Regular Paper

Sonic Home: Environmental Sound Collection Game for Human Activity Recognition

Akira Masuda[1,a) Kun Zhang[1,b) Takuya Maekawa[1,c)

Received: May 4, 2015, Accepted: November 5, 2015

Abstract: This paper presents a new smartphone game for collecting environmental sounds using a participatory sensing approach. In our game, a player becomes an owner of a virtual house and furnishes the house with in-game items that are obtained by collecting environmental sounds. The collected sounds are used to train an environmental sound recognition model for daily activity recognition. However, such participatory sensing systems have several issues related to, for example, motivation of users and reliability of collected data. To cope with the issues, we propose and implement gamified functions for controlling the quality and diversity of collected data. We conducted an experiment to evaluate the developed functions and confirmed the effectiveness of our developed functions.

Keywords: activity recognition, environmental sound recognition, gamification, participatory sensing

1. Introduction

The recent proliferation of smartphones has triggered many studies on activity recognition studies that employ smartphone sensors. For example, many studies try to recognize such activities as walking and running by using smartphone accelerometers [4], [8]. Also, several studies recognize daily activities with microphones in smartphones [15] by recognizing environmental sounds such as vacuuming sounds and sounds of running water. Such sensor-based activity recognition is a fundamental technology of context-aware systems and lifelogging. However, many activity recognition studies rely on supervised machine learning techniques and require labeled training data, which are costly to prepare. In particular, preparing sound training data is very costly because there are many kinds of sounds in our daily lives. Also, features of sounds may highly depend on environments. (Sounds of vacuum cleaners, for example, depend on manufacturers or product types.) Therefore, we should collect an enormous amount of labeled sound data.

In order to collect large amounts of sensor data, recent studies have employed a participatory sensing approach [5], which enables large numbers of people (end users or participants) to share the burden of sensor data collection. However, this approach, which relies on end users, has several issues related to, for example, motivation of users and reliability of collected data. To cope with the issues, we employ *gamification* techniques and develop a social game named *Sonic Home* designed for collecting training sound data.

Gamification can be described as the use of game design elements in non-game contexts [10], [22]. For example, Foursquare has a gamified check-in system where users compete for *ownership* of spaces by the frequency that they visit them [14]. Game players (end users) are reportedly to have several kinds of emotions [1], [2], for example desire for attaining first place, and the gamification approaches permit us to motivate the end users to accomplish a certain task by stimulating the emotions.

Based on the gamification techniques, this study develops functionality of our game for solving the following problems related to the participatory sensing approach.

- **Motivation**: The success of a gamified participatory sensing system is how much sensor data is collected. However, when motivation of game play is low, end users stop collecting sensor data and the participatory sensing system will consequently fail. Therefore, based on prior studies on gamified systems, we develop the following two functions to keep the motivation of game play high; ranking and social functions. Note that these two functions have already been implemented or discussed in the prior studies and these functions are not new. We propose new functions for coping with the following two problems.
- **Diversity**: In a participatory sensing system, because end users collect sensor data as they like, easy-to-collect sensor data are collected many times. For example, sound clips of flushing toilet are easy-to-collect in our daily lives. So, the number of collected sound clips of flushing toilet will be large. However, collecting sound clips of an ambulance siren, for example, in our daily lives is difficult and so the number of collected sound clips of an ambulance siren will be small. In this study, we propose and implement a campaign function in order to control the diversity of collected sound clips. With the campaign function, we can reward an end user who collected hard-to-collect sound clips and thus

1 Graduate School of Information Science and Technology, Osaka University, Suita, Osaka 565–0871, Japan
a) masuda.akira@ist.osaka-u.ac.jp
b) kun.l.zhang@gmail.com
c) maekawa@ist.osaka-u.ac.jp

motivate end users to collect hard-to-collect sensor data.

- **Quality**: One drawback of a participatory sensing system is that the quality of data depends on participant enthusiasm to reliably collect sensing data [13]. Also, a game user can intentionally collect bad quality or meaningless sound clips in order to obtain rewards with a small or no effort. That is, the quality of collected data by end users is not guaranteed. However, it is very difficult for an administrator of the game to evaluate/verify the quality of the large amounts of collected sound data. In order to cope with this problem, we design our game so that an end user evaluates the quality of sound clips collected by other end users. That is, we implement a function that permits an end user to evaluate the quality of sound clips collected by other end users. When an end user evaluates a sound clip, the end user receives a reward. By doing so, we can obtain evaluated sound clips with the administrator's small effort. We propose a method to calculate the reliability of a sound clip by using its evaluations and then construct a reliable environmental sound recognition model.

The contributions of this work are that (1) we actually developed a smartphone-based sound collection game named Sonic Home that is equipped with functions for controlling the diversity and quality of collected sensor data, (2) we conducted an experiment to evaluate the developed functions, and (3) we proposed a method for training a reliable environmental sound recognition model with collected sound clips and evaluations by end users.

