be conducted at the right time. Therefore, this study aims to develop a method for estimating the opportune moments for proactive voice information service by sensing the status of the user. Figure 1 represents the concept of this study.

To estimate the chance for information delivery, the acceptability of users for information delivered at unexpected times needs to be investigated first. It has been reported that the cognitive workload of office workers decreases at the breakpoint between tasks [3, 4]. Similarly, the intervals between activities may also be used as information delivery at home. Therefore, we investigated subjectively acceptable moments through a questionnaire targeting housewives, who are one of the conceivable target groups of proactive information delivery at home. We also prototyped an experimental system that estimates the presumable activity boundary based on user motion detected by a depth camera. Then, we conducted a one-week at-home experiment with three participants living alone. The results suggested that the activity pattern before and after activity transitions has some relation with the acceptability of proactive information delivery.

2 Related Work

In conjunction with the worldwide spread of smart speakers, the number of studies on the application of smart speakers is also rapidly increasing [5, 6]. Several studies have tackled the estimation of the opportune moment for information delivery at home. Takemae et al. experimentally investigated the user's preference of each room for notifications [7]. Cumin et al. constructed a simulated house environment and conducted a user study there. They also reported that the location in a room considerably affects acceptability rather than the type of activity [8]. Vastenburg et al. demonstrated that the degree of engagement in the activity, in addition to the urgency of the information, influences the acceptability through at-home experiments [9].

Cognitive workload is known to decrease at the breakpoint between tasks [3, 4]. Tanaka et al. focused on this and proposed an interruptibility estimation algorithm for an office environment [10]. Banerjee et al. also focused on the task breakpoint in a manipulation task. They detected the boundary of motion using a Kinect to estimate the suitable time for voice interruption by a robot [11]. Therefore, this study assumes that users are more acceptable to information delivery at the boundary of activities in a home environment.

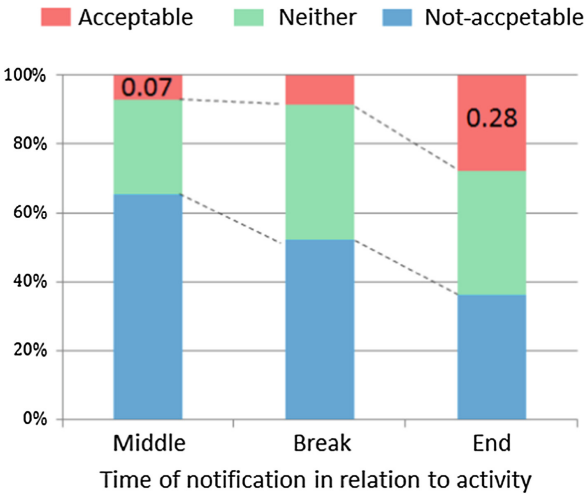
3 Questionnaire Survey and Hypothesis

3.1 Questionnaire on Acceptability of Information Delivery at Home

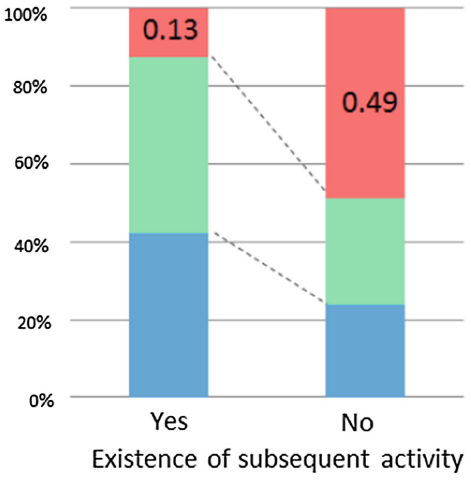
To investigate the variation of subjective notification acceptability at various timings of an activity, we conducted a questionnaire survey with 16 housewives [12]. We requested them to score the acceptability of 61 timings in three levels; acceptable, neither, or not acceptable. We asked them to imagine a proactive smart speaker delivering an advertisement. A few examples of timings are

- Just after cleaning a room.
- While looking at a variety of TV programs after dinner.

Figure 2(a) represents the summaries of answers in the middle, at the break, and the end of an activity. The rate of acceptance in the middle and end of an activity was 7% and 28%, respectively. Although the acceptance rate at the end of an activity is higher, the participants answered more than two-thirds of the cases as not acceptable. Then, we further divided the cases at the end of an activity into two groups based on the existence of a subsequent activity. As a result, the acceptance rate in the cases without subsequent tasks was 49% while only 15% of questions were answered as acceptable for cases with a subsequent to-do task, as shown in Fig. 2(b). These results suggest that the absence of a subsequent to-do task, i.e., the time allowed for rest, is the key to the acceptance of information delivery.



