Sound sensing using smartphones as a crowdsourcing approach

Sunao Hara*, Asako Hatakeyama*, Shota Kobayashi* and Masanobu Abe*
* Okayama University, Okayama, 700-8530, Japan
E-mail: hara@okayama-u.ac.jp Tel/Fax: +81-86-251-8250

***Abstract*—Sounds are one of the most valuable information sources for human beings from the viewpoint of understanding the environment around them. We have been now investigating the method of detecting and visualizing crowded situations in the city in a sound-sensing manner. For this purpose, we have developed a sound collection system oriented to a crowdsourcing approach and carried out the sound-collection in two Japanese cities, Okayama and Kurashiki. In this paper, we present an overview of sound collections. Then, to show an effectiveness of analyzation by sensed sounds, we profile characteristics of the cities through the visualization results of the sound.**

I. Introduction

Data collection and analysis are key technologies for a smart city [1]. For the success of data collection to cover a wide area, such as districts in the city, we need to elicit cooperation from residents [2]. Mobile phone sensing [3], [4] is a promising approach to sense a city's characteristics. Recently, mobile phones and smartphones equip a rich set of powerful embedded sensors. Especially, global positioning system (GPS) sensors and microphones are installed on almost all smartphones. Therefore, the use of smartphones as sensors is useful for collecting environmental sounds from many users with their locations as context information.

Noise mapping has attracted attention as the usage of the sound collected by mobile phone sensing. It is focused on noise pollution in a city and tried to visualize the noise environment of the city by collecting sound levels using crowdsourcing. EarPhone [5], NoiseSPY [6] and NoiseTube [7] are important research examples. These studies were utilize objective metrics from sound.

The Sound Around You Project [8] is an important work as well. Many subjective evaluations focusing on the soundscape aspect were collected in that study. Chatty maps [9] visualized 5 sound types on the map as social impressions extracted from Social Network Service. Subjective metrics had an important role in these studies, but objective metrics were not.

Environmental sound can be dealt with in two ways, namely, objectively and subjectively. Examples of the objective aspects are the sound of a car that makes us know a car is approaching from behind and the loudness of residential area, which is one of the factors we consider when buying a home. By contrast, the subjective aspect is not always extracted using only those measures but contexts also play an important role. In other words, humans interpret sounds differently based on their experiences and their current situation. For example, we may feel a sound is louder at night than at noon even if it is the same sound, and we may feel the sound of a car is louder in a quiet residential area than it is downtown. From these viewpoints, we aim to be visible information from environmental sound in terms of both objective and subjective aspects.

To construct models for extracting information about the subjective aspect, we collect environmental sound data with subjective judgments not only as much as possible but also from as many users as possible. To meet these requirements, we adopted a crowdsourcing approach by developing a smartphone application to collect environmental sounds [10]. We decided to use people's smartphones, which are equipped with microphones and GPS sensors, as recording devices since they are used very commonly nowadays. The application provides two modes for environmental sound collection: participatory [11], [12] and opportunistic sensing paradigms [13]. The participatory mode enables users to intentionally collect sound data, namely, the raw waveforms of sounds that participants are interested in or appreciate. The opportunistic mode automatically collects sound statistics, especially loudness levels. Subjective judgments are always and occasionally collected in the participatory and the opportunistic mode, respectively.

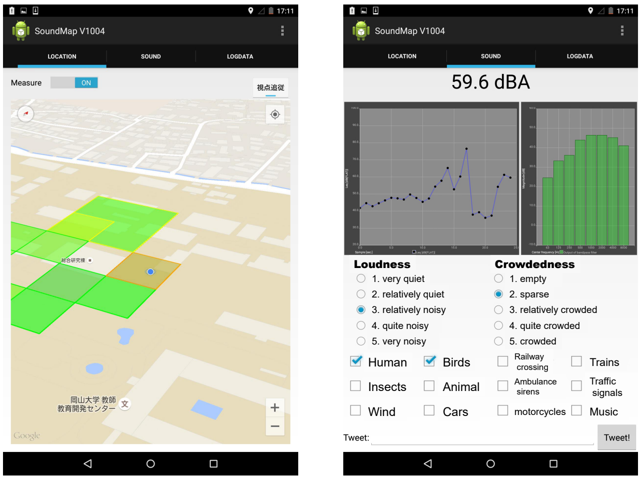
To confirm the values of the crowdsourcing sounds, we visualize the sounds as a sound map that contains objective and subjective informations. The sound map visualize two types of sound; loudness levels as a noise map and sound-type icons as a sound map. To evaluate the effectiveness of our sound map, we analyze characteristics of the target cities or of unusual events from the map.

