International Conference on Control, Automation and Systems 2007
Oct. 17-20, 2007 in COEX, Seoul, Korea

Probabilistic Sound Source Localization

Yoon Seob Lim[1] , Jong Suk Choi[1] and Mun-Sang Kim[1]

[1] Center for Cognitive Robotics Research, Korea Institute of Science and Technology, Seoul, Korea
(Tel : +82-2-958-6743; E-mail: himelys@kist.re.kr)

Abstract: Sound localization is one of ways that keep the relation with robot. Time Delay Of Arrival (TDOA) between two microphones through cross-correlation method has been used for sound localization in our robot platforms. However, since cross-correlation values are highly dependent on the upcoming sound signal and acoustic environment, time delay values and localization result are easily perturbed. Probabilistic method for sound source localization is presented in this paper. Markovian process and post filtering steps are applied for calculating time delay value and a reliable sound localization.

Keywords: Sound localization, Markov process, Bayes theorem.

1. INTRODUCTION

Sound localization is one of ways that keep the relation with robot. Numerous approaches exist which deal with a reliable sound source localization, employing a variety of features, tracking schemes and sensor setups. Kalman filters are commonly used to perform a single object under a Gaussian models and linear dynamics. These methods have been applied to audio and visual object tracking problems [2]. Using Kalman filters, adequate tracking performance can not be obtained when measurement is done in a noisy and cluttered environment.

To relieve this problem, probabilistic models such as Bayesian networks [3] and Sequential Monte Carlo method [4] have been proposed. Among them, particle filter shows a principled method for localization [5]. Particle filter is an approximation technique for a nonlinear and non-Gaussian situation. Particle filters have also been applied to audio source localization using a beamformer–based sound source localization [6].

Time Delay Of Arrival (TDOA) between two microphones has been used for sound localization in our robot platforms [1]. However, since cross-correlation values are highly dependent on the upcoming sound signal and acoustic environment, time delay between two microphones can be easily changed. Localization result is also perturbed by the error of calculated time delay value (Fig. 1). In this paper, we introduce a probabilistic sound localization method in time-domain to alleviate change in time delay value.

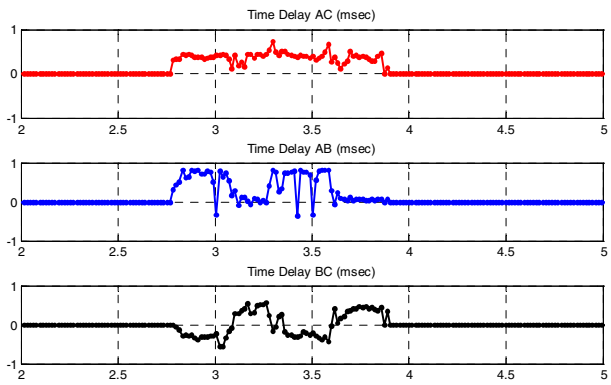

Fig.1 Example of time delay value

2. LOCALIZATION ALGORITHM

2.1 Microphone Arrangement

There are lots of possible causes that result in false localization of sound source, which are noises around the source, reverberant signals into the microphone and arrangement of microphones. The errors in sound localization due to noises and reverberant signals can be dealt with various filtering methods, one of which is proposed in this article. Errors by hardware configuration should be prevented in advance. If the interested location is only front side of robot, we can determine the location of sound source with time delay value with only two microphones. However, it is difficult to find out whether the source lies in the front side or back side of robot because the source located in the opposite of real source can generate the same time

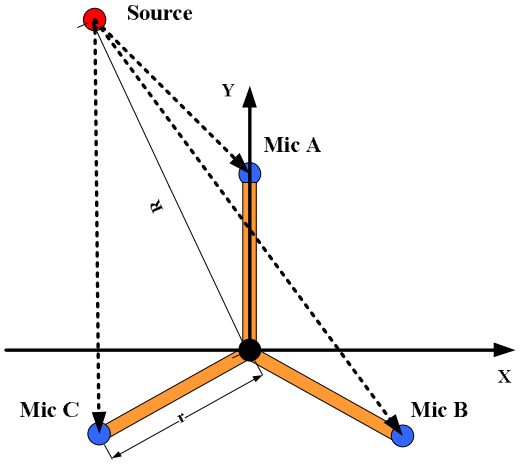

Fig.2 Microphone arrangement

978-89-950038-6-2-98560/07/$15 ⓒICROS

delay value. To clear the ambiguity in source location with only laterally positioned two microphones, additional microphones are needed and some researchers deploy several microphones in line. However, we just added one microphone and displaced them triangularly.