(a)



(b)

Fig. 2. Results of questionnaire for notification acceptability. (a) Distribution of acceptability scores for various times in relation to the activity being performed. (b) Detailed analysis of the answers for the times at the end of activities. Aggregated for each case with and without subsequent activity.

3.2 Model of Activity Transition and Opportune Moments for Information Delivery

In reference to the results of the questionnaire, this study defines the situation when no to-do task is at hand and the user is thus allowed to rest as a "mental goal" at home; we discuss a model to detect timings for information delivery. Here, we consider a state transition model as shown in Fig. 3.

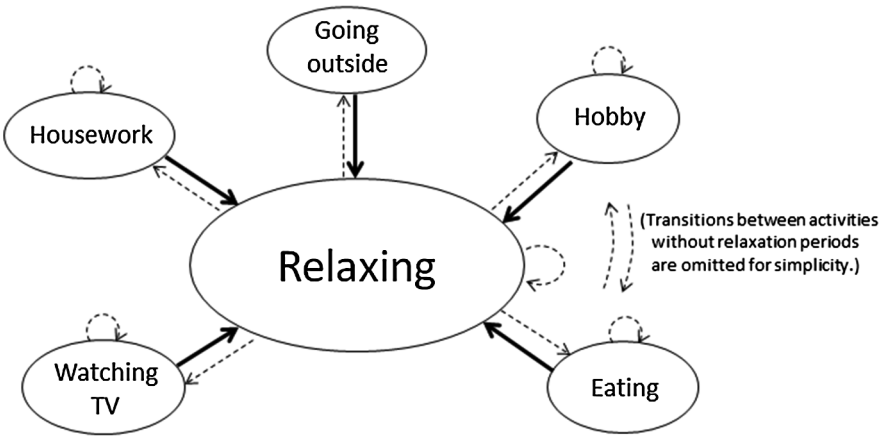


Fig. 3. Model of activity transition and opportune moments for information delivery. Bold solid arrows indicate better transitions for information delivery

The transitions to "mental goals," i.e., transitions to a relaxation state are represented as the bold solid arrows. This study discusses the feasibility to estimate chances for information delivery by detecting the transition of activities and judging whether the current state is a relaxation state. The transition in activities is expected to be related to the motion of the user. In particular, moving to another location in the room may be a sign of activity transition. For instance, we sometimes move to a desk to use a PC and sometimes move to a table to eat. The motion of arms can reflect activity transition. As for the recognition of the relaxation state, the posture and the location may provide some input. For instance, in a statistical sense, users that are standing or walking are thought to be engaged in an activity or busy, whereas users that are sitting or lying down may be considered to be free or in a relaxed state. Users lying on a bed are also presumed to be more relaxed.

4 Experimental System

4.1 System Overview

We developed an experimental system for collecting users' subjective acceptability scores together with their behavioral data at various timings at their home. Figure 4 illustrates the processing flow of the system. The system consists of a Kinect v2 and a laptop PC (OS: Windows10, CPU: Core i7 - 2.0 GHz, RAM: 8 GB). The system continuously monitors the user's position and motion based on the body tracking function of Kinect for Windows SDK 2.0 [13]. If any specific movements, which are presumed to relate with activity transition, are detected, it triggers the notification judgment process. Next, if the notification condition is fulfilled, the system vocally asks the participant "Do you have a minute?". Although we plan to estimate and reflect the type of activity on notification judgment, we only used the user's location in the room because the user's location is naturally related with the activity that he/she performs. We requested the participants to answer his/her acceptability at each notification time using finger gestures, and the system recorded it through an RGB image. The system also recorded the estimated body part locations twice per second to analyze the relationship between the user's motion and acceptability.

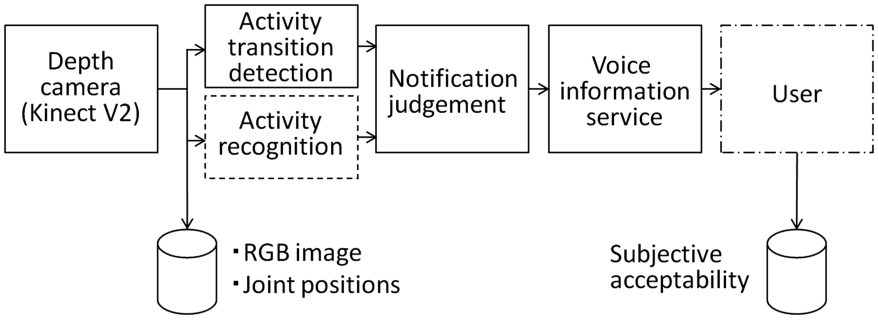


Fig. 4. Processing flow of experimental system.

