JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지(J. Korea Inst. Inf. Commun. Eng.) Vol. 21, No. 2 : 294~300 Feb. 2017

에너지 기반 영역 선택과 TDOA에 의한 3차원 음원 위치 추정

Mariam Yiwere · 이은주*

3-D Sound Source Localization using Energy-Based Region Selection and TDOA

Mariam Yiwere · Eun Joo Rhee*

Department of Computer Engineering, Hanbat National University, Daejeon 34158, Korea

요 약

본 논문에서는 에너지 기반 영역 선택과 TDOA에 의해 3차원에서 음원의 방위와 높이를 계산하여 음원 위치를 추정하는 방법을 제안한다. 본 연구의 목적은 음원 위치 추정에서 계산시간 감축으로, 수평면 3개 신호의 에너지 비교에 의한 영역 선택과 선택된 영역의 TDOA에 의해 방위각을 계산하고, 또 높이 계산을 위한 마이크로폰 신호와 가장 큰 에너지를 갖는 평면 신호와의 TDOA로 높이각을 추정하는 방법을 제안한다. 제안한 방법에 대한 음원 추정 실험 결과 수평 방위각 추정에서 평균 0.778°, 높이각 추정에서 1.296°의 오류를 보여 기존의 방법과 정확도에서 유사하고, 추정은 1회 신호 에너지 비교와 2회의 TDOA계산으로 가능하여 처리 시간이 단축된다.

ABSTRACT

This paper proposes a method for 3-D sound source localization (SSL) using region selection and TDOA. 3-D SSL involves the estimation of an azimuth angle and an elevation angle. With the aim of reducing the computation time, we compare signal energies to select one out of three regions. In the selected region, we compute only one TDOA value for the azimuth angle estimation. Also, to estimate the vertical angle, we choose the higher energy signal from the selected region and pair it up with the elevated microphone's signal for TDOA computation and elevation angle estimation. Our experimental results show that the proposed method achieves average error values of 0.778° in azimuth and 1.296° in elevation, which is similar to other methods. The method uses one energy comparison and two TDOA computations therefore, the total processing time is reduced.

키워드 : 방위각, 높이, 신호 에너지, TDOA

Key word : Azimuth, Elevation, Signal Energy, Time Difference of Arrival

Received 29 December 2016, **Revised** 30 December 2016, **Accepted** 10January 2017

* **Corresponding Author** Eun Joo Rhee(E-mail:ejrhee@hanbat.ac.kr, Tel:+82-42-821-1205)
Department of Computer Engineering, Hanbat National University, Daejeon 34158, Korea

Open Access **http://doi.org/10.6109/jkiice.2017.21.2.294** print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. INTRODUCTION

1.1.Background and Objective

Over the past decades, sound source localization which is the process of estimating the direction of a sound source has been extensively studied by researchers. The method 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 conference, etc.

The method 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 by the different microphones. With the help of trigonometric functions, the sound’s angle of incidence is estimated using the time delay value. The time delay value can be computed using either frequency domain [1] or time domain cross correlation [2-4].

The simplest way to perform 3-D sound source localization is by using four microphones; three microphones on the horizontal plane for estimating the horizontal angle and one elevated microphone for the estimation of the vertical angle of the sound source. 3-D sound source localization is important because it is very useful in modern day technology; for example, it enables robots to mimic the auditory mechanism of human beings to a higher degree becoming almost accurate; a service robot can look up to a speaker’s face rather than just looking to the direction of the speaker’s location.

1.2. Related Work

Generally, there are three broad categories of sound source localization methods [5]: 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. TDOA methods are less expensive to implement because they use fewer microphones and less computations.

3-D sound source localization methods based on TDOA have been suggested by many researchers [6-9] in the past few decades. They used different approaches and microphone array setups; while some used only three microphones, others used four or more. For example, Byoungho Kwon et al. proposed two different methods for 3-D localization [6]; one for closed microphone position and one for the open microphone position. In their first method, which is the open microphone position, they compute three TDOA values for the horizontal angle estimation. With this approach, the overall computation time is increase due to multiple computations.