2.2 Probabilistic Process

Even if microphones are arranged triangularly, environmental noise and acoustical condition around robot may cause error in calculating time delay between microphones. To deal with an inappropriate change of time delay between two microphones which is the main reason of error in localization, markovian process is applied to compute a time delay. Time delay value between two microphones is selected as a state variable.

$$\mathbf{X}_t^{ij} = \{\Delta T_{ij}\} \tag{1}$$

Where ΔT_{ij} is time delay between microphone i and j. What we want to calculate with this method is how much currently calculated time delay is reliable according to the observed sound signals. The probability of current time delay up to current observation of microphone pair ($\mathbf{Z}_{ij}^t = \{z_1^{ij}, z_2^{ij}, \cdots, z_t^{ij}\}$) can be written as equation (2).

$$Bel(t) = P\left(\mathbf{X}_t^{ij} \middle| \mathbf{Z}_{ij}^t\right) \tag{2}$$

Summarizing, to localize a sound source we need to calculate the probability recursively at each time step.

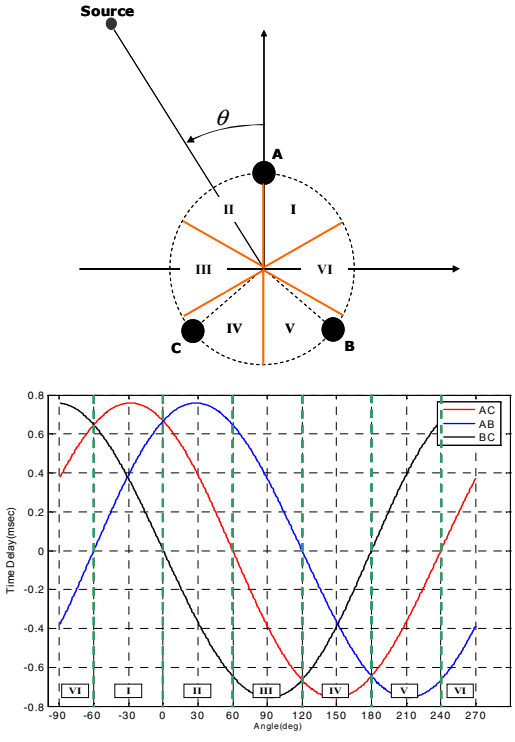

Fig.3 Characteristics of time delay value

This process is composed of two steps – prediction and update step.

2.2.1 Prediction Step

In prediction step, motion model is defined to predict the current state probability under the assumptions that a sound source can not move over a certain angle in one time frame and current state depends on the previous state. Motion model that we applied is as equation (3)

$$P\left(\mathbf{X}_t^{ij} \middle| \mathbf{X}_{t-1}^{ij}\right) \sim N\left(\mathbf{X}_{t-1}^{ij}, \sigma^2\right) \tag{3}$$

With this motion model, we can estimate the probability of each time delay state as equation (4).

$$P\left(\mathbf{X}_t^{ij} \middle| \mathbf{Z}_{ij}^{t-1}\right) = \sum_{\mathbf{X}_{t-1}^{ij}} P(\mathbf{X}_t^{ij} \middle| \mathbf{X}_{t-1}^{ij}) P(\mathbf{X}_{t-1}^{ij} \middle| \mathbf{Z}_{ij}^{t-1}) \tag{4}$$

Prediction of current state is presented as summation because time delay value between two microphones is discrete and is dependent on the sampling frequency.