4.2 Notification Judgement

As discussed later in Sect. 4.3, we focused on the position transfer and reaching action of users as index behaviors for activity transitions, which was verified in our previous study [14]. To avoid excessively frequent notifications, we set minimum blocking periods, which guarantees the participants will not be notified even if the notification condition is satisfied, to 8 min. Contrarily, in the case that no activity transition has been detected for 30 min, the system also provided a notification to the user for a comparison.

4.3 Activity Transition Detection

Most people will have several places in a room where they spend more hours, such as a desk, table, or sofa. We consider a situation when a person has just finished an activity and tries to start another one. If the person does not have a tool required for the activity such as a smartphone, PC, or book at hand, he or she will move to the place where the tool is and may come back to the original place. Alternatively, if the person tries to start an action that needs to be done at a specific place, he or she will also move, for instance, move to the table for lunch. Therefore, we used the position transfer in a room as an index for activity transition. The prototyped system detects the moment when the target user settles in a place after moving from another place farther than 0.6 m. We used the 3D position of the waist provided by Kinect v2 as the position of the user. The system detected the settlement of the user if the total travel distance in 5 s was less than 0.2 m.

When starting a new activity, we often extend our arms for taking something. We also extend our arms at the end of the activity to put the used object on the table or other places. Thus, the system used this reaching action as another indicator of activity transition. In particular, the system detected a sequence of hand motions, where the arm is extended to some extent and then returns close to the body.

5 Experiment for Analyzing the Relationship Between User's Acceptability to Voice Information Delivery and Activity Pattern Before and After Transition

5.1 Overview of Experiment

We conducted an experiment to collect users' acceptability scores at various timings at home using the prototyped experimental system as described in Sect. 4. We recruited three male volunteers who were studying at graduate school for the ease of experiment. Each of them was living alone in an apartment with a single living room. The experiment was conducted after an ethical review by a committee at the university.

Figure 5 illustrates an example of the experimental environment. Kinect v2 was installed at 1.6 m from the floor with 10° of depression angle to capture the entire room. On weekdays, the system recorded the data for 4 to 5 h from the time the user came home until the time the user went to bed. On holidays, it recorded the data for 6 to 7 h from the time the user awoke until the evening. We collected the data of 118 h for 21 days in total (a week for each participant).

During operation, the system intermittently requested vocally the participants to answer their acceptability at the timings as explained in the last section. We instructed the participants to imagine that the notification is the delivery of an advertisement, which might attract the interest of the participant. We requested the participants to score their subjective acceptability in five levels (1: low, 5: high) and indicate the score using finger signs to the camera when the notification is heard.

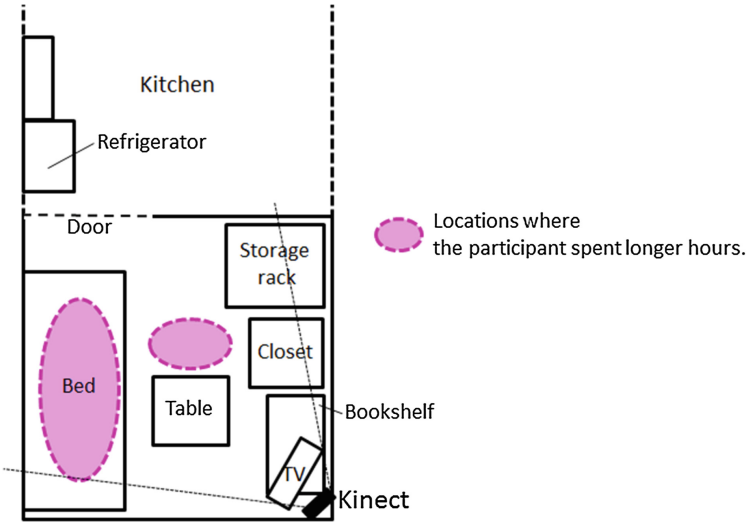


Fig. 5. An example of the experimental environment. Magenta-colored regions indicate the locations where the participant spent longer hours

We also instructed the participants to live as usual except for answering the acceptability as much as possible. In consideration of their privacy, we allowed them to temporally turn off the system when they do not want the data recorded. Furthermore, after the experiment, they checked the recorded RGB images to find and delete the ones that they do not want to share.

