J. Inf. Technol. Appl. Manag. 24(1): 157~167, March 2017
https://doi.org/10.21219/jitam.2017.24.1.157

ISSN 1598-6284 (Print)
ISSN 2508-1209 (Online)

Fast 360° Sound Source Localization using Signal Energies and Partial Cross Correlation for TDOA Computation

Mariam Yiwere* · Eun Joo Rhee**

Abstract

This paper proposes a simple sound source localization (SSL) method based on signal energies comparison and partial cross correlation for TDOA computation. Many sound source localization methods include multiple TDOA computations in order to eliminate front-back confusion. Multiple TDOA computations however increase the methods' computation times which need to be as minimal as possible for real-time applications. Our aim in this paper is to achieve the same results of localization using fewer computations. Using three microphones, we first compare signal energies to predict which quadrant the sound source is in, and then we use partial cross correlation to estimate the TDOA value before computing the azimuth value. Also, we apply a threshold value to reinforce our prediction method. Our experimental results show that the proposed method has less computation time; spending approximately 30% less time than previous three microphone methods.

Keywords : Partial Cross Correlation, Signal Energy, Time Difference of Arrival, Front-Back Confusion

Received : 2016. 12. 03. Revised : 2017. 03. 17. Final Acceptance : 2017. 03. 17.

* Department of Computer Engineering, Hanbat National University, e-mail : myiwere@yahoo.co.uk
** Corresponding Author, Department of Computer Engineering, Hanbat National University, 125-Dongseo-daero, Yuseong-gu, Daejeon, Korea, 34158, Tel : +82-42-821-1205, e-mail : ejrhee@hanbat.ac.kr

1. Introduction

1.1 Background and Objective

Sound source localization (SSL) is the process of estimating the direction of a sound source. A system consisting of a microphone array and SSL algorithm captures a multiple channel sound signal and then computes the signal's angle of incidence with respect to the system's reference point of 0°. It is useful in various fields including human-robot interaction where the robot is able to determine the position of a speaker, video surveillance where the surveillance camera automatically rotates when an event occurs outside its field of view, hearing aid systems, lecture archiving, video conferencing, etc.

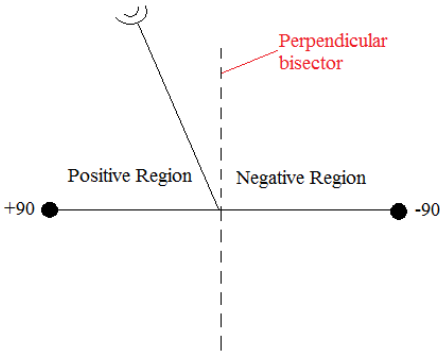
It involves capturing an audio signal using a microphone array which consists of at least two microphones, and then computing the time delay between the signals received at the different microphones. This time delay value is in turn used to compute the signal's angle of incidence using basic trigonometric functions. To compute the time delay value between two signals, a cross correlation method is usually implemented; for example the generalized cross correlation (GCC) [Knapp and Carter, 1976], or a time domain implementation [Murray et al., 2004; Broeck et al., 2012; Yiwere and Rhee, 2015] of cross correlation.

In using only two microphones, there is a difficulty in determining whether the sound source is in front of or behind the system. This is generally known as the front-back confusion. For effective use of any sound source localization method, the problem of front-back confusion must be eliminated. For example, in human-robot communication and hearing aid systems, it is important for the algorithm to have the ability to distinguish between front signals and back signals in order to help the robot or user to easily identify the location of a speaker.

1.2 Related Work

Sound source localization has been widely researched over the past few decades to solve the problem of front-back confusion, to improve upon the accuracy of time delay estimation as well as angle estimation and also to increase the speed of localization. The different sound source localization methods can be categorized as follows [Tashev, 2009] : The Eigen Value based methods, which implement complex Eigen value decompositions; the Steered Power Response based methods, which use large numbers of microphones together with complex rotations; and the Time Difference of Arrival (TDOA) based methods, which use interaural time difference of two signals. TDOA methods are similar to the mechanism in the mammalian auditory cortex [McAlpine and Grothe, 2003]; they use fewer microphones and therefore less expensive to implement.