2. Related Work

2.1 Participatory Sensing

Due to the recent proliferation of smartphones, many participatory sensing systems have been developed. For example, systems proposed in Refs. [20], [21] provide fine-grained traffic information on a large scale. Environmental monitoring is also an important application of participatory sensing systems. For example, systems for monitoring air pollution [16] and urban noise [18] have been developed. Also, a system for collecting photos of the content of recycling bins has been studied [6]. Furthermore, a system for collecting training acceleration data for activity recognition has been developed [12].

2.2 Motivation

As above, while various participatory sensing applications have been studied, many of them do not have the design of incentive mechanisms. Because the top motivation for participation is reported to be enjoyment [17], recent several studies have employed gamified techniques to motivate users (participants). For example, BudBurst is a mobile game for collecting plant life stage data [11]. Players gain points and levels within the game by finding and making qualitative observations on plants. Moreover, in EyeSpy [3], players tag geographic locations with photos or text. Similar to our study, EyeSpy also has a game design where an end user verifies other users' tags. In contrast, our study tries to calculate a score of a collected sound clip with its evaluations and sound data analysis, and train a sound recognition model based on the score.

Game players (end users) are reportedly to have several kinds of emotions and several player categories are proposed according to their emotions as follows [1], [2].

- **Explorers**: Players who prefer discovering areas, creating maps, and learning about hidden places.
- **Achievers**: Players who prefer to gain points, levels, ranks, and so on within a game. Several gamification studies try to stimulate the emotions of the players. For example, in the above mentioned BudBurst, an end user can obtain ranks after collecting sufficient points ranging from "Sprout," with the fewest points, through "Seedling," "Thriving," and "Deep-Rooted" as the highest rank.
- **Socializers**: There are many players who play games for social aspects. They gain the most enjoyment from a game by interacting with other players.

2.3 Environmental Sound Recognition

While there are many studies on environmental sound recognition, most of them develop standalone applications just work locally on smartphones or PCs [7], [15]. For example, in Ref. [7], bathroom activities such as showering, flushing, and urination are recognized by using microphone data. On the other hand, a method proposed in Ref. [19] employs an efficient way to collect training data by downloading sound clips from an audio sharing web site. Because some uploaded sound clips might be noisy, outlier detection techniques are used to detect and remove such noisy sound clips. In contrast, we compute a score (reliability) of a sound clip by using both evaluations by users and sound data analysis.

3. Sonic Home

3.1 Basics of Playing Sonic Home

Sonic Home is our developed game for Android smartphone users. An end user (player) becomes an owner of a virtual house as shown in **Fig. 1**. In the beginning of the game play, there are no in-game items (furniture) in the house. The purpose of this game is to furnish the house with in-game furniture by his/her own. By selecting (tapping) a room, the user can see a detailed view of the room. Icons in the left upper portion of the screen (Fig. 1) show a photo of the user's face, amount of in-game money, etc. Icons in the lower portion of the screen are menu buttons. Basically, the user furnishes the house with various kinds of in-game items by repeating the following procedures; (1) recording and uploading

Fig. 1 Screenshot of Sonic Home.

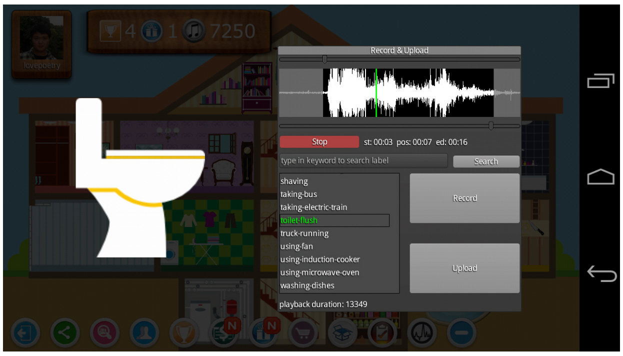

Fig. 2 Environmental sound collection view of Sonic Home.

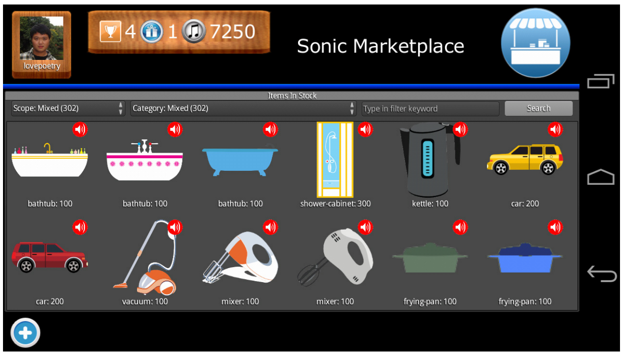

Fig. 3 Market view of Sonic Home.

a sound, (2) buying furniture in a market, and (3) furnishing the house with the furniture. We explain the procedures.

3.1.1 Recording and Uploading Sound

A user records environmental sound with a smartphone microphone. **Figure 2** shows a view for sound recording, which is launched from a menu button. With a waveform shown in the upper right portion of this view, the user can trim the sound by selecting anacoustic segments. Then, the user labels the sound by selecting a sound category and uploads an MP3 file converted from the sound by selecting an "upload" button. Finally, the user obtains in-game money in reward for the upload. Note that the amount of in-game money an end user obtains decreases as he/she uploads sound clips with the same label many times. Also, because our Android application discards the sound file just after the upload, the user cannot upload the same file again.