Nima Yousefian et al. proposed a 3-D localization using only 3 microphones [7]. First they estimate the approximate position of the sound source by computing TDOA values of the three microphone pairs and also the comparing the energy received by each pair of microphones, and then they apply an error criterion to the points in the vicinity of the estimated position. The multiple computations of TDOA values together with other estimations in this approach increases the algorithms computation time.

Also, Sangmoon Lee et al. proposed a new 3-D localization method using inter-channel time difference trajectory [8]. Their method was implemented with a two-channel rotating microphone array and by analyzing the source direction-dependent characteristics of the trajectories using the Ray-Tracing formula for 3-D models. This approach will also take longer processing time because of the analysis of multiple trajectories.

Yuki Tamai et al. also proposed a 3-D sound localization method using an arrangement of 32 microphones which forms an optimized pattern with three rings [9]. They apply Delay and Sun Beam Forming method and Frequency Band Selection to estimate the 3-D location, source distance as well as sound sources separation. This method uses a large number of microphones and this increase the cost of the system.

Also some researchers do generate a new signal in order to estimate the elevation of the sound source [5]. This step also adds up to the method’s computation time and it can be avoided to save time.

1.3. Suggestion

By studying some previous works, we realized that the use of multiple TDOA values which leads to increased computation time cuts across many localization algorithms and there is a need for an algorithm that will reduce the amount of computations and still achieve acceptable localization results. Such a method will be useful in certain applications where accuracy of the localization is not as critical as the processing time.

In order to solve the problem of too much computation, we suggest a 3-D localization method that will be based on region selection using signal energy comparison and also TDOA estimation.

Firstly, we use the signal energy comparison to select one region that is estimated to have the sound source, and then we proceed with the actual localization by computing the TDOA value using the two signals in the selected region. Next, we pair one of the microphones in the selected region with the fourth elevated microphone in order to estimate the elevation of the sound source. Depending on the selected region, the horizontal angle is converted to a 360° equivalent angle.

The organization of this paper is as follows; region selection based on energy comparison and the azimuth and elevation computation are described in sections II and III respectively, section IV describes the TDOA computation, experiments and discussion is presented in section V and section VI presents conclusion.

Ⅱ. Region Selection based on Energy Comparison

Since the power and energy of a signal reduces with distance, we know that all microphones in the microphone array will receive signals with different energy levels even though they are all from the same sound source. This is because the microphones in the array are spaced out and are not at the exact same position. Fig. 1 illustrates the relationship between signal energies and distance from source to microphone. Notice that microphone 2 is closer to the sound source, with a distance d, and therefore it receives a signal with a higher energy level than that of microphone 1.

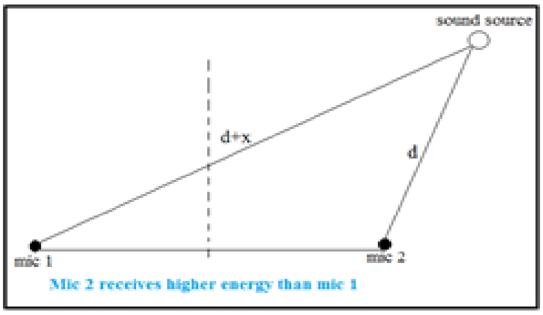

Fig. 1 Sound source localization with 2 microphones.

With this notion, when signals are receive from the sound source by the microphone array, we first estimate their energies using equation 1. Table 1 shows examples of estimated energies of input signals. These signal energy levels are then compared with one another and the two signals with the highest energy levels are selected for TDOA computation [10]. We estimate that the sound source will be located in this region because it is closer to the microphones in the region.