The TDOA methods can be implemented with at least two microphones and cross correlation implementation. They are relatively simpler to implement. When only two microphones are used, the above mentioned front-back confusion problem occurs. In order to eliminate the front-back confusion problem and to achieve a 360°



〈Figure 1〉 Source Localization with 2 Microphones

range localization, researchers have proposed various methods [Lee et al., 2005; Kwon et al., 2007; Li et al., 2012; Sreejith et al., 2015; Yiwere and Rhee, 2016] using at least three microphones.

Ji Yeoun Lee et al. used signal energy and zero crossing rates to select one microphone pair in the correct region for the TDOA estimation and the subsequent azimuth estimation [Lee et al., 2005]. With this approach, they compute only one TDOA value which saves computation time; however, the presence of noise can influence the section decision in a negative way, leading to incorrect azimuth estimation.

Byoungho Kwon et al. used the conventional three microphone method in their first approach [Kwon et al., 2007]. Using three microphones, they compute the cross-correlation values for all the three microphone pairs and then they sum up all the converted cross correlation values to determine the sound direction. In this method, using three TDOA computations first of all eliminates the front-back confusion and also it improves the accuracy of the azimuth computation; however, three computations of TDOA have the drawback of increasing algorithm's total computation time.

Xiaofei Li et al. proposed a method based on guided spectral-temporal position method using four microphones arranged in a cross manner [Li et al., 2012]. They used the methods of spectral subtraction and campestral mean normalization to remove noise from the input signal prior to the TDOA estimation. Their method focuses on reducing the effect of reverberation in order to improve accuracy; however the use of four microphones increases the hardware cost of the system.

Taking advantage of the Sign of Time Delay Estimate (STDE) values, Sreejith T. M. et al. proposed a Y-shaped microphone array arrangement for sound tracking [Sreejith et al., 2015]. They created nine regions based on the voroni regions of the different microphone pairs. Based on the STDEs computed for all six microphone pairs, they estimated which voroni region the sound source is in. The use of the nine voroni regions helps in the speaker tracking and home-region detection; however using four microphones increases the hardware cost of the system and also computing six different TDOA values increases the computation time.

M. Yiwere and E. J. Rhee proposed a TDOA Sign-based sound source localization method using an L-Shaped microphone array [Yiwere and Rhee, 2016]. They used the signs of TDOA values to determine which region the sound source is in. By focusing on two microphone pairs, they estimated only two TDOA values, thereby reducing the total computation time compared to other previous methods. This method saves some computation time by using fewer TDOA computations; however, TDOA Sign comparison and other checks do take up some extra computation time.

Also some researchers have proposed the use of Head Related Transfer Functions (HTRF) to achieve 360° localization using only two microphones [Hwang et al., 2005; Usagawa et al., 2011; Alan-Boyd et al., 2015]. For example Sungmok Hwang et al. [Hwang et al., 2005] proposed the use HRTF for sound source localization on a robot platform. They make use of the phase and magnitude information present in the HRTF database by introducing certain criteria to be satisfied by a set of signals from the microphones. But the presence of noise in any real environment can easily affect the magnitudes of real-time signals which may mislead the algorithm to make a wrong estimation of the sound source location.

1.3 Suggestion

Through our survey of previous related works, we found that most existing sound source localization methods involve too many avoidable computations due to multiple TDOA estimations. Also some methods use more than three microphones which increase the hardware costs as explained above. The main reason for this is the use of multiple microphone pairs which are actually necessary in order to solve the front-back confusion (i.e. to achieve 360° localization). Although these previous methods are successful in localizing sound sources, there is a need for their computational costs to be reduced so that such methods can be implemented on small or embedded systems.