3.1.2 Buying Furniture

The user can go to a market from a menu button (**Fig. 3**). In the market, the user can buy furniture with in-game money. Note that the user can buy only furniture related to sound clips that the users have already uploaded. For example, after a user uploads a sound clip of toilet flushing, the user can buy furniture or items related to the sound clip such as a lavatory basin and a sink. The purchased items will be stored in a storage.

3.1.3 Install Furniture

The user can get an item out of his/her storage and place the item at a room as he/she likes. The user first selects an item in his/her storage and then taps a location in a room where he/she wants to place the item. The user can also change the size of an item as he/she likes.

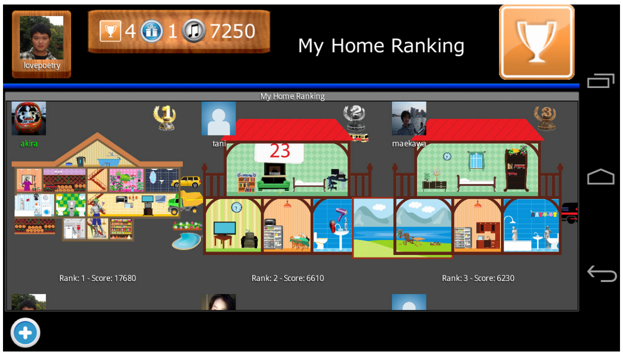

Fig. 4 Ranking view of Sonic Home.

3.2 Three Features of Sonic Home

As above, we have explained how to play Sonic Home. Our game also has several functions for solving the three issues mentioned in the introduction section as follows.

3.2.1 Motivation

As mentioned in the introduction section, a decrease in motivation is an important issue of participatory sensing systems. We try to motivate participants by employing ranks and social networks in a similar way to prior studies. As for social networks, our game has a function that enables a user to upload a screenshot of his/her house to existing social network services such as Twitter and Facebook. (For example, the user can attach a screenshot to his/her tweet.) With this function, the user can share a screenshot of his/her house with other people. This function is developed to stimulate the desire of the socializers described in Section 2.2. As for ranks, our game provides a ranking function as shown in **Fig. 4**. The function provides a list of houses of top-ranked users. The ranks of the users are computed according to the total price of furniture in their houses. Also, by tapping (selecting) a house in the ranking view, the user can visit the house and see the details of the house (like Fig. 1). That is, the top ranked houses will be visited by many other users. By providing the ranking function, we can motivate players, especially the achievers described in Section 2.2, to get high in the ranking. Also, the users are expected to feel motivated when they watch top-ranked fine houses.

3.2.2 Quality

As mentioned in the related work section, the quality of collected sensor data is one issue of participatory sensing systems. For example, a collected sound clip may include noises or the volume of a sound clip may be small. Also, an end user may upload bad quality or meaningless sound clips to obtain in-game money with no or small effort. When such kinds of poor sound clips are uploaded, a trained sound recognition model will also be poor. However, it is very costly for an administrator of the game to evaluate the quality of the large amounts of collected sound data.

In this study, we design our game so that an end user evaluates the quality of sound clips collected by other end users. That is, we implement a function that permits an end user to evaluate the quality of sound clips collected by other end users. When an end user evaluates a sound clip, the end user receives a reward. With this approach, we motivate an end user to evaluate sound clips uploaded by other users.

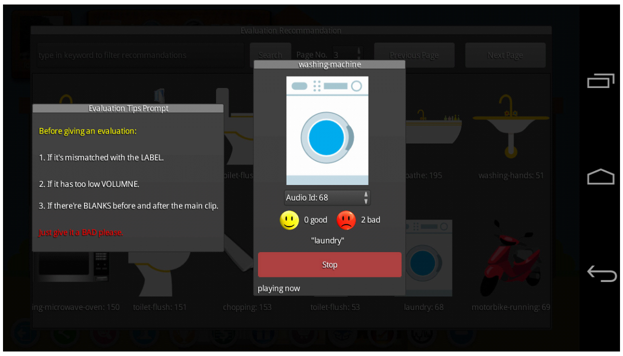

Fig. 5 Environmental sound evaluation view.

Here we explain how an end user evaluates a sound clip. As mentioned above, an end user can visit another user's house. When the user selects (taps) an item in the other user's house, the user can play sounds related to the item that were collected by the another user (e.g., sound of flushing toilet for lavatory basin) as shown in **Fig. 5**. With this view (Fig. 5), the user can play a sound clip and evaluate it by selecting a good or bad button. Criteria for the evaluation are indicated at the left portion of this screen (label correctness, sound volume, and presence of silent sound segments). By using the evaluations performed by end users, we estimate the quality of sound clips. We detail it in the next section.