Table. 1 Example of signal energy comparison

	Signal Energies		
	Mic 1	Mic 2	Mic 3
Region 1 (Mics 1&2)	159.983	172.403	141.984
Region 2 (Mics 2&3)	147.759	179.44	154.708
Region 3 (Mics 3&1)	131.713	118.523	169.116

Fig. 2 shows the positioning of microphones in the horizontal plane and the three regions that are formed by this arrangement. As mentioned earlier, by arranging the microphones in the form of a triangle with equal distances and angles apart, each microphone pair creates one region in which the sound source can be located.

$$Energy_x = \sum_{i=0}^{N-1} |x_i| \quad (1),$$

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

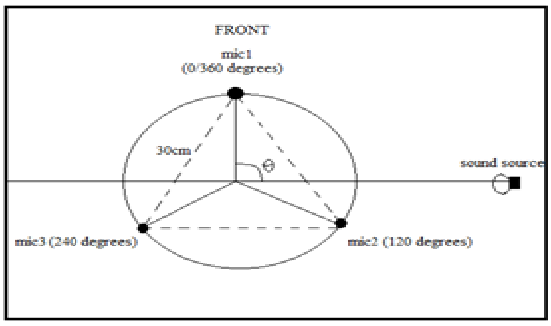

Fig. 2 Horizontal plane microphone positions and region formation.

Ⅲ. Azimuth and Elevation Computation

Fig. 3 shows a diagram of our microphone setup consisting of four microphones. Microphones 1, 2 and 3 are on the horizontal plane as shown in section II and microphone 4 is elevated at the center of the triangle. Notice that for the azimuth value, only microphone pairs 1-2, 2-3, and 3-1 can be used in computing the TDOA value, whereas for the elevation angle, only microphone pairs 1-4, 2-4 and 3-4 and be used.

To estimate the azimuth and elevation angles, we use the TDOA value together with other variables namely; the velocity of sound *(v)*, delay time *(t)* and the distance *(d)* between the two microphones involved. To get the delay time, the inverse of the signals' sampling rate is multiplied by the TDOA value as shown in equation 3. Equation 2 shows the inverse sampling rate, and equation 4 is used to get the angle value.

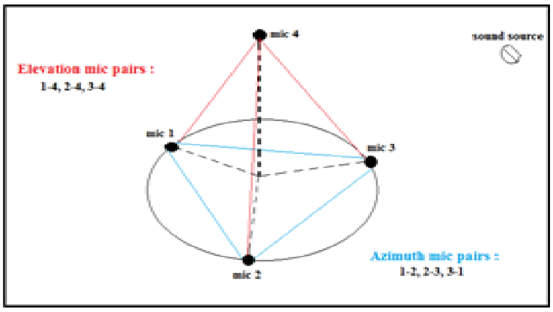

Fig. 3 3-D localization microphone setup.

As explained above, the azimuth angle is estimated by first selecting two microphones which make up one region by using a comparison of signal energies as shown in Fig. 4 (a). These two microphones' signals are used to compute a TDOA value, which is in turn used to compute the azimuth angle. Depending on the selected region, the estimated azimuth angle is converted to its 360° equivalent.

$$\Delta = \frac{1}{44100} = 2.2676 \times 10^{-5} \quad (2)$$

$$t = \Delta \times \tau \quad (3)$$

$$\theta = \mathrm{arcsine}\frac{vt}{d} \quad (4)$$

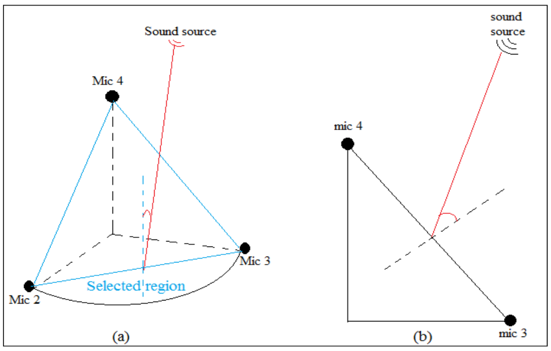

Fig. 4 Selected region and elevation angle. (a) Selected one of 3 regions (b) Microphone pair for elevation angle.

Next, to get the elevation angle, we choose from the selected region the signal that has the higher energy. It is expected that this signal is the closest to the sound source and therefore, it is suitable for the estimation of the elevation angle. Once it has been decided, the signal is paired with the fourth signal which is elevated at the center of the three horizontal plane microphones. In the Fig. 4 (b), the microphone 3 is chosen for pairing.

