SOUND COLLECTION SYSTEMS USING A CROWDSOURCING APPROACH TO CONSTRUCT SOUND MAP BASED ON SUBJECTIVE EVALUATION

Sunao Hara, Shota Kobayashi and Masanobu Abe

SOUND COLLECTION SYSTEMS USING A CROWDSOURCING APPROACH TO CONSTRUCT SOUND MAP BASED ON SUBJECTIVE EVALUATION

Sunao Hara, Shota Kobayashi and Masanobu Abe

Okayama University
Graduate school of Natural Science and Technology
3-1-1, Tsushima-naka, Kita-ku, Okayama, Japan.
hara@okayama-u.ac.jp

ABSTRACT

This paper presents a sound collection system that uses crowdsourcing to gather information for visualizing area characteristics. First, we developed a sound collection system to simultaneously collect physical sounds, their statistics, and subjective evaluations. We then conducted a sound collection experiment using the developed system on 14 participants. We collected 693,582 samples of equivalent A-weighted loudness levels and their locations, and 5,935 samples of sounds and their locations. The data also include subjective evaluations by the participants. In addition, we analyzed the changes in sound properties of some areas before and after the opening of a large-scale shopping mall in a city. Next, we implemented visualizations on the server system to attract users' interests. Finally, we published the system, which can receive sounds from any Android smartphone user. The sound data were continuously collected and achieved a specified result.

***Index Terms*—** Environmental sound, Crowdsourcing, Loudness, Crowdedness, Smart City

1. INTRODUCTION

Data collection and analysis are key technologies for a smart city [1]. For the success of data collection in a smart city, we need to gain the cooperation and participation of residents [2]. Mobile phone sensing [3, 4] is a promising approach for residents to sense a city's characteristics. Mobile phones and recent smartphones contain a rich set of powerful embedded sensors. Especially, Global Positioning system (GPS) sensors and microphones are installed on most smartphones, although the set of installed sensors varies among smartphones. Consequently, sound collection with location information using sensors is a hopeful approach for the success of data collection with cooperation from residents.

In this study, we developed a sound collection method that uses crowdsourcing to understand environmental sounds by considering contextual information. The sound collection is performed using an Android application [5] on a smart device. The collected data fall into two types, user-specific and statistical. We use two crowdsourcing paradigms to collect the sounds: participatory [6, 7] and opportunistic sensing [8]. Using the participatory sensing paradigm, we can collect sounds that participants are interested in or appreciate, therefore, we used this paradigm to collect the waveforms of sounds. Using the opportunistic sensing paradigm, we can collect sound statistics and, in particular, the loudness levels as statistics.

Moreover, we developed a visualization method for the sounds collected using participatory and opportunistic sensing. This visualization is one of the most important capabilities for interpreting environmental sounds. The waveforms of sounds are visualized as icons symbolizing the sounds at particular locations on a map, and the statistics of the sounds are visualized as colors on the same map. We also implemented system functions that can be used for crowdsourcing-based sound collection.

2. BACKGROUND

Sound properties are generally interpreted as having spectral and/or temporal parameters, such as spectrum, fundamental frequency, and loudness. However, these parameters only interpret the sound properties on the basis of a common understanding of human beings; this is insufficient. To understand environmental sounds in the real world, we need to consider contextual information, i.e., not only sound properties, but also the situation of the listener.

The data must be statistically processed or anonymized to reduce any privacy risk for public systems. From this perspective, EarPhone [9] and NoiseTube [10] are important examples. In these studies, the researchers attempted to collect environmental sounds as sound levels using crowdsourcing; in other words, they dealt primarily with the statistics of sounds. McGraw *et al.* [11] collected sound data using

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

Amazon Mechanical Turk as a crowdsourcing platform. Matsuyama *et al.* [12] conducted their sound-data collection using an HTML5 application and evaluated the performance of sound classifiers. Their study deals primarily with the raw waveforms of sounds, which cannot identify the listener. In contrast with these studies, the principal contribution of our paper is to enable sound-data collection that takes contextual information into account.