3.2.3 Diversity

There are several hard-to-collect sounds in our daily lives. For example, collecting sound clips of an ambulance siren is difficult. We employ gamified techniques to encourage end users to collect such hard-to-collect sounds. Our game has a campaign function that enables an administrator of the game to perform a campaign for a hard-to-collect sound class. A campaign has its start and end dates and if an end user collects or evaluates a sound clip on which the campaign focuses during the campaign period, the end user can obtain more in-game money than usual. Also, when an end user completes the predefined number of sound clips on which the campaign focuses, the end user obtains a bonus (in-game money and a special item). That is, by performing a campaign of a hard-to-collect sound class, an administrator of this game motivates end users, especially the explorers and achievers described in Section 2.2, to collect the sound.

3.3 System Implementation

Because our game has multiple end users, we implemented our game system as a server-client system. We implemented the server program on Apache Tomcat by using Java. The server program manages uploaded sound clips, uploaded evaluations, etc. The server program also manages user data (e.g., positions of furniture, in-game money, and ranks). We implemented the client program as an Android application by using Java. The client program provides game play functionality as mentioned in Section 3.1.

4. Reliability Estimation Using Game Play Data

4.1 Overview

As mentioned in Section 3.2.2, an end user evaluates a sound clip collected by another end user. We estimate a score (reliability) of a sound clip by using the evaluations and sound data analysis. A score of a sound clip is the sum of the following two sub-scores, a data score and an evaluation score. The data score is an objective score that is computed by analyzing collected data. The evaluation score is a subjective score that is computed based on evaluations by end users. We explain them in detail.

4.2 Data Score

This score is computed based on our assumption that a sound clip that is apparently different in acoustic features from others with the same label will be considered unreliable. In other words, if a sound clip is similar to other sound clips with the same label, then it is reliable.

When we compute a data score of a sound clip, we construct a statistical model with other sound clips with the same label and then compare the sound clip with the model. Before that we extract feature vectors from sound clips. Each feature vector consists of Mel-Frequency Cepstral Coefficients (MFCC) components, which are usually used in environmental sound recognition, extracted from a sliding time window of a sound clip. Then we build a model of the other sound clips by using the Gaussian Mixture Model (GMM) [9]. To construct the model, we estimate the parameters of the model by employing expectation maximization (EM) [9]. We can compute the likelihood of the model (M) for a feature vector at time t (v_t) as follows.

$$p(v_t|M) = \sum_i \pi_i \mathcal{N}(v_t, \mu_i, \Sigma_i), \quad (1)$$

where π_i is the mixture weight of the ith multivariate Gaussian distribution of the GMM, and μ_i and Σ_i are the mean vector and covariance matrix of the Gaussian distribution, respectively. That is, we can compute the likelihood of the model for a feature vector extracted from the sound clip of interest by using Eq. (1). Because we extract multiple feature vectors from the sound clip of interest, we compute the average likelihood for the vectors and the average likelihood will be the data score of sound clip s as follows.

$$d(s) = 1/T \sum_{t=1}^{T} p(v_t|M) \quad (2)$$

where T is the number of feature vectors (time windows) extracted from s.

4.3 Evaluation Score

This score is computed based on our assumption that a sound clip that is evaluated as good by many end users will be considered reliable. With this score, we can distinguish a correctly labeled sound clip dissimilar to other sound clips with the same label from an incorrectly labeled sound clip both of which have small data scores. So, the trained sound recognition models do

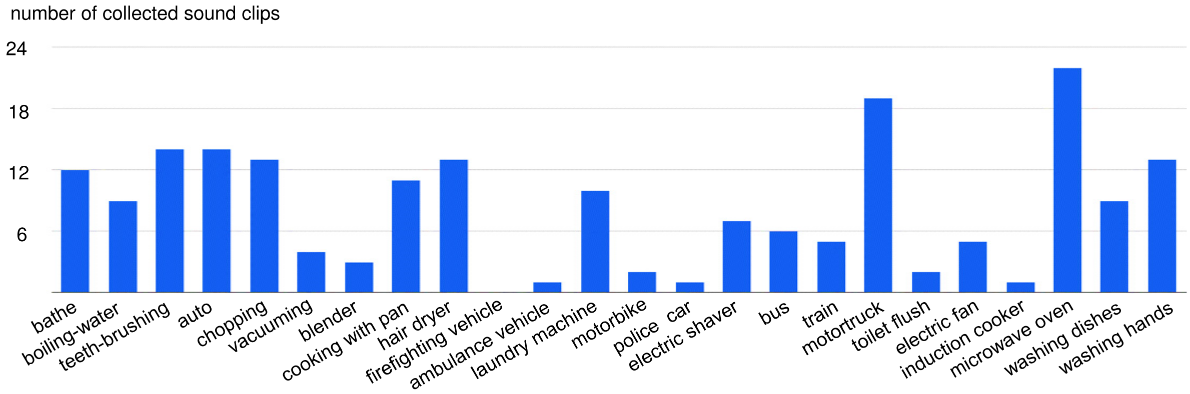

Fig. 6 Distribution of collected environmental sound clips.

not capture only similar sound patterns with high data scores but also various sound patterns with small data scores while reducing the weights of incorrectly labeled sound patterns. Note that, because the evaluations are performed by end users, the reliability of the evaluations is also not guaranteed. Therefore, we compute the evaluation score by taking into account the reliability of a user in terms of his/her evaluations.