The remainder of this paper is organized as follows. Section II presents a summary of the sound collection system, and Section III explains the database consists of sounds collected from real environments by using the system described in Section II. Section IV presents an analysis of the sound map constructed from the database to focus on some aspects. Finally, Section V summarizes this paper.

II. Sound collection system

A. *Overview of recording application*

We developed an application for recording environmental sound. We used a Google Nexus 7, a 7-inch touch screen tablet running Android OS. Figures 1(a) and (b) show screenshots of the location- and sound-logging screens, respectively. Data recording begins when a user slides the button on the upper side of the location-logging screen.



SoundMap V1004

LOCATION | SOUND | LOGDATA

Measure ON

視点追従

(a) Location-logging screen

SoundMap V1004

LOCATION | SOUND | LOGDATA

59.6 dBA

Loudness
- ○ 1. very quiet
- ○ 2. relatively quiet
- ● 3. relatively noisy
- ○ 4. quite noisy
- ○ 5. very noisy

Crowdedness
- ○ 1. empty
- ● 2. sparse
- ○ 3. relatively crowded
- ○ 4. quite crowded
- ○ 5. crowded

☑ Human ☑ Birds □ Railway crossing □ Trains

□ Insects □ Animal □ Ambulance sirens □ Traffic signals

□ Wind □ Cars □ motorcycles □ Music

Tweet: Tweet!

(b) Sound-logging screen

Fig. 1. Screenshots of developed Android application

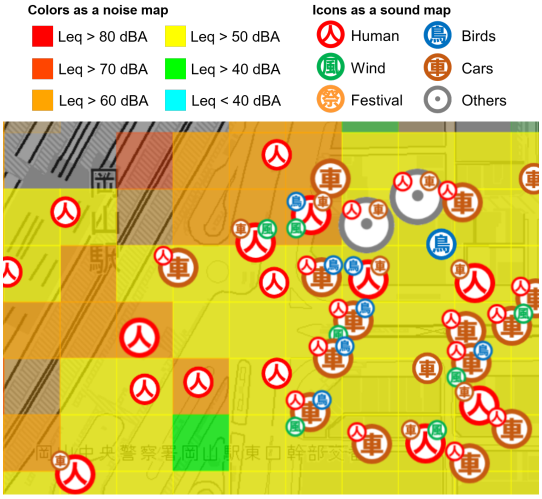


Fig. 2. Visualization loudness level and sound type distributions by colormap and icons

On the location-logging screen, the system can record highly accurate location information using GPS, Cell-ID, or Wi-Fi via the Android API. The default sampling period is 1 second, but the user can change this in the application settings. Color squares on a map on the screen show the history of the user's locations.

On the sound-logging screen, the system can record raw sound signals and calculate loudness levels using a microphone on the device. It always stores the sound data of the most recent 20 seconds in a ring buffer, and it analyzes the sound to calculate the equivalent loudness level and the levels of an eight-channel frequency filter bank at intervals of 1 second.

Users can attach annotations, such as subjective evaluations, sound-type selection, and free description, to a sound while recording. The subjective evaluation uses a five-grade scale for two metrics, subjective loudness level and subjective crowdedness level. The sound type is easy to annotate with a selection of 12 preset sound types. A free description can be used as a summary of features such as the recording environment, or users' feelings.

