

Implementation of a sound-source localization method for calling frog in an outdoor environment using a wireless sensor network

Yasuharu Hirano*, Takuya Iwai*, Daichi Kominami†, Ikkyu Aihara‡, and Masayuki Murata*

*Graduate School of Information and Science, Osaka University, Osaka 565-0871 Japan

†Graduate School of Economics, Osaka University, Osaka 560-0043 Japan

‡Graduate School of Life and Medical Sciences, Doshisha University, Kyoto 610-0394 Japan

Abstract—Wireless sensor networks have a wide scope of applications, one of which is to observe and examine behavior of animals including insects, and amphibian. Our target is localization of small animals that emit a sound in an outdoor field by using a wireless sensor network. However, due to the resource limitations of sensor devices, most of existing localization methods are not compatible with such devices. In this paper, we implement a localization method using time difference of arrival (TDOA) measurements into wireless devices with a microphone and show the results of experiments. We focus on the interesting behavior of the Japanese tree frog and previously conducted some fieldwork. We also demonstrate the pitfalls in the implementation with resource-limited wireless devices for an outdoor environment.

Index Terms—Wireless sensor network, localization, Japanese tree frog, outdoor environment

I. INTRODUCTION

Mathematical models inspired by biological mechanisms help us to develop robust and adaptive systems in the ICT field [1]. In the background of these interdisciplinary research progress, a lot of studies of the mathematical modeling of biological systems have been performed thanks to the development of experimental techniques and a computer performance. The cooperative behavior with sociality emerging from autonomous motion controls of individuals is called *swarm intelligence* and there are lots of research that apply swarm intelligence to the network control [2].

It is important to observe how individuals communicate with each other for investigating the mechanisms of their behavior. For modeling their communication, exploring *when* and *where* individuals interact with each other is necessary. Thus, the identification of individual positions is important. However, to find animals in an outdoor environment is hard because they often are small and conceal themselves in the environment.

Many localization techniques have been proposed so far, but most of them are based on the assumption that a radio transmitter or receiver is directly mounted on the target animals [3], [4]. However, it is hard to put such a device on a target in advance in an outdoor environment. Then, we localize each animal based on the information that are detectable with some devices. One of such information obtained from their communication behavior is their calling. It is a natural idea to make a localization system that utilizes microphones to record their calling communication.

Outdoor environments make it difficult to deploy the localization system that consists of a large number of devices with wired connections. Therefore, we implement a localization method into a small number of wireless devices with a microphone for reducing the deployment cost. However, due to the resource (processing power, memory capacity, and

communication capacity) limitations of sensor devices, most of existing sound-source localization methods may not be compatible with the localization system based on wireless sensor networks. Most of existing sound-source localization methods are classified into two methods: AOA-based (angle of arrival) methods and TDOA-based (time difference of arrival) methods [5]. AOA-based localization methods estimate the sound-source position using microphones' position and the angle of the signal arrival at each microphones. To get AOA information, a microphone array is generally used. Microphone arrays have high accuracy, but they are expensive and comparatively large [6]. This increases the *deployment cost* from a monetary and carrying-task perspectives. TDOA-based methods estimate sound-source positions using microphones' position and the time differences of the sound arrivals between all pairs of two microphone nodes. These information can be easily obtained if each wireless sensor device has a microphone and a clock-timer. Since we place a great deal of importance on the deployment cost, we use TDOA-based localization methods.

In this paper, we implement a real-time sound-source localization system using wireless sensor devices with a microphone aiming at biological research of the Japanese tree frog. Previously, we conducted some fieldwork for revealing spatio-temporal structures inherent in frogs' calling communication. Then, we found that it was difficult to detect the position of frogs because they call from inside of grass or underground, but found that their calling was loud and continued for a long time. According to the feedback of the fieldwork, we design the localization system. We carry out experiments to clarify the accuracy of the estimated sound-source position under the outdoor environment. We also present the problems to occur with the implementation of the present sound-source localization methods for outdoor environments and present the solutions for them. Figure 1 shows an overview of our system, where wireless sensor devices with a microphone record sounds emitted from a sound source and transmit the sound data to a lap-top computer. The lap-top computer calculates the time differences of the sound arrivals between all devices and then estimates the sound-source position.