We first define variables used in our calculation method. We assume that a set of "good" evaluations of sound clip s by all end users to be G_s, a set of "bad" evaluations to be B_s, and a set of sound clips that has been evaluated by user u to be S_u. We can compute a basic evaluation score of s as $\overline{e(s)} = \frac{|G_s|}{|G_s|+|B_s|}$. Based on them, we compute the reliability of user u in terms of his/her evaluations based on our assumption that, when the evaluations of user u do not coincide with those of other users, the user's evaluations are unreliable. The reliability of user u in terms of his/her evaluations is calculated by using the following formula.

$$ue(u) = \frac{1}{|S_u|} \sum_{s \in S_u} \exp\{-(e_u(s) - \overline{e(s)})^2\} \tag{3}$$

where $e_u(s)$ is 1 if user u evaluated sound s as good and $e_u(s)$ is 0 if user u evaluated sound s as bad. Based on the reliability of user's evaluations, we compute an evaluation score of sound clip s as follows.

$$e(s) = \frac{\sum_{u \in U} ue(u) \cdot e_u(s)}{|G_s| + |B_s|} \tag{4}$$

where U is a set of end users. Note that, when user u does not evaluate s, $e_u(s)$ becomes 0. The equation shows the weighted ratio of good evaluations by the reliability of user evaluation ($ue(u)$).

4.4 Training Sound Recognition Model

Based on a data score ($d(s)$) and an evaluation score ($e(s)$) of sound clip s, we compute a score (reliability) of s by using $d(s) + c \times e(s)$, where c is used to adjust the weight of a subjective score.

We construct a sound recognition model by using the above mentioned GMM(M in Eq. (1)). We compute a model of the ith environmental sound class M_i by using the EM algorithm mentioned in Section 4.2. Note that, when we construct the model, we weight feature vectors extracted from sound clip s according to a score of s in order to make the probability density of high scored instances higher and low scored instances lower, which is expected to be able to improve the quality of training dataset for building recognition models.

After training a model for each sound class, we determine which class feature vector v is classified into. We compute the likelihood of each model $p(v|M_i)$ for feature vector v by using Eq. (1), and a sound class corresponding to a model with the highest likelihood is the classified sound class.

5. Evaluation

In this evaluation, we confirm the effectiveness of functions implemented in our game system and the effectiveness of our reliability estimation method. To evaluate the functions, we performed a trial where participants actually played our game. By using uploaded sound clips and evaluations by the participants, we also evaluate our reliability estimation method. After the trial, we asked the participants to answer questionnaires to obtain feedback of the trial.

5.1 Trial of Sonic Home

The trial of Sonic Home was performed during 25 days from January 9 to February 2, 2014. We asked 14 users (our university students) to register the game and 11 users were active players who actually uploaded sound clips and evaluated clips to furnish their houses. During the experiment, the participants have collected 195 sound clips of 23 sound classes and performed 658 evaluations. Before the experiment, we have prepared 24 sound classes shown in **Fig. 6**. Figure 6 also shows the number of collected sound clips for each sound class. As shown in Fig. 6, the participants could not record "firefighting vehicle" sounds. We can see that the numbers of sound clips of "toilet flush" and "microwave oven" were large because they are easy to record. However, sounds of vehicles such as "firefighting vehicle"and "motortruck" seem to be very hard to collect. In addition, the number of items purchased was 296, which means each user's house has 27 items on average.

Figure 7 shows an example house furnished by our experimental participant. She furnished her house with girlie items. The "beauty sense" of users is, in fact, an important part of the game play to appeal personal preferences to other users. In this case, the participant has decorated her house carefully with pink and lovely items. Like this example, many of our participants enjoyed

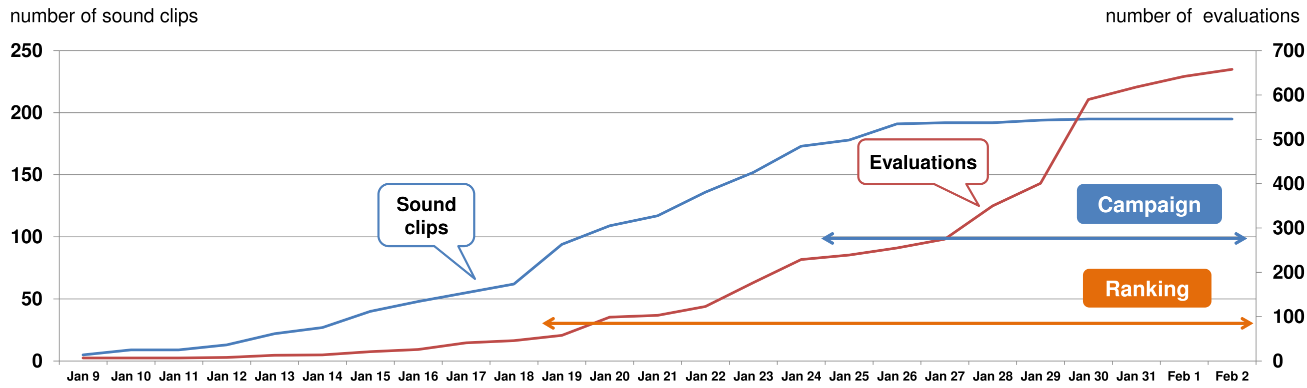

Fig. 8 Transitions of the cumulative numbers of uploaded sound clips and evaluations.