The application generates sound files and three types of log files in one session. The log files are a location history log file, loudness level log file, and tweet log file, each containing time information. We can merge these log files based on time information to realize both opportunistic sensing and participatory sensing.

B. Recording capability

The application calculates the statistics of recorded sounds and stores them as the loudness level log files in the device's storage. Note that the sounds are disposed after they are processed because of privacy reasons.

Sounds are recorded at a sampling frequency of 32,000 Hz and 16 bits over a single channel. They are analyzed at equivalent A-weighted loudness level [14], [15] L_{eq} per second.

We stored the recorded sounds as WAV files only when users pushed the "Tweet!" button on the interface. Other related log files include location history log files and tweet log files. The tweet log files contain subjective evaluations of recording environments and annotations of the recorded sound. These evaluations and annotations are on-site impressions of the listeners.

A WAV file containing the last 10 second of sound is created by pushing the tweet button on the sound-logging screen (Fig. 1 (b)). To add an annotation to the sound, participants select the sound type before pushing the tweet button. Twelve types of sound are preset for ease of use and users can select multiple types: T_1: human speech, T_2: birds, T_3: insects, T_4: cars, T_5: wind, T_6: motorcycles, T_7: railway crossing, T_8: trains, T_9: ambulance sirens, T_{10}: traffic signals, T_{11}: music, and T_{12}: animals.

Additionally, users can input free text to annotate the sound or the recording environment. They are not required to fill in all selections, but they can input just one part with an annotation if they want to check one or more metrics.

C. Visualization

A visualization system is implemented as a web application by using the open-source libraries, Leaflet[1] and D3.js[2]. The system can visualize two types of data in the server system described in Section II. An example of environmental sound visualization is shown in Fig. 2.

Visualization of the loudness data is provided through a color map of each area. The color index is calculated from the average loudness values in each squares. We can overview the loudness distribution of any district of interest on the map. The color indicates the average loudness level; for example, red indicates higher loudness than blue. The transparency shows the number of data in the area; for example, the weaker the transparency, the fewer the data. In other words, weak transparency indicates non-confident data.

[1] http://leafletjs.com/

[2] https://d3js.org/

TABLE I
CONDITION OF DATABASE RECORDINGS IN OKAYAMA CITY

Recording No. 1 (Nov. 2014)	
Date	Nov. 27 and 28, 2014 (as weekdays) Nov. 22 and 29, 2014 (as holidays)
# of subjects	8 (one subject for each area in each hour)
Areas	A1: Quiet residential area A2: Shopping street far from a station A3: Shopping street near a station A4: Downtown area near a station (recording in two areas in one day)

Recording No. 2 (Jan. 2015)	
Date	Jan. 14 and 27, 2015 (as weekdays) Jan. 24 and 31, 2015 (as holidays)
# of subjects	6 (one subject for each area for each hour)
Areas	(A1': Quiet residential area* A2: Shopping street far from a station A3: Shopping street near a station A4: Downtown area near a station (recording in two areas in one day)

* (A1') is another area from (A1)

TABLE II
TYPE OF ENVIRONMENTAL SOUNDS AND THEIR DISTRIBUTION

Class		**# of data for each area**				
		A1	A1'	A2	A3	A4
T_1	Human speech	387	297	906	498	707
T_2	Birds	343	555	107	430	148
T_3	Insects	54	0	1	1	1
T_4	Cars	659	574	1389	1051	880
T_5	Wind	55	334	119	120	108
T_6	Motorcycles	348	121	341	182	235
T_8	Trains	0	0	123	6	129
T_9	Ambulance sirens	26	33	14	24	15
T_{10}	Traffic signals	76	1	555	458	526
T_{11}	Music	15	16	356	111	343
T_{12}	Animals	55	85	3	4	3

Sound visualization is achieved using icons symbolizing sounds on the map, which help us see the types of sound in any district of interest. These sounds are distinguished by icons based on their subjective evaluations during recording. An icon can be clicked to browse the annotation associated with the sound. The neighboring icons are clustered using the Leaflet.MarkerCluster plugin[3]. The large icons denote the major sound classes in each cluster, and the small icons around the large icons denote other sound classes in the cluster.