2.2.2 Update Step

To obtain the posterior distribution $P\left(\mathbf{X}_t^{ij} \middle| \mathbf{Z}_{ij}^t\right)$, measurement likelihood model is proposed to incorporate sound information from microphones. We propose a measurement likelihood model as cross-correlation value spanning around possible time delay value of certain microphone pair.

$$P(z_t^{ij} \middle| \mathbf{X}_t^{ij} = \tau) = \int_{-\infty}^{+\infty} \frac{x_i(t) \cdot x_j(t-\tau)}{\sqrt{\int x_i(t)^2} \cdot \sqrt{\int x_j(t)^2}} dt \tag{5}$$

Where $x_i(t)$ is sound signal of microphone i at time t. The posterior distribution over $\mathbf{X}_t^{ij}$ is obtained by Bayse theorem:

$$P(\mathbf{X}_t^{ij} \middle| \mathbf{Z}_{ij}^t) = \frac{P(z_t^{ij} \middle| \mathbf{X}_t^{ij}) \cdot P(\mathbf{X}_t^{ij} \middle| \mathbf{Z}_{ij}^{t-1})}{P(z_t^{ij} \middle| \mathbf{Z}_{ij}^{t-1})} \tag{6}$$

The posterior represents the probability distribution of time delay and we may expect time delay of certain microphone pair as the value when this probability is the highest. And this whole process is independently applied to each microphone pair.

2.3 Post Processing

In our microphone configuration, time delay values are specifically determined according to the sound source location. Fig 3 shows this characteristic of our microphone configuration at each sound position. Since posterior distribution depends on the measurement likelihood model, undesirable time delay sets can be selected with only probabilistic process shown above.

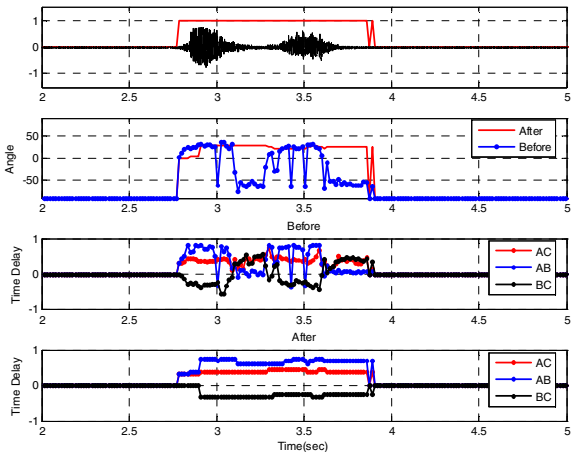

Fig.4 Sound localization result

To filter out those unwanted time delay set, if calculated time delay value corresponds to one of 6 regions, we adopt the time delay as current time delay value and vice versa after posterior computation of each microphone pair.

2.4 Sound Localization

Since we determine a highly probable time delay pairs, we can localize a sound source. A time delay grid for a sound source is pre-determined along 360 degrees and it is assumed to be situated at a specific distance. Localization likelihood is defined as (7).

$$L(\theta) = 1 / \sum_{i \neq j} (t_{ij}(\theta) - \hat{t}_{ij})^2 \qquad (7)$$

Where $t_{ij}(\theta)$ is time delay grid and $\hat{t}_{ij}$ is time delay value determined by probabilistic method. Sound location is determined where the likelihood has maximum value.

3. RESULTS

We apply proposed method to localize a sound source in several situations; when source stands still or moves at certain angular speed around a robot. Sound signal is sampled at 16 kHz and frame at each time step is composed of 512 samples and moved while some samples (ex. 50%) are overlapped. To have a fine observation result, we performed 3~6 times cross-correlation at each time frame. At each cross-correlation, 256 samples are used and samples are overlapped according to number of cross correlation. After the cross-correlation step, we performed a probabilistic process and post filtering step for each

Table.1 Sound localization performance

Distance	1m	2m	3m
Performance	100%	95%	76%

Input voice: "티롯 이리와", which means "Come on, Tirot"