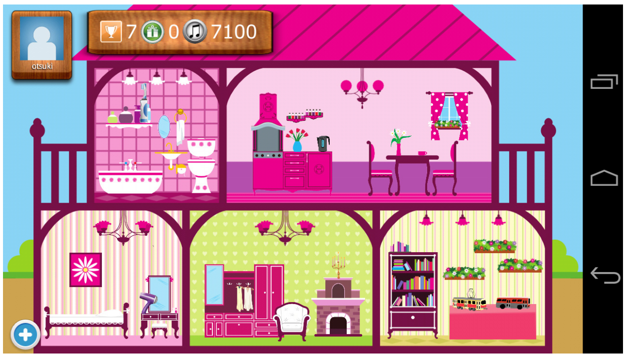

Fig. 7 Screenshot of a participant's house.

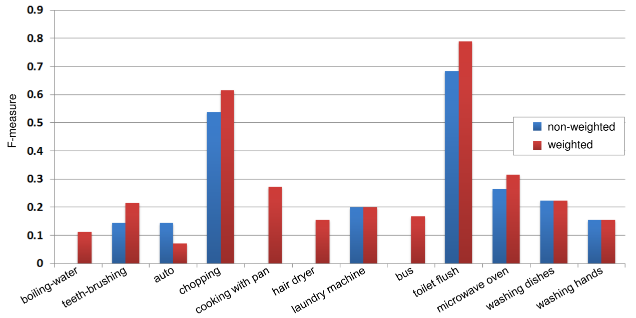

Fig. 9 Classification results when we train a recognition model on non-weighted data or weighted data.

furnishing their houses by using their inventiveness.

5.2 Effectiveness of Ranking and Campaign Functions

We first investigate the effect of the ranking function designed to cope with the motivation issue mentioned in Section 1. During the period of January 19 to February 2, we activated the ranking function. Therefore, we compare the number of collected sound clips (or evaluations) during this period with that during which the ranking function was not activated. **Figure 8** shows the transitions of the cumulative numbers of collected sound clips and evaluations. As shown in the figure, after we activated the ranking function, the numbers of collected sound clips and evaluations drastically increased. In fact, the number of sounds upload per day increased 1.4 times after we activated the ranking function. Also, the number of evaluations per day increased 6.9 times.

In Fig. 8, we can see that the number of sound uploads did not increase after January 26. This may be because houses of some participants were full of in-game items. Also, as mentioned in Section 3.1, the amount of in-game money an end user obtains decreases as he/she uploads sound clips with the same label many times. Therefore, the motivation of the participants for uploading decreased.

We then investigate the effect of the campaign function designed to cope with the motivation and diversity issues mentioned in Section 1. During the period of January 25 to February 2, we activated the campaign function. Note that we performed campaigns related to sound classes with small numbers of uploaded sound clips (six classes; "vacuuming," "motorbike," "hair dryer," "train," "bus," and "blender"). We compare the number of collected sound clips (or evaluations) related to the six classes during this period with that during which the campaign function was not activated. The number of sound uploads per day for the six classes was only 0.89 times compared with that before the campaigns. Note that the number of sound uploads per day for all the classes was 0.23 times compared with that before the campaigns. This means that, while the overall upload number decreased, the campaigns could keep the number of sound uploads for the six classes high. Also, the number of evaluations per day related to the six classes increased 12.5 times compared with that before the campaigns. In contrast, the number of evaluations per day for all the classes was 3.33 times compared with that before the campaigns.

5.3 Effectiveness of Reliability Estimation

In order to cope with the quality issue mentioned in Section 1, we calculate a score (reliability) of each sound clip collected by users and weight feature vectors extracted from sound clips according to their scores. Thus, for the evaluation of the reliability estimation, we compute the environmental sound recognition accuracy when weighted training data are used or non-weighted data are used. To compute the recognition accuracy, we employ leave-one-out cross validation. That is, we test one sound clip by using a classification model trained on other clips. Note that we evaluate only sound classes that our participants could collect sufficient quantities of sound clips. We also did not use the "bathe" sounds in our evaluation because our participants collected the sound by placing their smartphone outside the bathrooms and thus the volume of the collected sounds was small. **Figure 9** shows the classification results (F-measure) for the non-weighted data and the weighted data. (Note that, in general, the environ-

Table 1 Questions and answers for questionnaire survey of trial.