III. DATA COLLECTION

A. Data collection in Okayama City

1) Conditions of data collection: The detailed condition of data collection is summarized in Table I. Data were collected by 14 participants in four types of areas. The participants were instructed on how to use smart devices and the data collection applications. They were asked to collect sounds, annotations, and loudness levels. They were asked to travel around predefined routes in each area for 1 hour. The rounds were repeated from 8 a.m. to 9 p.m. Two of the participants were recording simultaneously.

The participants recorded loudness levels with the application running and sounds with annotations at various intervals. They held the devices in their hands during data collection, keeping them in an appropriate position for collecting clear sound samples. However, footstep noise could be mixed in with the recorded sound because the participants might have been handling the device while walking, which can cause bias in the loudness levels.

2) Summary of collected data: All collected data were synchronized with their time information, and we obtained 693,582 loudness data with tuples of latitude, longitude, and time. The sound data comprised 5,935 collected samples with 10 second of sound within a tuple.

The distribution of the collected sound data of each type is summarized in Table II. Note that, class T_7 is cleansed from the table because there are no railway crossing in the recording area. From the table, the residential areas (A1 and A1') have different distribution of sounds compared to the other areas (A2, A3 and A4). The residential areas contained more "birds", "insects", and "animals" but less "human speech", "cars", and "music" than the other areas. It means that they are quieter. The difference between "insects" and "wind" is potentially an indicator of the difference in area characteristics. However, this difference can reasonably be generated from a difference in the season in which the recordings were made. Between Area A3 and the Areas A2 or A4, Area A3 has high levels of "human speech" and "birds" but low levels of "music". This fact suggests that Area A3 is a quiet shopping area compared to the other areas.

B. Data collection in Kurashiki city

1) Conditions of data collection: The data collection conditions are summarized in detail in Table III. Basically, the recordings were carried out in the same manner as that mentioned at Section III-A.

In the case of Recording No. 4 on Oct. 16, 16 participants recorded at the same time. Eight of the participants were traversing around predefined routes and the other eight participants were standing at fixed locations. We attempted to capture details of the area by increasing the number of participants in the area.

Additionally, we recorded new sound classes "footstep noise (T_{12})" and "festival (T_{13})" in this round of data collection. The sounds "festival" and "footstep noise" might be used to refer to bustle in the area.

2) Summary of collected data: In recording No. 3, we obtained 249,480 loudness data and 975 sound data. In recording No. 4, we obtained 556,380 loudness data and 1,913 sound data.

The distribution of the sound data collected for each type is summarized in Table IV. Note that, classes T_7 and T_8 are cleansed from the table because there are no railway crossings or trains in the recording area. The classes "human speech"

[3] https://github.com/Leaflet/Leaflet.markercluster

TABLE III
CONDITION OF DATABASE RECORDINGS IN KURASHIKI CITY

Recording No. 3 (May 2016)	
Date	May 5, 2016 (as holidays*) May 25, 2016 (as weekdays) from 9 a.m. to 9 p.m.
# of subjects	8 (two subjects were recording at the same time)
Areas	A5: Kurashiki Bikan Historical Quarter - two predefined routes with the same start position - two static recording points (without subjects)

Recording No. 4 (Oct. 2016)	
Date	Oct. 16, 2016 (on the day of the festival) from 7 a.m. to 6 p.m.
# of subjects	34 (16 subjects were recording at the same time)
Areas	A5: Kurashiki Bikan Historical Quarter - one predefined route with eight start positions - eight static recording points (with subjects)

* May 5 is a public holiday in Japan.

TABLE IV
TYPES OF ENVIRONMENTAL SOUNDS AND THEIR DISTRIBUTIONS IN THE KURASHIKI RECORDINGS. THE NUMBERS IN THE PARENTHESES SHOW THE PROPORTIONS OF THE SOUND CLASSES IN THE ENTIRE SOUND DATA. NOTE THAT EACH SOUND CAN BELONG TO MULTIPLE SOUND CLASSES.