5.2 Results

We analyzed 407 responses collected in the experiment and compressed the acceptability scores of 1 and 2 to "Not acceptable," 3 to "Neither," and 4 and 5 to "Acceptable." The numbers of the answers for not acceptable, neither, and acceptable were 164, 39, and 204, respectively. As we expected, the acceptability for voice notifications in real-life situations varied depending on the timing of delivery.

Then, we checked whether the acceptability at the end of the activity was higher than at the middle of the activity, with an expectation that was obtained from the questionnaire survey described in Sect. 3. We categorized the notifications into stationary (non-moving) and non-stationary (moving) groups. The non-stationary group represents the notifications that were delivered within 30 s after the detection of the position transfer of the participant, which implies a transition in activity.

Figure 6 shows the rates of the answers. The rate of acceptance at stationary situations was 49% while the rate at non-stationary scenes was 54%. Contrary to our expectations, their difference was apparently negligible. However, because they should originally have different natures, we further analyzed them.

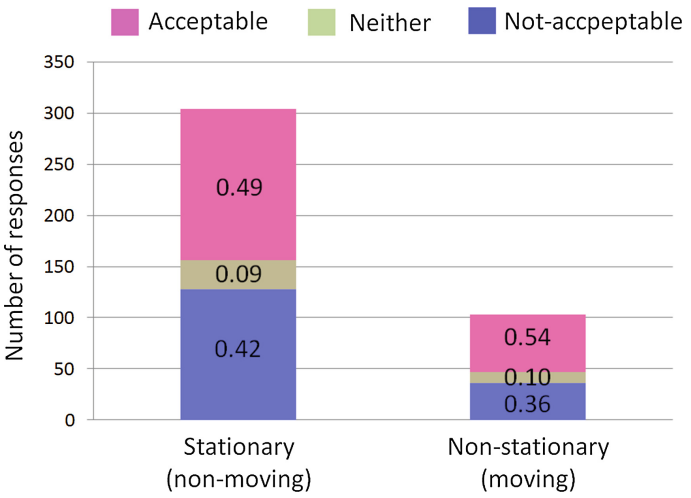


Fig. 6. Distribution of acceptability scores for notifications provided while in stationary (non-moving) and non-stationary (moving) situations.

At first, we focused on the 103 responses answered just after moving, which are summarized in the right bar in Fig. 6. Because our activity in the living room loosely relates with the location, we speculated that the acceptability has some relationship with the settled location after the transfer. Thus, we divided the responses for each location after moving as shown in Fig. 7.

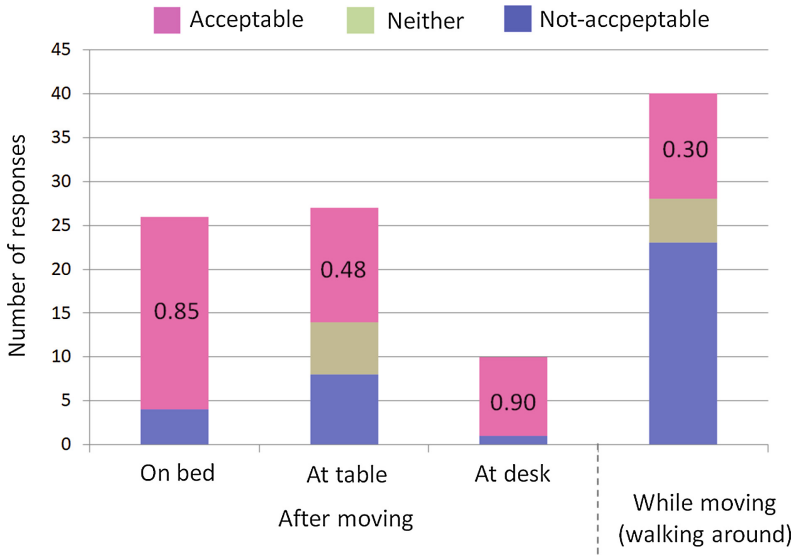


Fig. 7. Distribution of acceptability scores for notifications while in non-stationary (while or after moving) situations. Aggregated for each location being settled after moving.

The average rate of acceptance after moving was 69%, while the rate of moving was 30%. The situation while moving includes the cases that the participants were doing housework (e.g., cleaning with a vacuum cleaner and washing) and cases where they went to the restroom. The lower acceptance rate confirms that the timing while moving is not appropriate for information delivery; and the times staying at a specific location after moving is better.