Question	Answer
Do you know any other applications that collect environmental sounds except for Sonic Home?	Yes: 0, No: 9
Have you been conscious of environmental sounds as a data source that can recognize human's activity before?	Not at all: 2, Not quite: 7
Do you think you have enjoyed yourself in "Sonic Home" when collecting environmental sounds?	Yes: 2, Generally Yes: 4 No: 1, I'm not sure: 2
How long did you take to play the game once?	More than 5 min and less than 15 min: 7 Less than or equal to 5 min: 2
Which part of this game makes you feel enjoyable?	Watching others' homes: 1 Decorating home: 7 Listening various sounds: 1
What is the motivation for you to make game money?	To get items for decoration: 3 To rank high: 4 To show beautiful home to the others: 2
Have you been checking the "Ranking"?	Yes: 8, No: 1
Have you become willing to collect more items to make your home better than others after you checked the "Ranking"?	Yes: 3, A little: 3, No: 3
Have you been checking the "Campaign"?	Yes: 3, No: 7
Did you become more aware of the sound targeted in "Campaign"?	Yes: 3, A little: 3, No: 3
Do you feel that your privacy has invaded when environmental sounds of your daily life may be heard by other people?	A little: 2, No: 5, Not at all: 2

mental sound recognition accuracy has not yet been high [19].) While F-measures of many sound classes for the non-weighted data were zero, the weighted data could improve the F-measures. The average F-measure for the non-weighted data was 18.42% and that was 24.21% for the weighted data. That is, with our reliability estimation method, we could achieve 31.43% improvement in terms of recognition accuracy.

5.4 Feedback of Trial

To understand how the participants of the trial feel about playing Sonic Home, we performed a web-based questionnaire survey that consists of 34 questions. Invitation e-mails to the survey were sent to 14 participants and we received 9 submissions. **Table 1** shows some of the questions in the survey and the responses from the participants.

From the feedback, we found that all of the respondents have never played a game that encourages players to collect environmental sounds. Also, the feedback show that 67% of the respondents fairly or generally enjoyed playing Sonic Home. Almost all of them especially enjoyed decorating their homes rather than collecting or evaluating environmental sounds. 44% of the respondents were motivated to make game money to rank high in the ranking function and 33% of the respondents have become willing to collect more items after they checked a ranking. These facts indicate that our attractive game system motivated the participants to collect and evaluate environmental sounds through game play. Therefore, we believe that our game system has contributed to cope with the motivation issue mentioned in Section 1.

Because we found that more than half of the respondents did not check the campaign function frequently, a campaign underway should be heightened while playing Sonic Home. While privacy seems not to be an urgent issue in Sonic Home, a few respondents cared about privacy, indicating that there is room for further improvement. For example, a function that enables players to define the publicity levels of their collected sounds can be useful. We plan to improve Sonic Home according to these feedbacks.

6. Discussion

6.1 Diversity of Sound Classes

In our system, a user can label a sound clip using predefined 24 sound classes (categories). While there are many kinds of sound categories in our real environments, our game system does not allow users to add a new sound class. This is because the users can add meaningless and/or redundant sound categories. Also, in-game items associated with a new sound class should also be prepared. As a part of our future work, we plan to develop a new game function that permits a user to propose a new additional sound category as well as upload his/her created in-game items.

6.2 Private Information in Sound

A sound clip uploaded by a user can contain private information of the user, e.g., speech including a name and a telephone number. We believe that there are two approaches to cope with this problem. The first one is asking a user to confirm a sound clip before file upload. That is, before the user uploads the sound clip, a pop-up that asks the user to confirm the sound is shown. The second one is using speech recognition techniques. When a sound segment including a human voice is detected, the segment is automatically removed.

7. Conclusion

This paper proposed and implemented an environmental sound collection game named Sonic Home. Our game has functions for controlling the diversity and quality of collected sensor data. We conducted an experiment to evaluate the developed functions and confirmed the effectiveness of our functions. As a part of our future work, we plan to release our game to application delivery systems (e.g., Apple AppStore and Google Play) and conduct a large scale evaluation.

Also, investigating the effect of real-world recording conditions, e.g., an echo and reverberation, is an important future work.

Because such conditions can reduce the sound recognition performance, we should reduce the weight of a sound clip recorded under such conditions. In our current system, a user evaluates a sound clip based on the three criterias: label correctness, sound volume, and presence of silent sound segments. We plan to add a new criterion related to sound conditions to solve the problem. Note that we should investigate whether or not an end user can detect such recording conditions as echo and reverberation by just listening to a sound clip.

Furthermore, we plan to investigate the effectiveness of our implemented functions on a user belonging to a certain game player category. For example, the ranking function can stimulate the feelings of a user belonging to the achiever category.

Acknowledgments This work has been partially supported by Hayao Nakayama Foundation for Science, Technology and Culture.

References

[1] Bartle, R.: Bartle Test of Gamer Psychology, GamerDNA, Technical report, available from ⟨http://www.gamerdna.com/quizzes/bartle-test-of-gamerpsychology⟩ (accessed 2010-12-08).