In order to overcome the above mentioned weaknesses by achieving 360° localization using fewer computations, we suggest incorporating signal energy comparison into the process of sound source localization [Yiwere and Rhee, 2015]. To further reduce the amount of computations, we also implement a partial cross correlation method in time domain [Yiwere and Rhee, 2015] instead of the conventional cross correlation method for TDOA estimation. We do not perform TDOA estimation for all microphone pairs.

Firstly, we predict the direction (i.e. left or right) from which sound originates by comparing the energies of the signals received at left and right microphones, and then we compute the TDOA value using the partial cross correlation method. Depending on the TDOA value, we use one of two methods to determine whether the signal is in front of or behind the microphone array. Finally we compute the azimuth value using the delay value from the partial cross correlation.

The organization of this paper is as follows; Initial source direction prediction and partial cross correlation are described in sections 2 and 3 respectively, section 4 describes the azimuth computation, experiment is presented in section 5 and section 6 presents conclusion.

2. Initial Source Direction Prediction

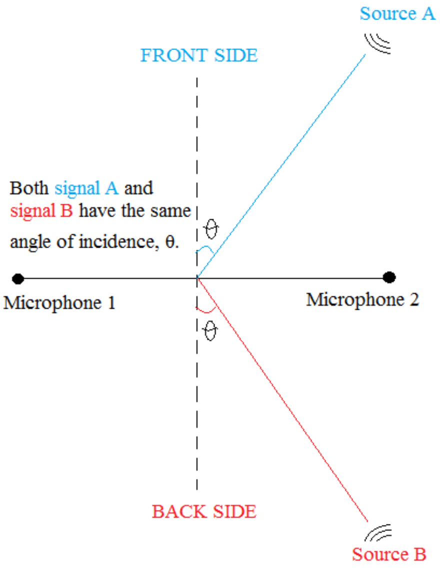
Time difference of arrival is simply the difference between the times a signal from a single source reaches two different sensors-microphones in the case of sound source localization.

The distance between a pair of microphones is divided equally by a perpendicular bisector at the point where the TDOA (delay) value is zero. All TDOA values at one side of the bisector are positive values while all TDOA values at the other side of the bisector are negative values [Yiwere and Rhee, 2015], as shown in <Figure 1>. Also the microphone on the side where the sound source is, usually records the higher energy signal. Taking advantage of this, we first predict which side of the perpendicular bisector the sound is originating from by comparing the left and right signal energies. This predicted direction is later used in the partial cross correlation implementation.

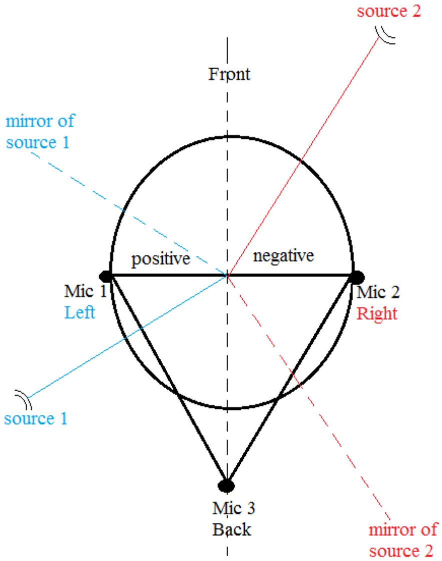
<Figure 2> illustrates the front-back confusion problem which is solved next using one of two methods. In the figure, sound source A and sound source B are at the front and back side of the microphone array set-up respectively. However both signals are incident at the same angle theta with respect to the perpendicular bisector. This implies that a two-microphone array system will have a difficulty in determining whether any sound source is in front of or behind it.

To set up our microphone array, we position three microphones-left, right and back microphones-at an equal distance apart as shown in <Figure 3>. By making the assumption that most of the sound signals are located in front of the system (e.g. in front of a robot), we focus on microphones 1 and 2 as the primary microphone pair, using the back microphone-microphone 3-to distinguish the back signals from the front signals.