II. SYSTEM REQUIREMENTS

A. Characteristics of the Japanese tree frog

Japanese tree frogs have unique and interesting characteristics. Male Japanese tree frogs vocalize *advertisement calls* early at night to inform their existence to a conspecific female (Fig. 2). When one frog begins to call, other frogs that hear it also begin to call following the frog (chorus). The

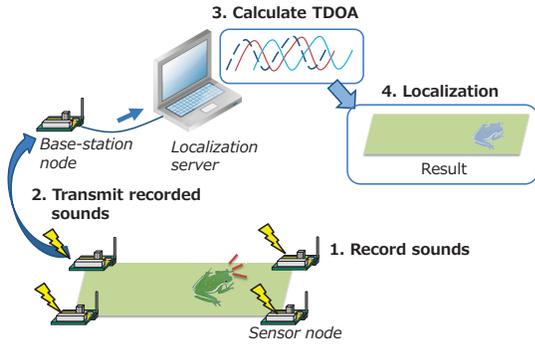


Fig. 1. Summary of the implemented system



Fig. 2. Japanese tree frog calling at the ridge of rice paddies

chorus of a few frogs synchronizes in anti-phase so that their calling does not overlap. Their chorus behavior is considered to be for letting female frogs distinguish them individually.

The body length of Japanese tree frogs is 2.0–4.5 cm. They inhabit rice paddies or forests and their positions are sparsely-distributed. They do not call under the water and do not move while it calls. Once the Japanese tree frog begins a chorus, it often continues more than several minutes. The chorus of Japanese tree frogs can be observed in rice paddies in the spring of Japan, around which there are not the tall trees, but are growing thick grass.

The ridge of the rice paddies is muddy, and most part of it is not flat. Therefore, it is desirable to use a light and small device. The authors of [7] implement a sound-source localization system for the tree frogs using a special device with a light-emitting diode (LED) that turns on response to a nearby sound. This system can acquire the position of frogs with a precision of about 10 cm since in this system, respective devices are deployed in the observation area at 10 cm intervals, which takes comparatively much time to prepare and manage the system. Our final goal is to achieve the precision comparable to that of [7] in real-time by a less number of wireless sensor devices with a microphone.

III. TDOA-BASED SOUND-SOURCE LOCALIZATION

A. Overview

TDOA is a time difference of sound arrivals between all pairs of microphones. TDOA is obtained by phase-difference measurements as shown in Fig. 3. The possible positions of a sound source are obtained as two hyperbolas from the TDOA between two microphones. Then, the intersection that all hyperbolas obtained from all sets of microphones meet is the estimated position of the sound source. However, due to the errors of the timer in a sensor node as well as various environmental noises, all hyperbolic curves do not intersect on

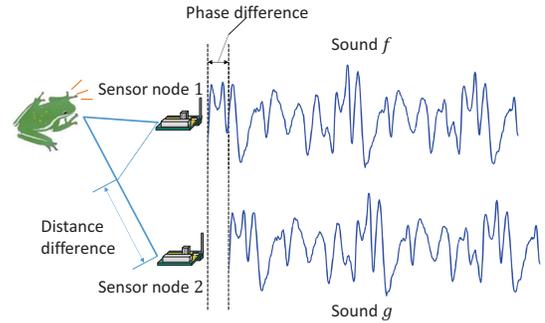


Fig. 3. Phase difference of sounds

the same point. Therefore, to determine the estimated location of the sound source, optimization methods are generally used, e.g., the least squares method. In the following subsections, we describe how to estimate the TDOA between two microphones and how to estimate the sound-source position using the TDOA in our implementation.