As for the relationship between the settled place after moving and the acceptability, the rate of acceptance while on the bed was higher (85%) than that at the table (48%). The post-survey revealed that most of the activities on the bed were for fun (e.g., gaming and video streaming on a smartphone), while the measurable portion of activities at the table were for life or work, such as having lunch or using the PC. These activities for life and work appeared to be the major causes for the lower acceptance rate at the table. Contrary to our expectations, the rate of acceptance at the desk was high (90%). However, in this study, only one participant had a desk in his home, and there were only 10 samples. Therefore, further experimentation is needed for obtaining a general result on this point.

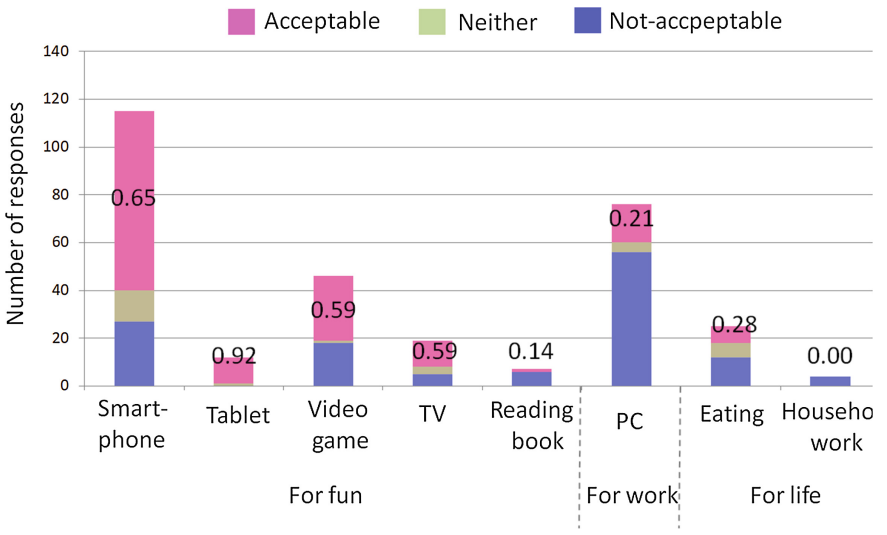


Fig. 8. Distribution of acceptability scores for notifications while in stationary (non-moving) situations. Aggregated for each activity being performed at the notification.

To further analyze the 304 responses answered at stationary (non-moving) situations, as shown in Fig. 8, we divided all observed activities into three types; activities for fun, work, and life. We refer to the use of a smartphone, tablet, video game, TV, and reading a book as a fun activity. We refer to the use of a PC as a work activity since each participant worked on their research activity on their PC. Finally, we refer eating and housework such as cleaning and washing as a life activity. The number of life activities were the smallest among the three types (29). The rate of acceptance while

doing housework was 0% (0/4), and the rate while eating was only 22% (7/22). These results confirmed that the notification during a life activity is inappropriate in terms of acceptability. The rate of acceptance while during a work activity (i.e., PC) was also only 16/76 (21%). It suggested that the notification during work activities is inappropriate as well. On the other hand, the rate of acceptance during a fun activity was higher than 60%, and all the activities except reading a book have high acceptability (smartphone: 75/115, tablet: 11/12, video game: 27/46, TV: 11/19, and reading book: 1/7). It suggested that the times while being engaged in a fun activity are the potential opportune moments for proactive vocal information delivery.

6 Discussion

The result suggested that when the user settles at a place after moving from another position is the chance for information delivery. Furthermore, in the cases that users were settled on a bed, which would be one of the typical places where a user relaxes, the voice notifications appeared to be more accepted. The result also suggested that while users are engaging in a fun activity, such as using a smartphone, is a chance for notification. Furthermore, some of the feedback provided through the post-survey suggested that users tend to accept notifications at higher probability during a fun activity after eating, taking a bath, and PC work. These example cases correspond to both the time at the activity transition and the time while relaxing; that is the "mental goal" discussed in Sect. 3.2.

However, even in those situations, 20% to 30% of the notifications were evaluated as not acceptable. In practical use, two or three times of inappropriate notifications out of 10 will be too much and unaffordable. Therefore, a further exploration for more appropriate timing is needed. In this study, we focused only on the activity transitions that can be judged based on externally-observable information (i.e., moving). Meanwhile, smartphones and PCs occupied most activities, at least in this experiment. Such multi-functional devices are used for various activities; and thus, the acceptability of the user might differ from the activities performed on the device. For instance, the post-survey revealed that the notifications while using application software for communication are not acceptable most of the time.