With regards to the primary microphone pair, sound sources 1 and 2 as shown in the <Figure 3> both have a mirror sound source; that is, for each of the two sound sources, there is another position on the other side of the system that has equal TDOA and azimuth values. In order to tell the difference between sources 1 and 2 and their respective mirror sources, the back microphone's



〈Figure 2〉 Front-Back Confusion Illustration



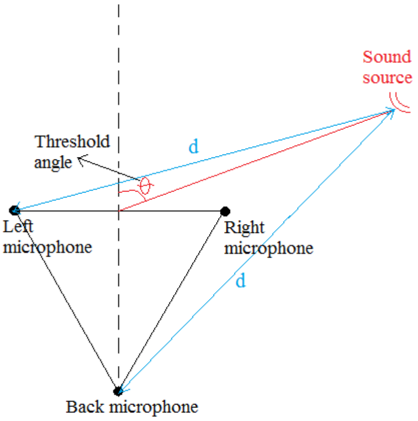
〈Figure 3〉 Three Microphone Array Set-Up signal is employed.

First the energies of all three signals are estimated using equation 1. The left and right signal energies are compared and the direction of the higher energy signal is selected for the partial cross correlation implementation. In the next step, we compare the signal with lower energy to the back signal's energy. If the back signal's energy is higher, then source is behind the system, else the sound source is in front of the system.

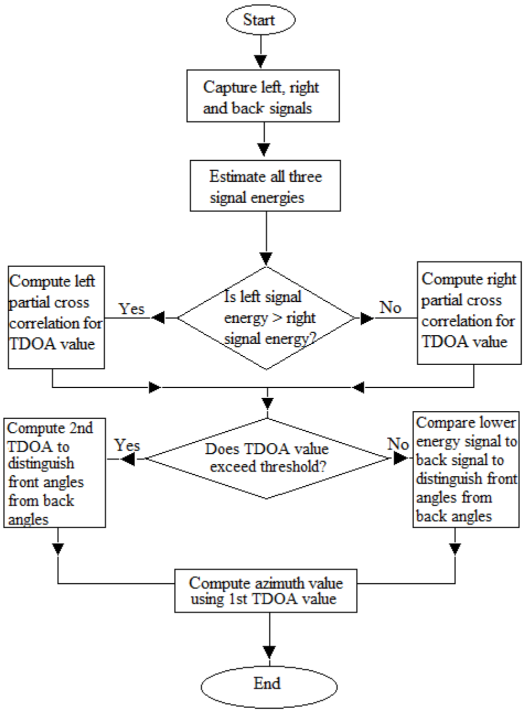
$$Energy\ of\,X: E_x = \Sigma_{i=0}^{N-1}|x_i| \qquad (1)$$

where N is the signal length and x_i is the samples of signals.

Beyond a certain point, the comparison between the lower energy signal and the back signal become ineffective. This is because the two energies become comparable to each other as shown in <Figure 4>. At this point, the distances between the sound source and the two microphones in question become almost equal and this causes their energy estimates to be very similar or become the reverse of our expectation.



<Figure 4> Illustration of Threshold Point



<Figure 5> Flow Diagram for the Proposed Method

In order to solve this problem, the delay value from the partial correlation is first compared to a threshold value. If the delay does not exceed the threshold value, we compare the lower energy signal to the back signal. But if the delay value exceeds the threshold, we employ a different method to decide whether the signal is in front of or behind the microphone set-up. We perform another TDOA computation using the back signal and the signal with the lower energy. This time, we perform a full cross correlation in order to take advantage of the sign of the TDOA values. If the TDOA value has a negative sign, the sound source is behind the microphone array; and if the value has a positive sign, the sound source is in front of the set up. <Figure 5> shows a flow diagram of our proposed method described above.