B. TDOA estimation using a cross-correlation method

For the localization of a sound source in the 2D-plane, the positions of more than two microphones and TDOAs between more than one pairs of microphones are required. Although the arrival times of sounds from a sound source to microphones differ according to the distance between them, the sound waveform observed in each node is very similar if the influence of noises and sound echoes are small enough. In other words, if each sensor node's timer is synchronized, we can get the TDOA between two sensor nodes from the phase difference of the observed sounds (Fig. 3).

As a general method for calculating the phase difference of two signals, the cross-correlation function is computed at each unit-time lag. The cross-correlation function returns a value that indicates how similar two signals are to each other when one signal is shifted by a lag (denoted by n). When the function has the maximum value with a lag of n^* , n^* is regarded as the phase difference of two signals. Equation (1) shows the cross-correlation function for the discrete two sound signals f and g observed in two sensor nodes.

$$R_n^{(fg)} = \frac{\sum_{i=0}^{N-1} f_i g_{i+n}}{\sqrt{\sum_{i=0}^{N-1} f_i^2} \sqrt{\sum_{i=0}^{N-1} g_{i+n}^2}} \quad (|n| \leq N), \quad (1)$$

where f and g are signal sequences, which correspond to $\{f_0, f_1, \dots, f_{N-1}\}$ and $\{g_0, g_1, \dots, g_{N-1}\}$ ($|f_k| \leq 1$ and $|g_k| \leq 1$ for any k), respectively. In addition, f_k and g_k are zero when $k \leq 0$ or $k \geq N$. The phase difference of two signals denoted by n^* is defined by using $R_n^{(fg)}$ as following:

$$n^* = \arg \max_n (R_n^{(fg)}). \quad (2)$$

Finally, the TDOA is calculated as $n^* \times R_s$, where R_s is a time period for the unit-time lag. It is a premise in Equation (2) that signals can be acquired at regular intervals without loss. However, this premise is not always satisfied. In this case, we interpolate these data using the linear interpolation.

C. Position estimation using TDOA

As discussed above, due to the timer errors of sensor nodes and various environmental noises, the estimated position is underspecified. In this paper, we use the approximate technique proposed in [8]. Reference [8] represents a hyperbolic curve

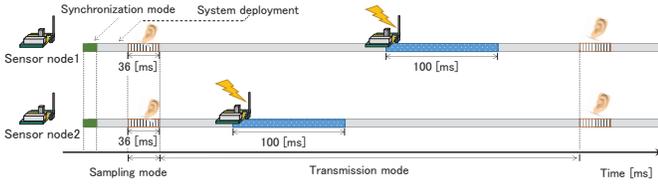


Fig. 4. Time sequence of sensor node's behavior

obtained from a pair of two microphones (this pair is denoted by m) on the $x - y$ plane as $f^m(x, y) = 0$. It defines a cost function $J = \sum_{m \in M} |f^m(p, q)|^2$ for an estimate of the error between the true and estimated positions, where M means a set of all pairs of microphones, and p and q are coordinates. The estimated position is chosen so that J is minimized. The computational complexity of this technique is $O(N^2)$ for the number of the nodes (denoted by N). This enables a real-time estimation of the sound-source position.

IV. LOCALIZATION SYSTEM USING IRIS MOTE

A. System design

In our localization system, there are three types of devices: sensor nodes, a base-station node, and a localization server (Fig. 1). A sensor node is responsible for recording sounds and sending the recorded sounds to the base-station node. For that, sensor nodes are deployed to cover a target area and they should be within the communication range of the base-station node. The base-station node, which is connected to the localization server, transmits the received sound data to the localization server with serial communication. The localization server interpolates data, calculates the cross-correlation function, and estimates the position of the sound source.

B. Devices

We use IRIS Motes as a *sensor node*. An IRIS is widely used in the field of the wireless sensor network. Its clock cycle is 8 MHz and flash memory size is 8 KByte. An IRIS has the RF230 chip as a wireless communication interface and its transmission rate is 250 kbps. For the communication protocol, we use the ZigBee protocol. TinyOS is installed in our IRIS, which is a free open-source software and is the platform targeting wireless sensor networks. For the position estimation, we use an ASUS ZENBOOK UX31A whose CPU is Core i7 3517U and memory size is 4 GByte. We call this laptop a *localization server*. We prepare an IRIS as a *base-station node* that has a serial connection to the localization server. Since this laptop cannot understand the ZigBee protocol, this node is necessary (details are below). For recording sounds, we use the MTS310 sensor board. The AD converter of the MTS310 can get a 10-bit information for the amplitude of a sound at a certain point.