Fig. 5 shows a flow diagram of the described method. First we capture the four channel signal and estimate the energies of the individual signals at microphones 1,2 and 3. Based on the energy values we select two signals for the TDOA and azimuth estimations. Next, we choose the signal with the highest energy in the selected region, and pair with the elevated microphone 4 for the estimation of the elevation angle.

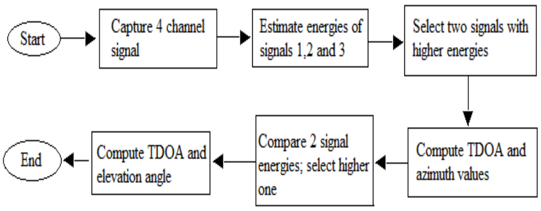
Start
Capture 4 channel signal
Estimate energies of signals 1,2 and 3
Select two signals with higher energies
Compute TDOA and azimuth values
Compare 2 signal energies; select higher one
Compute TDOA and elevation angle
End

Fig. 5 Flow diagram of the proposed method.

Ⅳ. TDOA Computation

TDOA value is the estimated delay in times of arrival between two signals. It is estimated by performing a cross correlation of the two signals involved. We use the time domain cross correlation in our method to estimate the TDOA values[2]. In order not to compute the cross correlation for all possible delay values, which means saving some computation time, we first compute the relevant range of delay values, that is, the maximum and minimum delay values possible. 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) can be estimated is *df/v*, and the minimum delay is *-df/v* as shown in equation 5. Equation 6 is used to compute the cross correlation of two signals, after which the index of the maximum correlation coefficient in the output array is selected as the delay value.

$$Range = [\min\tau, \max\tau] = [\frac{-df}{v}, \frac{df}{v}] \quad (5)$$

$$Corr(x,y)(j) = \sum_{k=0}^{N-1} x_{(j+k)} y_{(k)} \quad (6),$$

where x and y represent the two signal and N is the length of one signal.

Ⅴ. Experiments and Discussion

The proposed method was implemented on an Intel PC using Visual C++ and Portaudio library for real-time sound capture. We used four dynamic cardioid microphones in our experimental setup and connected them to a TASCAM US 4x4 audio interface. Three microphones on the horizontal plane are arranged in the form of a triangle and spaced equally, with a distance of 30cm and an angle of 120° between each pair. The fourth microphone is positioned at the center of the triangle and elevated to a distance of 30cm from the horizontal plane microphones. Sound for the experiment were generated 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.

During our experiments, we performed multiple tests for each test angle in order to confirm the performance of our proposed method; at least 10 repetitions for several test angles and we recorded their average value. Tables 2 and 3 show experimental results of the horizontal and vertical angles estimations respectively. We realize from the values that our proposed method functions with high accuracy and minimal error values[11].

The proposed method uses very few computations compared to other 3-D sound source localization methods and as a result, it spends a shorter processing time. Since we use region selection to choose one microphone pair, we compute only one TDOA value for azimuth estimation, unlike the methods which compute TDOA for three microphone pairs [6-9]. In terms of complexity, our method performs two TDOA computations resulting in $2(N_l \times N_s)$ computations and $3N$ summations, whereas other methods perform three TDOAs, i.e. $3(N_l \times N_s)$ and more computations for signal generation [12].

Table. 2 Results of horizontal angle estimation.

Actual Angle(°)	Estimated Angle(°)	Error Angle(°)
0	0.00	0.00
30	29.6932	0.3067
60	61.9211	1.9211
80	80.1291	0.1291
120	118.177	1.823
150	148.7667	1.2333
180	179.7524	0.2476
210	209.5115	0.4885
240	241.8228	1.8228
260	260.1291	0.1291
300	300.00	0.00
330	328.7667	1.2333

Table. 3 Results of elevation angle estimation.

Actual Angle(°)	Estimated Angle(°)	Error Angle(°)
0	0.7428	0.7428
5	5.2067	0.2067
10	9.70256	0.2974
15	17.3482	2.3482
20	19.6967	0.3033
30	33.87855	3.8785

In some instances, environmental noise affects estimation of the signal energy levels and this in turn, affects the estimation of the elevation angle. To overcome this drawback, environmental noise needs to be suppressed during usage of the method.