3. Partial Cross Correlation

In order to speed up the TDOA estimation time, we use the partial cross correlation method, equation 2. Frist we compute the relevant range of delay values, which includes positive and negative lags. Using the distance between the microphones (d), the sampling frequency rate (f) and the velocity of sound (v), we determine the minimum and maximum delay in samples as follows; the maximum delay (τ) that can be estimated is df/v, and the minimal delay is −df/v as shown in equation 2.

$$\text{Range} = [\min\tau, \max\tau] = \left[\frac{-df}{v}, \frac{df}{v}\right] \quad (2)$$

Then we use the energies of the signals to predict which part of the correlation function will contain the TDOA value, and then we compute the cross correlation in the predicted direction using equation 3. This saves a lot of computation time because only a few relevant correlation coefficients are computed.

$$\text{CrossCorr}(x, y)(j) = \begin{cases} \sum_{k=0}^{N-1-j} x_{(k+j)} \cdot y_{(k)}, & E_{x(left)} < E_{y(right)} \\ \sum_{k=0}^{N-1+j} x_{(k)} \cdot y_{(k-j)}, & E_{x(left)} > E_{y(right)} \\ 0 & Else \end{cases}$$

$$\begin{aligned} &\text{if} \quad E_{x(left)} > E_{y(right)}, \quad \min\tau \le j \le 0 \\ &\text{if} \quad E_{x(left)} < E_{y(right)}, \quad 0 \le j \le \max\tau \end{aligned} \quad (3)$$

where $\max\tau$ and $\min\tau$ are df/v and -df/v, and $E_{x(left)}$ and $E_{y(right)}$ are the energies of signals x and y respectively.

4. Azimuth Computation

The azimuth calculation is dependent on the Time Difference of Arrival value [Murray et al., 2004]. It is basically a computation of the angle of incidence of the sound. After the TDOA value is obtained by taking the offset of the maximum correlation value, the following variables are employed to get the value of the angle. The first variable is the sampling rate of the signals. In our system, we used a sampling rate of 44.1 kHz, that is 44,100 samples per second and our delta value is shown in equation 4.

$$\text{Delta} = \frac{1}{44{,}100} = 2.2676 \times 10^{-5} \quad (4)$$

$$t = \Delta \times \tau \quad (5)$$

$$\theta = arcsine\frac{vt}{d} \quad (6)$$

Other variables are the velocity of sound, delay time and the distance d between the two microphones. The velocity of sound, v, is assumed to be *343m/s* and the delay time is the delay in samples multiplied by delta, see equation 5. Azimuth (angle) value Θ, is derived based on the trigonometry function arcsine, as shown in equation 6. After computing the azimuth of the signal, it is then converted into its 360° equivalent using <Table 1> as reference.

<Table 1> Angle Conversion Reference Table

Side of system	Sign of TDOA	Conversion to 360°
Front	Positive	angle
Back	Positive	180°-angle
Back	Negative	180°-angle
Front	Negative	360°+angle

5. Experiment and Discussion

We implemented our proposed method on an Intel PC using Visual C++ and Portaudio library for real-time sound capture. Our experimental setup includes three dynamic cardioid microphones connected to a TASCAM US 4×4 audio interface. The microphones are placed in the form of a triangle and spaced equally, with a distance of 30cm between each pair. The experimental data were generated by clapping at a distance of at least one meter away from the microphone set-up and they were captured in real-time using a sampling rate of 44.1 kHz.

As we have assumed that most of the sound sources are located in front of the system, the primary microphone pair is first used to compute the TDOA value using the partial cross correlation. Then, in order to distinguish between the front and back signals, we include a secondary microphone pair consisting of either microphone 1 or 2 and the back microphone 3. The goal is to eliminate the need to select one of three sections as in previous research [Lee et al., 2005].

To confirm the performance and functionality of our proposed method, we performed multiple tests for each test angle. Specifically, we performed the experiment 10 times for each of the 14 angles and then we recorded the average of all 10 estimates for each angle. Mostly the system functioned well as expected; however in some few cases we observed that the system produces incorrect values due to inaccurate energy comparison results. This inaccuracy is attributed to the presence of noise and other distractions in the environment of the experiment.