The sampling theorem said that the original sound wave can be perfectly reconstructed from its samples if the sampling rate is set as twice of the frequency of the original sound. The fundamental frequency of the advertisement calls of the Japanese tree frog is 2,000 Hz. Since the CPU clock frequency of the IRIS is 8 MHz, it is possible to sample the sound of the Japanese frog sound with 4,000 Hz.

C. System implementation

1) *Implementation of sensor nodes*: Sensor nodes have three modes of operations: the synchronization mode, the sampling mode, and the transmission mode. In the synchronization mode, all sensor nodes communicate with the base-station

node to synchronize their clock timer. In the sampling mode, sensor nodes store voltage values that their microphone outputs and also store the time of sampling. This task is conducted at the same time among all sensor nodes. In the transmission mode, they transmit the recorded sounds to the base-station node. At this time, sensor nodes transmit packets in different time slot so that packets from sensor nodes do not collide in the base-station node. When the transmission of the recorded sounds is completed, sensor nodes change their mode to the sampling mode again. Figure 4 shows an example of the behavior of two nodes that change their mode. The details of the three modes are described in the following paragraphs.

a) *Synchronization mode*: In our implementation, sensor nodes take two types of synchronization methods, one is roughly accurate synchronization and the other is highly accurate calibration. The former is for starting to sample sounds at the roughly same timing among all sensor nodes. And the latter is for the estimation of the source position.

For the rough synchronization, the base-station node sends its current time and all sensor nodes set their own time to it. This rough synchronization is done only at the beginning of the system operation to prevent sensor nodes from spending their computation and communication resources. Then, the calculated TDOA includes the sum of the time difference of sound arrivals and the error of the system time of sensor nodes. Therefore, it is essential to remove the system time error from the calculated TDOA for estimating the source position. To do so, after the rough synchronization, we gather sensor nodes in one place so that the time when a sound comes in each sensor node is the same. Then, we make hand clappings and record their sound. Now, the TDOA of a pair of sensor nodes for the hand-clapping sounds only contains the difference between their system time. The localization server stores these time differences, and when it estimates the sound-source position, it adds offsets to obtained TDOAs. Note that clapping sounds include various frequencies in a relatively short time, and therefore, they are suitable for time calibration. After deploying sensor nodes, these sensor nodes commence the sampling mode.

b) *Sampling mode*: In the sampling mode, sensor nodes store the sets of a voltage value that their microphone outputs and the time of sampling the value. After storing a fixed number of the sets, sensor nodes start the transmission mode. In TinyOS, the maximum size of the data field in one packet is 114 Byte. Because one sample consists of sound data (the voltage value) of 2 Byte and time data (the time of sampling) of 4 Byte, a sensor node can transmit at most 19 samples in a packet. In our implementation, each packet includes 10 samples and a sensor node transmits 8 packets to the base-station node in the transmission mode, which are experimentally determined. In other words, sensor nodes accumulate 80 samples and transmit them to the base-station node.

c) *Transmission mode*: In the transmission mode, sensor nodes transmit the data to the base-station node. Since the transmission speed of an IRIS is a little low to transmit sounds as streaming data, it transmits a certain number of samples (80 samples in our implementation). Each sensor node's transmission timing differs by 100 ms so that the packet collision among sensor nodes does not occur. The transmission schedule of nodes are determined by the base-station node. In our system, with consideration for the number of sensor nodes, each sensor node returns to the sampling mode after 500 ms

of the transmission mode.