In this paper, we refine a system that was implemented by Hara *et al.* [5]. In particular, A-weighted loudness levels [13, 14] can be recorded in the application.

3. SOUND COLLECTION SYSTEM

3.1. Recording application for environmental sound

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

On the location-logging screen, the system can record highly accurate location information using GPS, Cell-ID, or Wi-Fi networks via the Android API. The default sampling rate is 1 s, but the user can change this through the settings. Pin icons on a map on the screen can show the history of the user's locations.

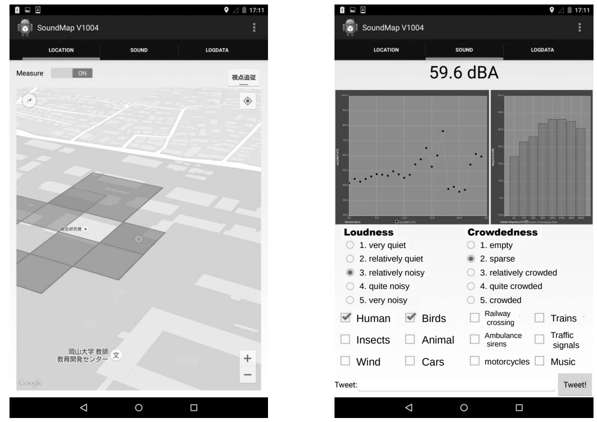
On the sound-logging screen, the system can record raw sound signals and calculate loudness levels using a microphone on the device. It always stores the sound data of the most recent 20 s using a ring buffer, and it also analyzes the sound to calculate the equivalent loudness level and the levels of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

Users can attach annotations, such as subjective evaluation, 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 five preset sound types. A free description can be used as a summary of such features as the recording environment, or feelings.

All of the annotations are recorded in log files with time information, and a WAV file including 10 s of sound is created at the same time. These can be sent to a server, if the application settings permit. The sent log files are parsed on the server and shown in a timeline view that is similar to that of Twitter, and is shared for all users in the implementation.

3.2. Specification of the data collected by the application

The application generates sound files and three types of log file in one session. The log files are a location history log file, loudness level log file, and tweet log file, each containing time information, which is triggered.



(a) Location tab page (b) Sound tab page

Fig. 1: Screenshots of Android application

Sounds are recorded at a sampling frequency of 32,000 Hz and 16 bits per second with a single channel. They are analyzed at equivalent A-weighted loudness level [13, 14] L_{eq} per second:

$$X[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \cdot e^{-2\pi jkn/N} \tag{1}$$

$$L_{eq} = B\left(10 \log_{10} \frac{1}{K} \sum_{k=0}^{K-1} \left| A[k] \cdot X[k] \right|^2 \right) \tag{2}$$

where $x[n]$ is a sampled signal, N is the signal length, $A[k]$ is an A-weighting filter, K is an FFT length ($K > N$), and $B(\cdot)$ is a transform function from a power of quantized waveform to a sound pressure level. In this paper, N is fixed to 32,000, which is equivalent to 1 s. $B(\cdot)$ is detected through a preliminary examination to compare with values of a sound level meter, RION NL–42.

The microphone specification must be calibrated appropriately if it is to be used in a real crowdsourcing environment. For this purpose, we measured sound properties and prepared $B(\cdot)$ functions for 22 devices. Automatic detection of the calibration parameter is future work.

In addition to L_{eq}, this system can also record filter bank output levels in eight-channel, which is related to octave band filter analysis. The filter is implemented using triangle windows. The central frequencies of the filter are $f_c = [63, 125, 250, 500, 1000, 2000, 4000, 8000]$.

3.3. Server application for collection and exploration of sounds

The client and server applications communicate via HTTP protocols. The server implements APIs for receiving and browsing data and the browsing API can create not only a general HTML view for web browsers, but also a JSON (JavaScript Object Notation) view for advanced applications.