Class		**# of data for each recording**			
		Recoding No. 3		Recording No. 4	
T_1	Human speech	631	(65%)	1,598	(84%)
T_2	Birds	190	(19%)	534	(28%)
T_3	Insects	28	(3%)	61	(3%)
T_4	Cars	412	(42%)	539	(28%)
T_5	Wind	108	(11%)	207	(11%)
T_6	Motorcycles	72	(7%)	72	(4%)
T_9	Ambulance sirens	16	(2%)	29	(2%)
T_{10}	Traffic signals	129	(13%)	82	(4%)
T_{11}	Music	109	(11%)	170	(9%)
T_{12}	Animals	5	(1%)	25	(1%)
T_{13}	Footstep noise	490	(50%)	1,388	(73%)
T_{14}	Festival	—		239	(12%)

and "footstep noise" are relevantly increased the proportion in Recording No. 4 than Recording No. 3. This difference may or may not have been caused by the festival. From the result, we can deduce the bustle caused by the festival.

C. Questionnaires about the data collection

We have asked the participants of the Recording No. 4 to answer two questionnaires. The respondents are 33 of participants excluding the experimenter. They answered after finishing their recording.

The first questionnaire is that "what kind of information are you looking for when participating a festival?" The subjects are selected multiple choices from 15 options. This question is designed to know the effective application that is able to work with our sound recording applications, though, three situations are provided to them. The situations are (1) planning the attendance of the festival, (2) preparing at the day before the festival, and (3) attending the festival.

Figure 3 is population for each option of the first questionnaire. Note that we merged the similar result, and shows as the result of 7 selections. From the figure, we can see that on-site information is heavily required at the situation (3). In contrast, the abstract of the festival, the weather, the venue, and transportation information are needed for the situations (1) and (2). To see another viewpoint, the information of "entertainments" is constantly needed for any situation. This result suggested that a system for entertainment information guide is able to motivate users to install and use the application. Furthermore, if we can design a collaborative application between the guidance system and our sound recording, the sound sensing as a crowdsourcing approach might be more successful.

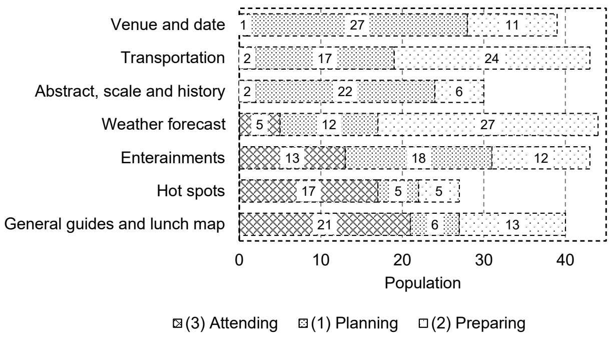


Fig. 3. Demands of the festival's information for each situation

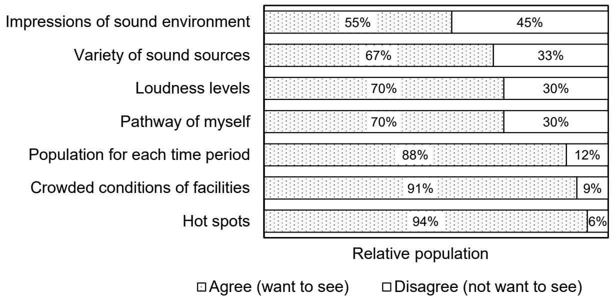
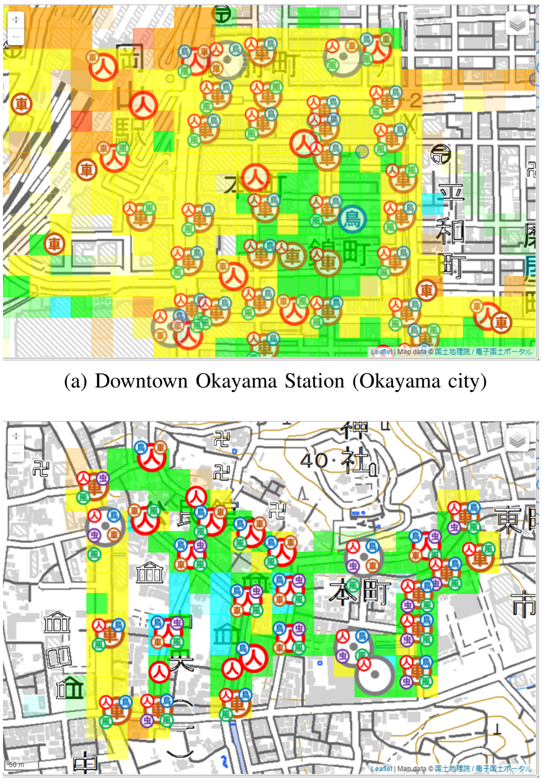