2) *Implementation of the base-station node:* The base-station node informs its own current system time to sensor nodes at the beginning of the synchronization mode. After that, each sensor node sets its system time to the time contained in the received information. This is conducted for the rough synchronization of all sensor nodes. The base-station node also determines the time slot of each node to transmit samples in the transmission mode. In addition, the base-station node has a role to transmit the received sounds from sensor nodes to the localization server via a universal serial bus (USB) connection. This is because IRIS Motes and the localization server share no communication protocol in common.

3) *Implementation of the localization server:* When the localization server receives the data that each sensor node sampled for the same period from the base-station node, it checks whether there is not data loss and the data has enough volume sounds. If the volume is large enough and data loss is above the acceptable level, the localization server interpolates data. Then, it calculates the cross-correlation function for each pair of two sensor nodes to obtain a TDOA. Finally, it estimates the source position from obtained TDOAs.

D. Problems and solutions

Here, we show the problems with our implementation and our solutions for them. These are helpful tips for the implementation of sound-based localization systems.

1) *Non-parallel execution of sampling and wireless communication:* A sensor node does not sample sounds while transmitting the recorded sounds to the base-station node. Wireless communication generates an interrupt to the CPU and a critical section of any process occupies the CPU resource. These decrease the sampling rate or cause sampling jitters. Thus, sensor nodes store 80 consecutive samples and then transmit them in the transmission mode.

2) *Sampling-rate limitation and fluctuation:* When using an interrupt of sensor's timer to sample sounds, its sampling rate depends on the minimum time between interrupts. In the TinyOS, the timer interrupt cycle is more than milliseconds (less than 1,000-Hz sampling). Therefore, we implement the sound sampling by the program loop. Then the minimum interval of sampling is about 450 μ s in the system time. This means that the maximum sampling rate is about 2,200 Hz. In fact, the fluctuations of the sampling interval occur (from 403 μ s to 545 μ s). We use an interpolation method with 1- μ s unit-time lag for dealing with this jitter.

3) *Low clock accuracy:* One second in the timer of a sensor node does not correspond to actual one second. Main reason of this is the precision of the CPU's clock in an IRIS mote is comparatively low. The localization server, which has a high accurate clock timer, corrects this gap by preliminarily finding out the ratio of actual one second to the devices' one second.

4) *Design of sampling periods:* We calculate the cross-correlation function for two sound signals that are observed in two sensor nodes. To calculate the cross-correlation function, both signals need to include the same part of the wave form emitted by the source. On the other hand, too long sampling periods cause longer time to transmit the recorded sounds to the base-station node, which evokes packet losses. We carefully choose the number of samples in the sampling mode for our experiment.

5) *Network congestion and packet loss:* As the number of samples obtained in the sampling mode increases because of the increase in sampling rate or the increase in the number of sensor nodes, the amount of data transmitted in the transmission mode increases. To avoid congestion caused by such transmission, these are decided in consideration of the sensor node's transmission speed and the amount of sample data. Additionally, if all sensor nodes send data at the same timing, packet loss can occur due to the packet collisions. Thus, sensor nodes transmit the recorded sound to the base-station node at the different timing for avoiding packet collisions in the transmission mode.

6) *Narrow sound-collecting range of a microphone:* To obtain an accurate cross-correlation function, a sufficient signal-to-noise ratio (SNR) is required. However, if the sound collecting range of the microphone is narrow, such a sufficient SNR cannot be obtained. Preliminarily, we play a frog calling sound by a loudspeaker. The range that the microphone of MTS310 can collect the sound is about 3 m. This is the largest bottleneck in the scalability of our system. Thus, microphones are necessary to have a sufficient amplifier.

7) *Inaccuracy in positions of sensor nodes:* For the calculation of a TDOA, accurate positions of sensor nodes are required. Since we manually deploy sensor nodes in the experimental field after the synchronization mode, it is difficult to obtain the accurate positions of sensor nodes in an actual outdoor field. In our experiments, sensor nodes are placed at the corners of the predetermined square area. Although there are many techniques for self-localization of sensor nodes, applying these techniques is our future work.