Table 1: Condition of database recordings

Recording 1 (Nov. 2014)	
Date	Nov. 27 and 28, 2014 (as weekdays) Nov. 22 and 29, 2014 (as holidays)
# of subjects	eight subjects (`U11–U18`) (one subject for each area at one hour)
Areas	(1) Quiet residential area (2) Shopping street far from a station (3) Shopping street near a station (4) Downtown area near a station (recording at two areas for one day)

Recording 2 (Jan. 2015)	
Date	Jan. 14 and 27, 2015 (as weekdays) Jan. 24 and 31, 2015 (as holidays)
# of subjects	six subjects (`U21–U26`) (one subject for each area at one hour)
Areas	(1') Quiet residential area* (2) Shopping street far from a station (3) Shopping street near a station (4) Downtown area near a station (recording at two areas for one day)

* (1') is another area from (1)

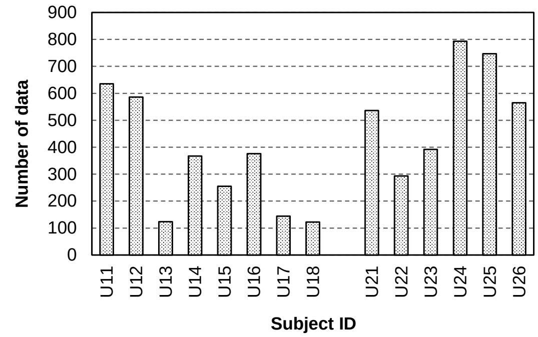


Fig. 2: Number of sound data as a function of a subject ID

The server system includes several open-source software applications. The server OS is a Debian GNU/Linux 7.5 (Wheezy). The web application framework is Mojolicious[1] with Perl. The back-end database software is MongoDB[2]. The application runs on Mojolicious Hypnotoad with an nginx front-end server[3]. The system is used for the crowdsourced sound recordings; hence, a large number of users will use the system, and it must have the appropriate processing capacity. These software have a distributed computing architecture that might provide an answer to problems of heavy usage.

4. DATABASE CONSTRUCTION

4.1. Conditions of data collection

The detailed condition of data collection is summarized in Table 1. Data was collected by a total of 14 participants at four types of areas. The participants were instructed on how to use smart devices and the data collection applications. They were asked to collect the sounds, annotations, and loudness levels. They were asked to travel around static routes for each area in 1 h. The rounds were repeated from 8 a.m. to 9 p.m.

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

The subjective loudness level is evaluated on a five level scale: L_1: very quiet, L_2: relatively quiet, L_3: relatively noisy, L_4: quite noisy, and L_5: very noisy. The subjective crowdedness level is also evaluated on a five level scale: S_1: empty, S_2: sparse, S_3: relatively crowded, S_4: quite crowded, and S_5: crowded. The subjective evaluations are recorded as annotations.

The sound file containing the last 10 s 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. Five types of sound are preset for ease of use and users are allowed to select multiple choices: 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, participants can input free text to annotate the sound or recording environment. They are not required to fill in all of the selections, but can input just one part with an annotation if they want to check one or more metrics.

4.2. Summary of collected data

All of the 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 s of sound with the same tuples. The number of collected data for each user is shown in Fig. 2. A distribution of the sound data collected for each type is shown in Table 2.

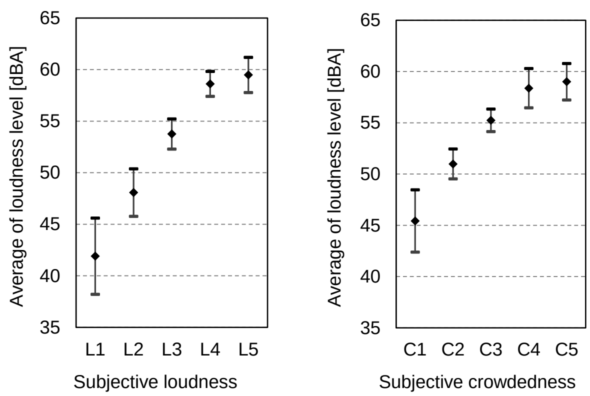
4.3. Analysis focused on subjective evaluations