〈Table 2〉 Results of Angle Estimation

Actual Angle(°)	Estimated Angle(°)	Error(°)	Side
0	0.00	0.00	Front
30	29.5211	0.4789	Front
60	62.48842	2.48842	Front
80	80.1291	0.1291	Front
100	99.8709	0.1291	Back
120	118.177	1.823	Back
150	148.7667	1.2333	Back
180	179.7524	0.2476	Back
210	209.5115	0.4885	Back
240	241.8228	1.8228	Back
260	260.1291	0.1291	Back
280	279.8709	0.1291	Front
300	298.1772	1.8228	Front
330	328.7667	1.2333	Front

<Table 2> shows results of the angle estimation experiments with the predicted directions. Our proposed method computes angle estimates correctly with a high accuracy in most cases. The maximum and average error values recorded were 2.48842° and 1.012918° respectively, which are very low compared to some previous works. The method also determines the correct side of the microphone array where the sound source is (i.e. front or back), as shown in the last column.

Compared to previous methods, our proposed method spends a shorter computation time in that, we mostly implement only the partial cross correlation which uses only half of the time required by the conventional cross correlation implementation. Although our previous work [Yiwere and Rhee, 2016], achieves a shorter computation time by computing only two TDOA values, our current proposed method outperforms it in terms of computation time since it uses only one TDOA computation most of the time. <Table 3> shows a comparison of our proposed method to other previous methods [Lee et al., 2005; Yiwere and Rhee, 2016] in terms of number of TDOA computations.

As mentioned above, in the presence of too much noise, the energy estimation is sometimes affected, leading to some incorrect predictions and estimation. However, when measures are taken to reduce the effect of noise, the proposed method will function relatively better.

〈Table 3〉 Comparison of TDOA Computations

SSL Method	Number of Microphones	Number of TDOA Computations
Conventional 3-Microphone Method	3	3
TDOA Sign with L-Shaped Array Method	3	2
Proposed Method	3	1.5

6. Conclusion

This paper describes a simple sound source localization method using signal energies and partial cross correlation for TDOA computation. Our focus in this work was to reduce the overall localization time by cutting down the amount of computations involved. In order to achieve this, we mainly used only one microphone pair and we implemented the partial cross correlation method which uses half the conventional cross correlation time. Also we used energy comparison to eliminate the front back confusion, thereby achieving 360° range of localization.

Results of our experiments confirm that our proposed method is comparable to other three microphone array localization methods in terms of accuracy; but it uses fewer computations which imply shorter computation time. Also, the algorithm is able to distinguish between front and back sound sources with very little computation. Due to the presence of noise, the energy comparison is sometimes affected leading to incorrect TDOA and angle estimations. However, by suppressing the effect of noise in the environment, our proposed method will function accurately. It can be used to speed up sound source localization in video surveillance systems, robot hearing components etc.

In future we plan to introduce a few more steps in the process to remove or suppress noise in the captured signals. This will help to increase the reliability of our energy comparison-based side prediction.

References

[1] Alan-Boyd, A. W., Whitmer, W. M., Brimijoin, W. O., and Soraghan, J. J., “Biomimetic direction of arrival estimation for resolving front-back confusions in hearing aids”, *The Journal of the Acoustical Society of America*, Vol. 137, No. 5, 2015, pp. EL360-EL366.

[2] Broeck, B. V. D., Bertrand, A., Karsmakers, P., Vanrumste, B., Van hamme, H., and Moonen, M., “Time-Domain GCC-PHAT Sound Source Localization for Small Microphone Arrays”, Education and Research Conference (EDERC), *2012 5th European DSP*, 2012, pp. 76-80.