8) *Estimation error:* There is an error between an estimated position provided in the localization server and the actual sound-source position. To reduce the error of the sound-source localization, the estimation of the position is conducted multiple times and we regard the center of gravity of all estimated positions as the conclusive estimated result. To this end, we regard the set of samples obtained in one sampling mode as a *sample block*, and estimate a position with each sample block.

V. EXPERIMENTAL RESULTS

We examined the estimation accuracy of our system. In a 1.8 m \times 1.8 m outdoor square area, we set four sensor nodes on the each corner of the area and one sound source (loudspeaker) in the area. The loudspeaker played three types of sounds: artificial sounds that had fundamental frequency of 500 Hz or 2,000 Hz, and advertisement sounds of the Japanese tree frog recorded in the indoor room. Here, the fundamental frequency of the sound of the Japanese tree frog is about 2,000 Hz. For the artificial sounds, the loudspeaker repeated sounds that included a 0.3-second sound and a 0.02-second pause as a Japanese tree frog calls. The position of the loudspeaker was set on the four positions P1 (0.225, 0.225), P2 (0.225, 0.675), P3 (0.675, 0.675), or P4 (0.900, 0.900).

We evaluated an absolute error between the true position and the conclusive estimated position, and a normalized root-mean-square error (NRMSE). NRMSE is defined as $\sqrt{\sum_{s \in S} (X_{true} - X_{est}^s)^2 + \sum_{s \in S} (Y_{true} - Y_{est}^s)^2} / A$, where (X_{true}, Y_{true}) is the true position of the sound source, (X_{est}^s, Y_{est}^s) is the estimated position in sample block s , and S is the set of sample blocks. A is the one side length of the observation area (1.8 m).

TABLE I
RESULTS OF ESTIMATION

Sound source	Sound-source position [m]	Estimated position [m]	Error [m]	NRMSE [m]	Example: 500 Hz (P2)
500 Hz	P1 (0.225, 0.225)	(0.823, 0.858)	0.871	0.632	
	P2 (0.225, 0.675)	(0.488, 0.724)	0.268	0.371	
	P3 (0.675, 0.675)	(0.815, 0.765)	0.167	0.496	
	P4 (0.900, 0.900)	(0.827, 0.743)	0.173	0.216	
2,000 Hz	P1 (0.225, 0.225)	(0.590, 0.528)	0.474	0.383	
	P2 (0.225, 0.675)	(0.810, 0.902)	0.627	0.560	
	P3 (0.675, 0.675)	(0.815, 0.765)	0.167	0.305	
	P4 (0.900, 0.900)	(0.827, 0.897)	0.073	0.275	
Japanese tree frog	P1 (0.225, 0.225)	(0.567, 0.534)	0.461	0.547	
	P2 (0.225, 0.675)	(0.552, 0.748)	0.336	0.475	
	P3 (0.675, 0.675)	(0.694, 0.851)	0.178	0.456	
	P4 (0.900, 0.900)	(0.837, 0.820)	0.101	0.343	

A. Results

We carried out the experiment in the quiet outdoors. There was a faint noise produced by insects and wind. The outdoor air temperature was 11 degrees Celsius, so the sonic speed was 338.21 m/s. In this paper, the position of the loudspeaker means the position of its diaphragm and the position of sensor nodes means the position of their microphone. Localization computation for one sample block in the localization server took about 0.3 s.

Table I shows the estimation results. At first, an estimated position of the Japanese tree frog is with an error between 10.1 cm to 46.1 cm. Since we set out to achieve an error within about 10 cm, this system is still insufficient in terms of the accuracy. Comparing the NRMSEs with the 500 Hz sound and the 2,000 Hz sound, there is no remarkable difference. We investigate the causes of errors in the TDOA calculation with these sounds. The sampling rate of sensor nodes (about 2,200 Hz) is not enough for restoring the 2,000 Hz original sounds. In addition, since both sound signals repeat the same waveform, the cross correlation functions with them have multiple peaks.