Figures 3(a) and (b) are the average loudness levels as functions of the subjective loudness and crowdedness levels, re-

[1] `http://mojolicio.us/`
[2] `http://www.mongodb.org/`
[3] `http://nginx.org/`

Table 2: Type of environmental sound and its distribution

Class		# of data
T_1	Human speech	2,882
T_2	Sound of birds	1,632
T_3	Sound of insects	58
T_4	Sound of cars	4,697
T_5	Sound of wind	760
T_6	Sound of motorcycles	1,270
T_7	Sound of railway crossing	1
T_8	Sound of trains	267
T_9	Sound of ambulance sirens	115
T_{10}	Sound of traffic signals	1,679
T_{11}	Sound of music	873
T_{12}	Sound of animals	153



(a) Subjective loudness (b) Subjective crowdedness

Fig. 3: Average loudness level as a function of a subjective evaluation. The error bars indicate 90% confidence intervals as estimates of average loudness levels for the subjective level.

spectively. The average value is calculated as the average of the data from the 14 participants, and the error bars are indicative of 90% confidence intervals. We find overlapping error bars in Fig. 3(a) at L_4 and L_5. Figure 3(b) has a similar tendency at C_4 and C_5. This is not trivial because of the design of our questionnaire, but its long error bars show the importance of listener-specific information for sound interpretation.

Table 3 shows frequency for the subjective crowdedness and loudness data as a contingency table. The number of data is zero or very small at low loudness levels and high crowdedness level, e.g., C_4 and L_1. However, this is not the case for small values at high loudness levels and low crowdedness level, e.g., C_1 and L_4. This asymmetric property indicates that estimating loudness levels from a crowdedness level is much easier than estimation crowdedness levels from a loudness level.

Table 3: Frequency of sound data for subjective crowdedness and loudness

	L_1	L_2	L_3	L_4	L_5	(null)	TOTAL
C_1	390	417	221	32	3	3	1,066
C_2	134	1,437	1,228	199	44	7	3,049
C_3	2	111	745	241	21	6	1,126
C_4	0	4	79	178	39	1	301
C_5	0	0	2	26	26	1	55
(null)	9	36	52	16	4	221	338
TOTAL	535	2,005	2,327	692	137	239	5,935

4.4. Analysis focused on the time series at hour of day

Figure 4 shows time series of loudness level for each area. The sequences at a quiet residential area (Fig. 4(a)) show little difference. We can imagine that area (1), recorded on November 2014, is noisy at night, but that area (1'), recorded on January 2015, is quiet at night. It is vice versa from 2 p.m. to 3 p.m.

We can see in Fig. 4(d) that the loudness level undergoes a major change from night to morning. One of the reasons is the opening of a large-scale shopping mall at the beginning of November 2014. The impact of the attraction of a large-scale retail store might be what is shown in this loudness level chart.

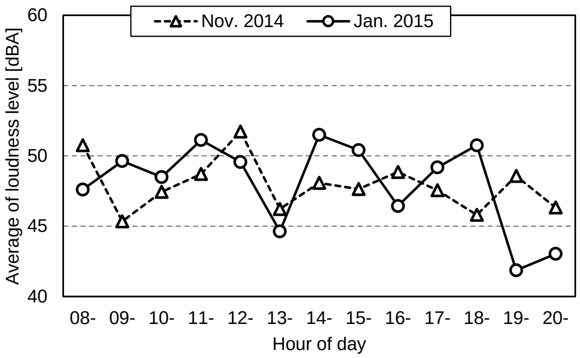
We must consider from Figures 4(b) and (c) the impact of the opening of the shopping mall on the existing shopping streets. The figure shows that there are fewer changes in the morning and evening. These are commuting times to school and work and thus show no change from before to after the opening of the shopping mall. On the other hand, it shows a non-negligible impact at noon and night. Needless to say, the charts show only that the loudness level decreases in the area. However, this might indicate that there are fewer people in the area than there were previously.