[3] Hwang, S., Park, Y., and Park, Y., “Sound

Source Localization using HRTF database", Proceedings of International Conference on Control, Automation, and Systems (ICCAS 2005), 2005, pp. 751-755.

[4] Knapp, C. H. and Carter, G. C., "The generalized correlation method for estimation of time delay", *IEEE Transactions on ASSP*, Vol. 24, No. 4, 1976, pp. 320-327.

[5] Kwon, B. G., Kim, G. G., and Park, Y. J., "Sound Source Localization Methods with Considering Microphone Placement in Robot Platform", 16th IEEE International Conference on Robots and Human Interactive Communication, 2007, pp. 127-130.

[6] Lee, J. Y., Chi, S. Y., Lee, J. Y., Hahn, M., and Cho, Y. J., "Real-time sound localization using time difference for human-robot interaction," IFAC Proceedings Volumes (IFAC-Papers Online), Vol. 16, 2005, pp. 54-57.

[7] Li, X., Shen, M., Wang, W., and Liu, H., "Real-time Sound Source Localization for a Mobile Robot Based on the Guided Spectral-Temporal Position Method", *International Journal of Advanced Robotic Systems*, Vol. 9, No. 78, 2012, pp. 1-8.

[8] McAlpine, D. and Grothe, B., "Sound Localization and delay lines-do mammals fit the model?", *TRENDS in Neuroscience*, Vol. 26, No. 7, 2003, pp. 347-350.

[9] Murray, J. C., Erwin, H., and Wermter, S., "Robotic Sound-Source Localization and Tracking using Interaural Time Difference and Cross-Correlation", AI workshop on Neuro Biotics, Germany, 2004.

[10] Sreejith, T. M., Joshin, P. K., Harshavardhan, S., and Sreenivas, T. V., "TDE Sign Based Homing Algorithm for Sound Source Tracking Using a Y-shaped Microphone Array", 23rd European Signal Processing Conference (EUSIPCO), 2015, pp. 1207-1211.

[11] Tashev, I., Sound Capture and Processing, John Wiley and Sons, 2009.

[12] Usagawa, T., Saho, A., Imamura, K., and Chisaki, Y., "A Solution of Front-back Confusion within Binaural Processing by an Estimation Method of Sound Source Direction on Sagittal Coordinate", TENCON 2011-2011 IEEE Region 10 Conference, 2011, pp. 1-4.

[13] Yiwere, M. and Rhee, E. J., "A TDOA Sign-Based Algorithm for Fast Sound Source Localization using an L-Shaped Microphone Array", *Journal of Information Technology Applications and Management*, Vol. 23, No. 3, 2016, pp. 87-97.

[14] Yiwere, M. and Rhee, E. J., "Fast Time Difference of Arrival Estimation using Partial Cross Correlation", *Journal of Information Technology Applications and Management*, Vol. 22, No. 3, 2015, pp. 106-114.

■ Author Profile

Mariam Yiwere

She received her B.S. degree in Computer Science from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana in 2012. In 2015, she received her M.S. degree in Computer Engineering from Hanbat National University, Daejeon, Korea, where she is currently pursuing Ph.D. degree. She is currently conducting research in the area of sound source localization in the Artificial Intelligence and Computer Vision Lab in the Graduate School of Information and Communications, Hanbat National University. She is interested in computer vision, digital signal processing and artificial intelligence

Eun Joo Rhee

He is a Professor of Department of Computer Engineering at College of Information Technology, Hanbat National University, Daejeon, Korea. He has the degree of Ph.D in Electronics Engineering from Chungnam National University. His research interests include in image processing, pattern recognition, computer vision and artificial intelligence. His papers have appeared in IEICE Trans. on Information and Systems, Journal of KIISE, Journal of the Institute of Electronics and Information Engineers, Journal of Information Technology Applications and Management, Journal of Information Technology Application, Journal of the Modern Linguistic Society of Korea, Journal of Korea Multimedia Society, Journal of the Korea Academia-Industrial cooperation Society.