Fig. 4. Demands of visualization using the environmental sound

The second questionnaire is that "what kind of visualized information do you want from today's recordings?" The participants are presented seven kinds of informations and asked to answer them with four levels: that is, (1) strongly disagree, (2) somewhat disagree, (3) somewhat agree, and (4) strongly agree.

Figure 4 is a relative population for answers to each visualization of the second questionnaire. The four-level answers are merged in two options, i.e. agree or disagree. From the result, visualizations of hot spots and crowded conditions of facilities are needed to the participants. The visualization of the population for each time period is also needed. This result suggested that the system to visualize such an information might be acceptable for the sound recording users, and such a function is needed to implement a newly developed visualization system.



(a) Downtown Okayama Station (Okayama city)

(b) Kurashiki Bikan Historical Quarter (Kurashiki city)

Fig. 5. Sound map of two cities in Okayama prefecture

IV. Analysis of sound map

In this section, we explain the power of our sound map in analyzing the characteristics of the city. The sound map contains two types of information, namely, a noise map as a color map and a sound-type map as icons. These types of information mutually energize the analysis.

A. *Difference between Okayama and Kurashiki*

As an example, we attempted to analyze and compare the characteristics of cities from our sound maps, as shown in Fig. 5. The area shown in Fig. 5(a) contains shopping streets, business buildings, main roads, and a large-scale shopping mall. Visitors can enjoy the vibrant and bustling atmosphere of the area. By contrast, Kurashiki city is known as an important historical city of the prefecture. The area shown in Fig. 5(b) contains one of the most famous sightseeing area in Okayama, namely, Bikan Historical Quarter [4]. Visitors can enjoy the quiet and peaceful atmosphere of the area.

The figure shows that Okayama is noisier than Kurashiki. However, by focusing on the distribution of sound icons, we can see that Kurashiki contains many "human speech" icons, whereas, Okayama contains many "cars" icons. This fact supports the characteristics of the Kurashiki Bikan Historical Quarter as a tourist spot. That is, Kurashiki is usually a silent place, but there are many people around the area.

B. *Analysis of the effect of an unusual event*

A festival was held on Oct. 16, 2017 at the Achi Shrine, which is located at the center of Kurashiki Bikan Historical Quarter. Figure 6 shows the sound maps for 2 hour during the festival. In the early morning (Fig. 6(a)), there are few people. The area is very quiet, so mostly "bird" sounds were recorded. This characteristic is similar to the usual characteristic of Kurashiki city, as mentioned in Section IV-A. Then, the area started bustling from 8 a.m. to the noon. (Fig. 6(b). "Human speech" started becoming visible. Next, the peak of the festival is arrived from the noon, and it is continued until about 4 p.m. (Fig. 6(c)). There are several short events during the festival, and they are denoted by red rectangles on the sound map as noisy areas. Finally, the bustle appears to start clearing, but the "human speech" remain (Fig. 6(d)). Some tourists might have still been charged up in that period.