In most of the results, estimation errors are more smaller when the sound source is nearer the center of the area. This is because as a sensor node is farther from the sound source, the SNR of received sound signal is smaller, which results in incorrect TDOAs. Some results are subject to echoes and environmental noises and show a bad accuracy (in particular, the result of 500 Hz (P2)). In the table, we show a figure that represents estimated positions in sample blocks (result (single)) and the conclusive estimated position (result (average)).

On the scalability of our system, we made a simulation where the field size was 20 m \times 20 m. In the simulation, a random value generated by the normal distribution with the average of 0 ms and the standard deviation of 2.0 ms was added to the true TDOA. This value of the standard deviation was measured in our experiment (500 Hz (P4)). When a sound source was installed at (2.5, 2.5), (2.5, 7.5), (7.5, 7.5), or (10, 10), absolute error was 0.070 m, 0.050 m, 0.084 m, 0.541 m, respectively. This result indicates that the estimation errors do not differ from those in the result of our experiment even though the field size decuples.

Due to the limitation of the sensor-node processing capability, the sampling rate is at least 2,200 Hz and this is not enough to restore 2,000-Hz artificial sounds and the original sounds of the Japanese tree frog calling. This causes an error in the calculated TDOA. Meanwhile, a highly frequent sampling rate causes big traffic, which occupies the network bandwidth. Another problem is that the MTS310's microphone chip has a narrow recording range. Then, high-end microphones or more sensor nodes are necessary for the wide area observation.

However, an increase in the number of sensor nodes occupies the more wireless channel. Therefore, in that case, it is important to choose an appropriate sampling rate and the number of sensor nodes considering the transmission rate and the target sound's frequency.

VI. CONCLUSION

In this paper, we implemented a sound source estimation method using a wireless microphone-sensor network. We also showed the problems faced with the implementation of the localization system into a wireless sensor network for an outdoor environment and showed the solutions for them. Experimental results of localization presented that our system can estimate the Japanese tree frog's position with an error of 10.1–46.1 cm. Since our goal is an accuracy less than 10 cm, improvement in accuracy is needed. Through the experiments, we found a clue to the improvement. To reduce the error, it is necessary to reduce both errors in the TDOA calculation and the position estimation. The error in the cross-correlation calculation was caused by the low sampling rate, the environmental and device-internal noises, and the time synchronization error. These will be improved by using sensor nodes that have more rich resources and the high-end microphones. On the error in estimation of the source position, a larger number of sensor nodes can improve it, which we validated by simulation. Our future work is to complete the system for more accuracy and for multiple sound sources for the actual observation.

ACKNOWLEDGMENT

This research was supported in part by "Grant-in-Aid for Scientific Research (A) 15H01682."

REFERENCES

- [1] S. Barbarossa and G. Scutari, "Bio-inspired sensor network design," *IEEE Signal Processing Magazine*, vol. 24, pp. 26–35, May 2007.
- [2] F. Dressler and O. B. Akan, "A survey on bio-inspired networking," *Computer Networks*, vol. 54, pp. 881–900, Oct. 2010.
- [3] A. Wilson, M. Wikelski, R. Wilson, and S. Cooke, "Utility of biological sensor tags in animal conservation," *Conservation Biology*, Mar. 2015.
- [4] R. E. Floyd, "RFID in animal-tracking applications," *Potentials, IEEE*, vol. 34, pp. 32–33, Oct. 2015.
- [5] Z. M. Saric, D. D. Kukolj, and N. D. Teslic, "Acoustic source localization in wireless sensor network," *Circuits, Systems and Signal Processing*, vol. 29, pp. 837–856, Oct. 2010.
- [6] J.-M. Valin, F. Michaud, J. Rouat, and D. Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1228–1233, 2003.
- [7] T. Mizumoto, I. Aihara, T. Otsuka, R. Takeda, K. Aihara, and H. G. Okuno, "Sound imaging of nocturnal animal calls in their natural habitat," *Journal of Comparative Physiology A*, vol. 197, no. 9, pp. 915–921, 2011.
- [8] A. Canclini, E. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, pp. 439–443, Feb. 2013.