In contrast, acceptability for the notifications while playing a game depends on the concentration on the activity. Therefore, to identify the more appropriate times, we need to explore the effective information related to smartphone use such as application types for more accurate estimation. Furthermore, the use of smartphones as a media for information delivery is another promising option. Smartphones are also useful in detecting user activity. However, they are sometimes stored in a bag or left for charging. Consequently, it appears promising to coordinate smart speakers installed in a room and smartphones carried and occasionally used by a user.

In addition, the development of an algorithm for judging appropriate timings using the explored indices, in conjunction with the user position and body motion used in this study, are also needed. Although this study discussed the acceptability of proactive

vocal information delivery at various times in real-life scenarios, the small number of participants limits the discussion on the generality of the observed tendencies. A further experiment with more participants should be conducted to reveal the common rules at the opportune moments for proactive vocal information delivery.

7 Conclusion

In this study, based on the questionnaire survey, we prototyped an experimental system which detects the activity transition of a user and notifies the user vocally for a proactive information service. We also collected the subjective acceptability scores for the notification delivered at various timings in an actual living environment with three participants living alone over 21 days.

The results revealed that the notifications are more accepted when the user is settled on a bed after moving from another place and while the user is using a smartphone after eating, housework, or working on a PC. It suggested the necessity for detecting the user activity and its transition for estimating an opportune moment for information delivery. One promising direction is to utilize the information on smartphone use in combination with the sensor data provided by a smart speaker installed in a room. It is also necessary to conduct experiments with a greater number of participants for the further discussion on general tendency and individual differences.

References

1. Google Home. https://madeby.google.com/intl/en_us/home/. Accessed 1 Dec 2018
2. Amazon Echo. https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-with-WiFi-Alexa/dp/B00X4WHP5E. Accessed 1 Dec 2018
3. Mark, G.J., Gonzalez, V.M., Harris, J.: No task left behind? Examining the nature of fragmented work. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 321–330. ACM, New York (2005)
4. Iqbal, S.T., Bailey, B.P.: Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1489–1492. ACM, New York (2005)
5. Porcheron, M., Fischer, J.E., Reeves, S., Sharples, S.: Voice interface in everyday life. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, p. 640. ACM, New York (2018)
6. Luria, M., Hoffman, G., Zuckerman, O.: Comparing social robot, screen and voice interfaces for smart-home control. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 580–592. ACM, New York (2017)
7. Takemae, Y., Chaki, S., Ohno, T., Yoda, I., Ozawa, S.: Analysis of human interruptibility in the home environment. In: Extended Abstracts on Human Factors in Computing Systems, pp. 2681–2686. ACM, New York (2007)
8. Cumin, J., Lefebvre, G., Ramparany, F., Crowley, J.: Inferring availability for communication in smart homes using context. In: IEEE International Conference on Pervasive Computing and Communications (PerCom) Workshops (2018)

9. Vastenburg, M.H., Keyson, D.V., de Ridder, H.: Considerate home notification systems: a field study of acceptability of notifications in the home. Pers. Ubiquit. Comput. **12**(8), 555–566 (2008)
10. Tanaka, T., Fukazawa, S., Takeuchi, K., Nonaka, M., Fujita, K.: Study of uninterruptibility estimation method for office worker during PC work. J. Inf. Process. Soc. Jpn. **53**(1), 126–137 (2012)
11. Banerjee, S., Silva, A., Feigh, K., Chernova, S.: Effects of interruptibility-aware robot behavior. arXiv:1804.06383 (2018)
12. Fujimoto, Y., Komori, M., Xu, J., Tasaka, K., Yanagihara, H., Fujita, K.: Preliminary study on modeling opportune time for proactive auditory information service. In: Proceedings of the 80th National Convention of Information Processing Society of Japan, 5E-03. IPSJ, Tokyo (2018)
13. Shotton, J., et al.: Real-time human pose recognition in parts from single depth images. Commun. ACM **56**(1), 116–124 (2013)
14. Fujimoto, Y., Nagasawa, Y., Xu, J., Tasaka, K., Yanagihara, H., Fujita, K.: Possibility of estimation on information provision based on body movement toward push-type information service. In: Proceedings of Human Interface Symposium, 7D1-5. HIS, Kyoto (2017)