V. Conclusion

In this paper, we introduce sound collection experiments in two cities, Okayama and Kurashiki. The collected data were visualized as a noise map and a sound-type map, which were constructed from the distribution of loudness levels and sound types, respectively. Finally, we analyzed the characteristics of the cities from our sound maps to evaluate the effectiveness of the map. As a result, the visualization of sound-type icons on the noise map gave us more information than just looking at the noise-only map.

The effectiveness of the analysis using our sound maps was demonstrated through the experiments, but there remains work to be done in the future. For example, microphones must be calibrated appropriately for another smartdevices, if a greater number of participants join our sound collection. Furthermore, raw waveforms of sound were not used because of privacy concerns. If privacy protection methods for raw waveforms of sound are established, more information can be visualized on our sound map.

Acknowledgment

This work has been supported by Strategic Information and Communications R&D Promotion Programme (SCOPE) from Ministry of Internal Affairs and Communications, Japan, and JSPS KAKENHI Grant Number JP15K01069.

References

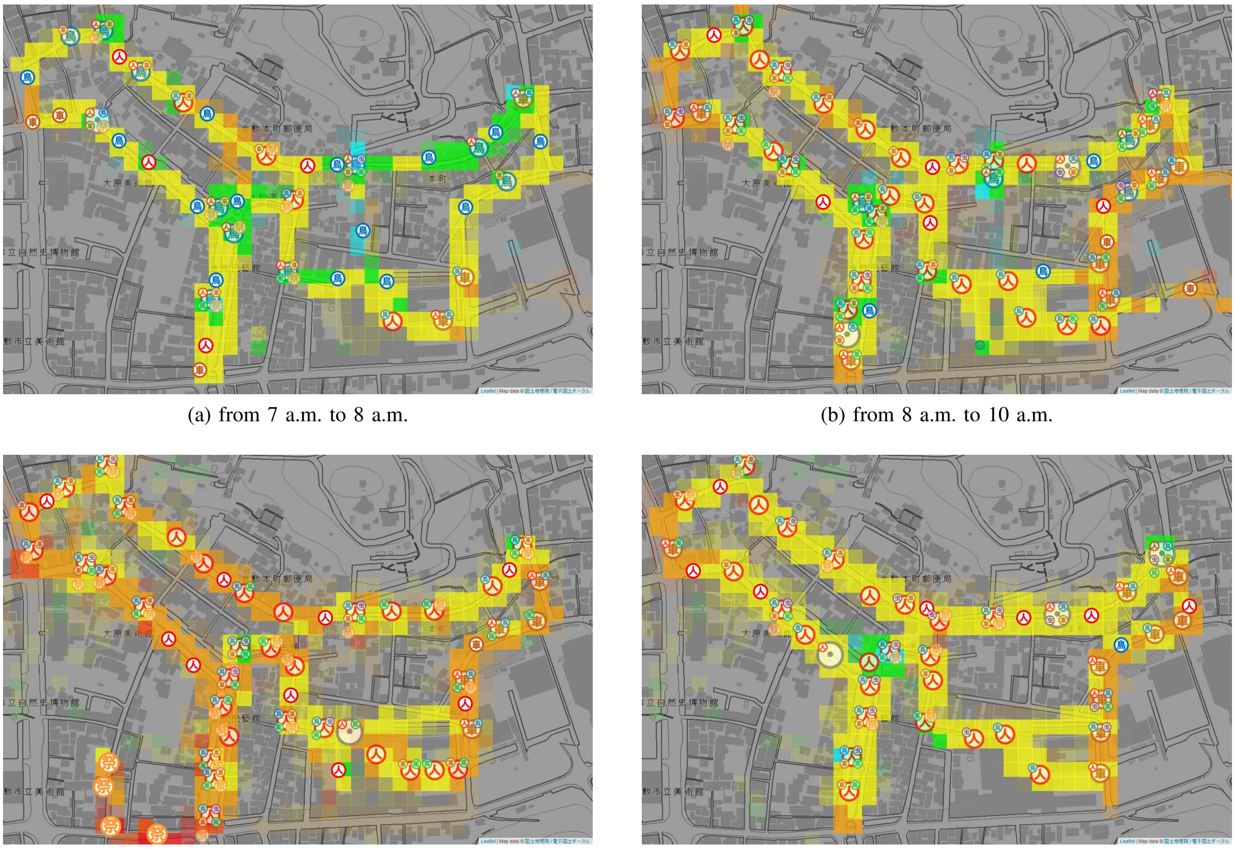
[1] M. Matsuoka, N. Ueda, H. Tokuda, R. Lea, and L. Muñoz, "SmartCities15: International workshop on smart cities: People, technology and data," Proceedings of UbiComp/ISWC15 Adjunct, pp.1509–1513, Sept. 2015.