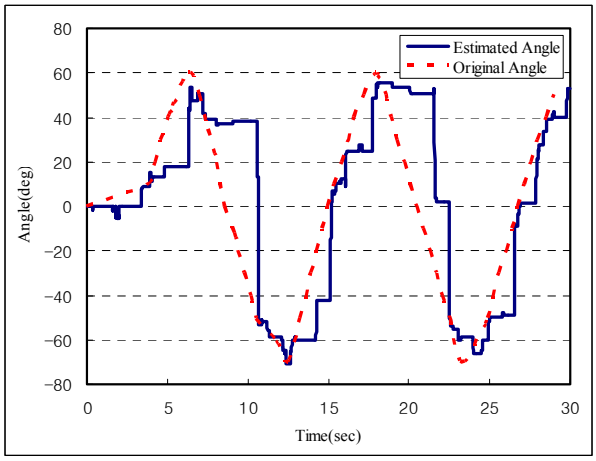

Fig.5 Localization of moving source

cross-correlation value. SNR of input signal is around 13~15dB during experiment.

Fig 4 shows a comparison result when we applied this algorithm or not. The sound source is positioned on 30 degree. Proposed method shows more stable time delay set than those that are determined only by cross correlation. This means that proposed algorithm can filter out unwanted signals such as environmental noise or wind noise at the end of input sound signal, which causes fluctuations of time delay values. Table 1 shows performance of proposed algorithm according to the distance between sound source and robot when speaker is standing at 0^o. Since robot is assumed to interact with human in a room or a house, maximum distance between human and robot is not over 3m. The error range for performance measure is ± 15^o. Even if the speaker is 3m away from robot, robot can localize sound source with over 70% correctness. However, full angular range performance is needed to be evaluated for overall efficiency of proposed method.

Localization for moving sound is also important for a reliable interaction with a robot for speaker. Fig 5 shows a localization result when a person is moving around a robot from -70^o to 60^o. Solid line is estimated position of source by proposed method and dot line is a real position of the source. Overall, tracking is acceptable even though there is some delay for localization, which is mainly because of motion model in the prediction step under the assumption that the source would not move in one time frame (duration is about 16msec).

4. CONCLUSION

Sound is one of the essential modalities for a reliable interaction with robot, which delivers a substantial amount of spatial, temporal information between human and robot (location, time delay). According to the localization performances shown in this article, spatial information through the improved computation of time delay between two microphones is feasible in a probable situation. In sum, we have enhanced a sound localization using a probabilistic method for a reliable interaction with a robot via acoustic information. Using this method, a robot can localize a certain speaker and

can use the result while interacting with human or robot.

5. ACKNOWLEDGEMENT

This research is supported by Development of Active Audition System Technology for Intelligent Robots through Center for Intelligent Robotics in South Korea.

REFERENCES

[1] H.D. Kim, J.S. Choi, and M.S. Kim, "Human-Rob ot Interaction in Real Environments by Audio-Vis ual Integration," *International Journal of Control*, Automation, and Systems, vol. 5, no. 1, 2007, pp. 61-69

[2] D.E. Sturim, M.S. Brandstein, and H.F. Silverman, "Tracking multiple talkers using microphone-array measurements," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997.

[3] V. Pavlovic, A. Garg, and J. Rehg, "Multimodal speaker detection using error feedback dynamic Bayesian networks," *Proceedings of the IEEE CVPR*, Hilton Head Island, SC, 2000.

[4] J. Vermaak, M. Gangnet, A. Blake, and P. Perez, "Sequential Monte Carlo fusion of sound and vision for speaker tracking," *Proceedings of IEEE ICCV*, Vancouver, July 2001.

[5] A. Doucet, N. de Freitas, and N. Gordon, Sequenti al Monte Carlo Methods in Practice, Springer-Verl ag, 2001

[6] D.B. Ward, E.A. Lehmann, and R.C. Williamson, "Particle Filtering Algorithms for Tracking an Acoustic Source in a Reverberant Environment," *IEEE Transactions on Speech and Audio Processing*, vol. 11, 2003, pp. 826-836.