[2] Bartle, R.: Hearts, clubs, diamonds, spades: Players who suit MUDs, *Journal of MUD Research*, Vol.1, No.1, p.19 (1996).

[3] Bell, M., Reeves, S., Brown, B., Sherwood, S., MacMillan, D., Ferguson, J. and Chalmers, M.: EyeSpy: Supporting navigation through play, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.123–132, ACM (2009).

[4] Berchtold, M., Budde, M., Gordon, D., Schmidtke, H. and Beigl, M.: ActiServ: Activity recognition service for mobile phones, *International Symposium on Wearable Computers*, pp.1–8 (2010).

[5] Burke, J.A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S. and Srivastava, M.B.: Participatory sensing, *ACM SenSys Workshop on World-Sensor-Web* (2006).

[6] CENS/UCLA: Participatory Sensing / Urban Sensing Projects, available from ⟨http://research.cens.ucla.edu/⟩.

[7] Chen, J., Kam, A.H., Zhang, J., Liu, N. and Shue, L.: Bathroom activity monitoring based on sound, *Pervasive Computing*, pp.47–61, Springer (2005).

[8] Choudhury, T., Consolvo, S., Harrison, B., Hightower, J., LaMarca, A., LeGrand, L., Rahimi, A., Rea, A., Bordello, G. and Hemingway, B., et al.: The mobile sensing platform: An embedded activity recognition system, *IEEE Pervasive Computing*, Vol.7, No.2, pp.32–41 (2008).

[9] Dempster, A., Laird, N. and Rubin, D., et al.: Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B* (*Methodological*), Vol.39, No.1, pp.1–38 (1977).

[10] Deterding, S., Dixon, D., Khaled, R. and Nacke, L.: From game design elements to gamefulness: Defining gamification, *15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pp.9–15 (2011).

[11] Han, K., Graham, E.A., Vassallo, D. and Estrin, D.: Enhancing motivation in a mobile participatory sensing project through gaming, *IEEE 3rd International Conference on Privacy, Security, Risk and Trust* (*PASSAT 2011*) *and IEEE Third Inernational Conference on Social Computing* (*SocialCom 2011*), pp.1443–1448 (2011).

[12] Hattori, Y., Inoue, S. and Hirakawa, G.: A large scale gathering system for activity data with mobile sensors, *International Symposium on Wearable Computers* (*ISWC 2011*), pp.97–100 (2011).

[13] Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T. and Campbell, A.T.: A survey of mobile phone sensing, *IEEE Communications Magazine*, Vol.48, No.9, pp.140–150 (2010).

[14] Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J. and Zimmerman, J.: I'm the mayor of my house: Examining why people use foursquare-a social-driven location sharing application, *CHI 2011*, pp.2409–2418 (2011).

[15] Lu, H., Pan, W., Lane, N., Choudhury, T. and Campbell, A.: SoundSense: Scalable sound sensing for people-centric applications on mobile phones, *MobiSys 2009*, pp.165–178 (2009).

[16] Mun, M., Reddy, S., Shilton, K., Yau, N., Burke, J., Estrin, D., Hansen, M., Howard, E., West, R. and Boda, P.: PEIR, the personal environmental impact report, as a platform for participatory sensing systems research, *MobiSys 2009*, pp.55–68 (2009).

[17] Nov, O., Arazy, O. and Anderson, D.: Dusting for science: Motivation and participation of digital citizen science volunteers, *2011 iConference*, pp.68–74 (2011).

[18] Rana, R.K., Chou, C.T., Kanhere, S.S., Bulusu, N. and Hu, W.: Earphone: An end-to-end participatory urban noise mapping system, *ACM/IEEE International Conference on Information Processing in Sensor Networks*, pp.105–116 (2010).

[19] Rossi, M., Troster, G. and Amft, O.: Recognizing daily life context using web-collected audio data, *International Symposium on Wearable Computers* (*ISWC*), pp.25–28 (2012).

[20] Thiagarajan, A., Ravindranath, L., LaCurts, K., Madden, S., Balakrishnan, H., Toledo, S. and Eriksson, J.: VTrack: Accurate, energy-aware road traffic delay estimation using mobile phones, *SenSys 2009*, pp.85–98 (2009).

[21] UC Berkeley/Nokia/NAVTEQ: Mobile Millennium, available from ⟨http://traffic.berkeley.edu/⟩.

[22] Von Ahn, L. and Dabbish, L.: Designing games with a purpose, *Comm. ACM*, Vol.51, No.8, pp.58–67 (2008).

Akira Masuda is a student at Graduate School of Information Science and Technology, Osaka University, Japan. His research interests include sensor data mining and web data mining.

Kun Zhang received his M.E. from Graduate School of Information Science and Technology, Osaka University, Japan in 2014. His research interests include human activity recognition.

Takuya Maekawa is an associate professor at Osaka University, Japan. His research interests include ubiquitous and mobile sensing. He received his Ph.D. in information science and technology from Osaka University in 2006.