[2] D. Gooch, A. Wolff, G. Korteum, and R. Brown, "Reimagining the role of citizens in smart city projects," Proceedings of UbiComp/ISWC15 Adjunct, pp.1587–1594, Sept. 2015.

[3] N.D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A.T. Campbell, "A survey of mobile phone sensing," IEEE Communications Magazine, vol.48, no.9, pp.140–150, Sept. 2010.

[4] W. Z. Khan, Y. Xiang, M. Y Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," IEEE Communications Surveys and Tutorials, vol.15, no.1, pp.402–407, Feb. 2013.

[4] https://www.kurashiki-tabi.jp/for/en/bikan.html



(a) from 7 a.m. to 8 a.m.

(b) from 8 a.m. to 10 a.m.

(c) from noon to 2 p.m.

(d) from 4 p.m. to 6 p.m.

Fig. 6. Sound map at Kurashiki on Oct. 16, 2016. A festival was held on this day.

[5] R.K. Rana, C.T. Chou, S.S. Kanhere, N. Bulusu, and W. Hu, “Ear-Phone: An end-to-end participatory urban noise mapping system,” Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks, pp.105–116, April 2010.

[6] E. Kanjo, “NoiseSPY: A real-time mobile phone platform for urban noise monitoring and mapping,” Mobile Networks and Applications, vol.15, no.4, pp.562–574, Aug. 2010.

[7] E. D’Hondt, M.A. Stevens, and A. Jacobs, “Participatory noise mapping works! an evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring,” Pervasive and Mobile Computing, vol.9, no.5, pp.681–694, Oct. 2013.

[8] C. Mydlarz, I. Drumm, and T. Cox, “Application of novel techniques for the investigation of human relationships with soundscapes,” Proceedings of INTERNOISE 2011 congress, pp.738–744, Sept. 2011.

[9] L.M. Aiello, R. Schifanella, D. Quercia, and F. Aletta, “Chatty maps: constructing sound maps of urban areas from social media data,” Royal Society Open Science, March 2016.

[10] S. Hara, M. Abe, and N. Sonehara, “Sound collection and visualization system enabled participatory and opportunistic sensing approaches,” Proceedings of CASPer-2015, pp.390–395, March 2015.

[11] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M.B. Srivastava, “Participatory sensing,” Proceedings of ACM workshop of World-Sensor-Web, ACM Sensys, pp.117–134, Oct. 2006.

[12] J. Goldman, K. Shilton, J.A. Burke, D. Estrin, M. Hansen, N. Ramanathan, S. Reddy, V. Samanta, M. Srivastava, and R. West, “Participatory sensing: A citizen-powered approach to illuminating the patterns that shape our world.” Woodrow Wilson International Center for Scholars, Washington, D.C., May 2009.

[13] A.T. Campbell, S.B. Eisenman, N.D. Lane, E. Miluzzo, and R.A. Peterson, “People-centric urban sensing,” Proceedings of WICON-06, Article No. 18, Aug. 2006.

[14] C.A. Kardous and P.B. Shaw, “Evaluation of smartphone sound measurement applications,” Journal of Acoustical Society of America Express Letters, vol.135, no.4, pp.EL186–192, April 2014.

[15] H. Fletcher and W.A. Munson, “Loudness, its definition, measurement and calculation,” Journal of Acoustical Society of America, vol.5, no.82, pp.82–108, Oct. 1933.