5. SOUND COLLECTION SYSTEM FOR A CROWDSOURCING APPROACH

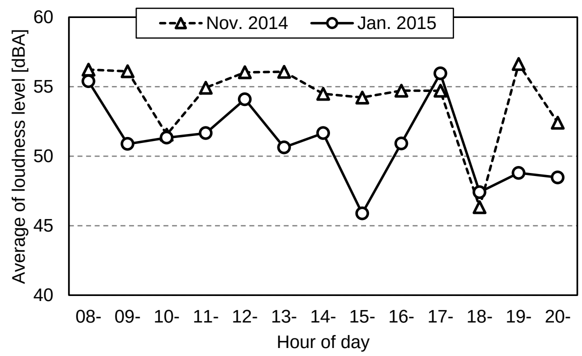
5.1. Functionality of crowdsourcing applications

The system statistically processes loudness data with spatio-temporal indices; latitude, longitude, and time. The number, sum, and squared-sum of data for each index are calculated as a sufficient statistic of Gaussian distribution. The calculation is implemented using the Map-Reduce function of MongoDB to make scaling out the collection system easy. This scalability is important for successful crowdsourcing.

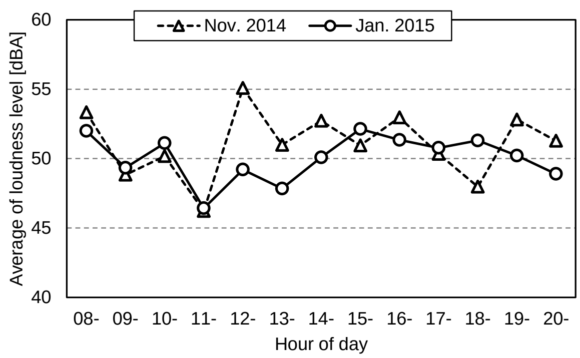
The statistics are updated on demand by uploading data from users. Users can see their contribution to the collection on our sound visualization map after a few minutes. The visualization of the contribution and the quick response are also important factors in crowdsourcing.



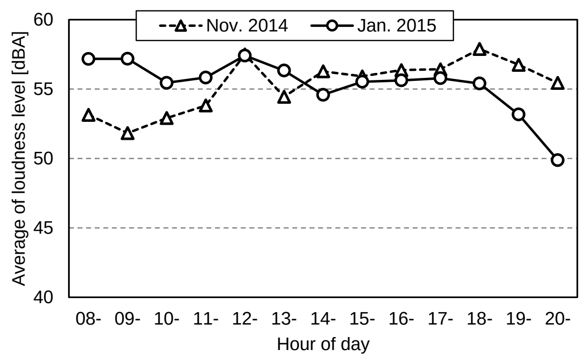
(a) Quiet residential area



(b) Shopping street far from a station



(c) Shopping street near a station



(d) Downtown area near a station

Fig. 4: Average of loudness level for hourly period

5.2. Visualization system for loudness and environmental sounds

A visualization system is implemented as a web application with the open-source libraries; Leaflet[4] and D3.js[5]. The system can visualize the data in the server system described in Section 3.

Visualization of the loudness data is provided through a color map of each area. The color index is calculated from the average loudness. We can overview the loudness distribution of any district of interest on the map. The color indicates the average of the loudness level; for example, red indicates a 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.

Sound visualization is achieved using icons symbolizing sounds on the map, enabling us to see the sound types in any district of interest. An example of environmental sound visualization is shown in Fig. 5. The sounds are distinguished by icons on the basis of their subjective evaluations during recording. An icon can be clicked to browse the associated sound's information and listen to it. The right side of the map interface shows the histogram of the sound types as statistics in the current viewing area.