Ⅵ. Conclusion

To solve the problem of too many computations, we have proposed a simple 3-D sound source localization method using region selection and TDOA computation. We select the region of the sound source using energy comparison before computing the TDOA value to estimate the horizontal angle. The horizontal angle is converted to a 360° equivalent based on the selected region. Also for the estimation of the elevation angle, we pair one microphone from the selected region with the fourth elevated microphone for TDOA estimation.

Our experimental results show that our proposed method performs well and it can be compared to other 3-D localization methods in terms of accuracy; however, our method uses fewer computations because we use a total of two TDOA estimations and also, we do not generate any new signal for the estimation of elevation angle. 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 and more consistently. The method can be used in applications where accuracy is not as critical as the speed of processing.

In the future, we plan to incorporate some preprocessing steps that will remove or reduce noise in the captured signals, so as to attain more accurate signal energy estimations.

REFERENCES

[1] C.H. Knapp and G.C. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Transactions on ASSP*, vol. 24, no. 4, pp. 320-327, Aug. 1976.

[2] M. Yiwere and E. J. Rhee, “Fast Time Difference of Arrival Estimation using Partial Cross Correlation,” *Journal of Information Technology Applications & Management*, vol. 22, no. 3, pp.106-114, Sept. 2015.

[3] J. C. Murray, H. Erwin and S. Wermter, “Robotic Sound-Source Localization and Tracking using Interaural Time

Difference and Cross-Correlation," in *Proceedings of the AI workshop on NeuroBiotics*, Germany, pp. 89-97, Sept. 2004.

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

[5] I. J. Tashev, *Sound Capture and Processing*, West Sussex, John Wiley and Sons, 2009.

[6] B. G. Kwon, G. G. Kim and Y. J Park, "Sound Source Localization Methods with Considering Microphone Placement in Robot Platform," in *Proceedings of the 16th IEEE International Symposium on Robots & Human Interactive Communication*, South Korea, pp.127-130, Aug. 2007.

[7] N. Yousefian, M. Rahmani, and A. Akbari, "Sound Source Localization in 3-D Space by a Triple-Microphone Algorithm," in *Proceedings of the 14th Asia-Pacific Conference on Communications*, Japan, pp. 1-5, Oct. 2008.

[8] S. M. Lee, Y. J. Park and Y. S. Park, "Three-dimensional Sound Source Localization Using Inter-Channel Time Difference Trajectory," *International Journal of Advanced Robotic Systems*, vol.12, no. 12, pp. 1-15, Dec. 2015.

[9] Y. Tamai, Y. Sasaki, S. Kagami, and H. Mizoguchi, "Three Ring Microphone Array for 3D Sound Localization and Separation for Mobile Robot Audition," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Canada, pp. 4172-4177, Aug. 2005.

[10] J. Y. Lee, S. Y. Chi, J. Y. Lee, M. Hahn, and Y. J. Cho. (2005, July). Real-time sound localization using time difference for human-robot interaction. *IFAC Proceedings Volumes* [Online]. 16(1), pp. 54-57. Available: http://dx.doi.org/10.3182/20050703-6-CZ-1902.01411.

[11] A. Pourmohammad and S. M. Ahadi, "Real Time High Accuracy 3-D PHAT-Based Sound Source Localization Using a Simple 4-Microphone Arrangement," *IEEE Systems Journal*, vol. 6, no. 3, pp.455-468, Aug. 2012.

[12] D. Hale, "An efficient method for computing local cross-correlations of multi-dimensional signals," Center for Wave Phenomena, Colorado School of Mines, Technical Report CWP-544, 2006.

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. 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.

이은주(Eun Joo Rhee)

He is a Professor of Department of Computer Engineering at College of Information Technology, Hanbat National University, Daejeon, Korea, since 1989. He has the degree of Ph.D in Electronics Engineering from Chungnam National University in 1989. He was a postdoctoral fellow in Graduate School of Imaging Science and Technology of Tokyo Institute of Technology in Japan from 1994 to 1995, and a visiting professor in Oregon Graduate Institute of Science and Technology in America from 1998 to 1999. His research interests are in image processing, pattern recognition, computer vision and artificial intelligence.