[4]http://leafletjs.com/
[5]https://d3js.org/

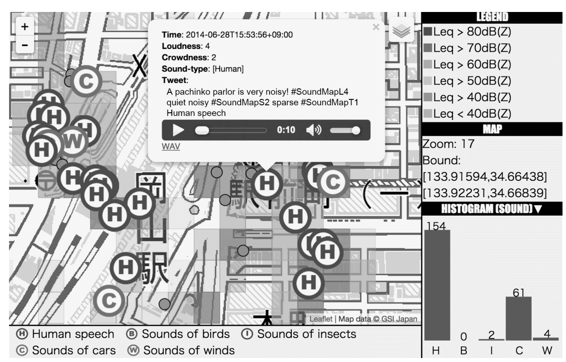


Fig. 5: Sound map visualizing sound type by icons

5.3. Data collection using crowdsourcing

All of the collected data are shown in Figure 6 as a histogram of the number of data. Huge numbers of data were recorded in November 2014, January 2015, June 2015 and November 2015. As a result of the data collection experiments described in Section 4, the number of data in November 2014 and January 2015 is substantial. We published the recording application in June 24, 2015, and consequently, the number of data in

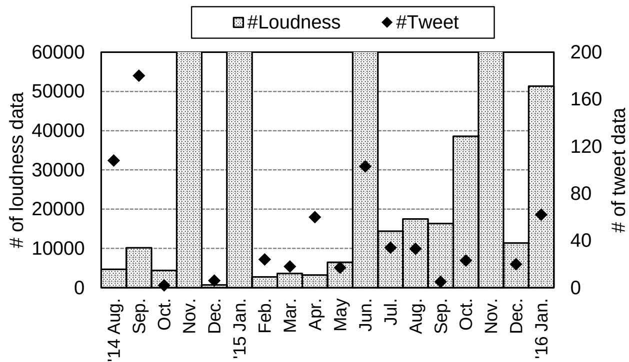


Fig. 6: Number of collected data

June 2015 is also large. We conducted a recording experiment with our application from November 28, 2015 to November 29, 2015, resulting in a large number of data in November 2015 as well.

6. CONCLUSION

In this paper, we developed a client-server application for collecting environmental sounds using smart devices, and we used the developed application to conduct a sound collection experiment with 14 participants. The collected data were analyzed for the distribution of loudness levels and sound types. In particular, we can find an effect of the opening of a large shopping mall in the city from time series charts. Finally, we attempted to collect sound data using crowdsourcing.

The effectiveness of the system has been demonstrated through the experiments, but there remains future work to be done. For example, the microphone specification must be appropriately calibrated, if it is to be used for more users. The calibration parameters of an unknown device might be estimated from the parameters of known devices and a small amount of data recorded by the unknown device.

7. REFERENCES

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

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

[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.

[5] S. Hara, M. Abe, and N. Sonehara, "Sound collection and visualization system enabled participatory and opportunistic sensing approaches," in *Proceedings of CASPer-2015*, Mar. 2015, pp. 390–395.

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

[7] 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.

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

[9] R. Rana, C. Chou, S. Kanhere, N. Bulusu, and W. Hu, "Ear-Phone: An end-to-end participatory urban noise mapping system," in *Proceedings of IPSN-2010*, Apr. 2010, pp. 105–116.

[10] 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.

[11] I. McGraw, C.-y. Lee, L. Hetherington, S. Seneff, and J. R. Glass, "Collecting voices from the cloud," in *Proceedings of LREC 2010*, May 2010, pp. 1576–1583.

[12] M. Matsuyama, R. Nisimura, H. Kawahara, J. Yamada, and T. Irino, "Development of a mobile application for crowdsourcing the data collection of environmental sounds," in *Human Interface and the Management of Information. Information and Knowledge Design and Evaluation*, S. Yamamoto, Ed., vol. 8521, pp. 514–524. Springer, 2014.

[13] 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, Apr. 2014.

[14] 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.