Sound Source Localization Methods with Considering of Microphone Placement in Robot Platform

Byoungho Kwon¹, Gyeongho Kim², Youngjin Park³,
¹ Byoungho Kwon, Center for Noise and Vibration Control, Dept. of Mech. Eng., KAIST, Daejeon, Korea e-mail: bhkwon@kaist.ac.kr
² Gyeongho Kim, Korea Atomic Energy Research Institute, Daejeon, Korea e-mail: kimgyeongho@kaist.ac.kr
³ Youngjin Park, Center for Noise and Vibration Control, Dept. of Mech. Eng., KAIST, Daejeon, Korea e-mail: yjpark@kaist.ac.kr

Abstract—many different methods for sound source localization have been developed. Most of them mainly depend on time difference of arrival (TDOA) or analytic head related transfer functions (HRTF). In real implementation, since the direct path between a source and a sensor is interrupted by obstacles as like a head or body of robot, it has to be considered the number of sensors as well as their positions. Therefore, in this paper, we present the methods, which are included sensor position problem, to localize the sound source with 4 microphones to cover the 3D space. Those are modified two-step TDOA methods. Our conclusion is that the different method has to be applied in case to be different microphone position on real robot platform.

I. INTRODUCTION

In the field of acoustic, the method to localize the sound source in 3D space has been approached in several ways. There are the methods; the one uses the time difference of arrivals (TDOA) of each microphone pair using the multi-microphone (one-step TDOA and two-step TDOA) [1], another utilizes a diffraction sphere to generate intensity information, in addition to time/phase information available in popular free field microphone arrays [2], and the other demonstrates the ability of precise azimuth and elevation estimation, using a generic head related transfer function (HRTF) database [3].

Also, there have been several researches to apply to real robot platform to confirm the performance of their system [4, 5, and 9]. The humanoid robot SIG [4] uses two pairs of microphones, one pair installed at the ear position of the head to collect sound from the external world, and the other placed inside the head to collect internal sounds for noise cancellation. Like human, this uses binaural localization, i.e., the ability to locate the sound source in 3D space. In LABOIIUS, using an array of 8 microphones, there has been implemented a system that accurately localizes sounds in 3D space with short-duration sound at a mobile robot [5]. Ref. [8] proposes the method which can estimate the direction and the position using the delay and sum beam forming (DSBF), the frequency band selection (FBS) and random sample consensus (RANSAC) algorithms with 32 microphones.

However, it is difficult for above mentioned methods to apply a general robot, because those have some demerits which are to require storage spaces of HRTF database and to need many microphones. Even though the use of microphone array including many sensors increases the resolution of the localization procedure and its robustness to ambient noise as like [8], there are some problems to consider the robot platform design and limitation of installing microphones. To improve these shortcomings, there need a use of the minimal number of microphones to estimate the sound direction in 3D space and TDOA method to increase the efficiency when applied to real robot.

In addition, there is considered the microphone positions to be installed on a real robot platform because they influence the performance. In case of more than two microphones, time delays of microphone pairs are used to estimate the sound direction. But if direct path from source to microphone is interrupted by a head or body of robot, this causes a lot of error in calculating the time delay with distorted signals. Considering that, the opened microphone position is defined by the case to exist the direct path from source to microphone and the closed microphone position is defined by the case not to exist that due to obstacles. Therefore, different algorithm is applied as each case.

In this paper, we present the methods to estimate the sound direction as considering the microphone positions; opened microphone position, closed microphone position.

The rest of the paper is organized as follows: in Section 2, we describe a general two-step TDOA method for sound source localization in 3D using the generalized cross correlation method. Also, we present the modified two-step TDOA method. In Section 3, we explain the proposed approaches which are the opened microphone position and the closed microphone position. In Section 4, we conduct experiments for each case in the anechoic chamber, and compare the performance of the proposed approaches. The results demonstrate the need of appropriate method as each case.

II. SOUND SOURCE LOCALIZATION IN 3D SPACE

We have to know the azimuth and elevation angle to localize the sound source in 3D. Usually, these are estimated using time delays of microphone pairs, which are described from the coherence of signals. The most common coherence measure is a simple cross-correlation between the signals perceived by expressed by:

$$R_{xy}(\tau) = \sum_{n=0}^{N-1} x_n[n] y_n[n-\tau]$$  \hspace{1cm} (1)
where, $x[n]$ is the signal received by microphone $i$ and $r$ is the correlation lag in samples. The cross-correlation $R_r(r)$ is maximal when $r$ is equal to the offset between the two received signals. Generally, it is approximated using the inverse Fourier transform of a cross spectral density function of two signals to reduce computing to $N \log_2 N$ because that of equation (1) is proportional to $N^2$. The correlation approximation is given by:

$$R_r(r) \approx \sum_{n=0}^{N-1} X_r(k)X_r(k)^* e^{j2\pi r/N} \quad (2)$$

where, $X_r(k)$ is the discrete Fourier transform of $x[n]$ and $X_r(k)X_r(k)^*$ is the cross-spectrum of $x[n]$ and $x[n]$. A limitation of the method is that the correlation is strongly dependent on the statistical properties of the source signal. Since our target signals are mainly speech signals, frequency region of which is generally less than 5 kHz, the correlation between adjacent samples is high and generates cross correlation peaks that can be very wide. Therefore we adopt the weighting function PHAT (the phase transform) to avoid that problem [6].

$$R_r(r) \approx \sum_{n=0}^{N-1} \psi_r(k)X_r(k)X_r(k)^* e^{j2\pi r/N} \quad (3)$$

where, $\psi_r(k) = 1/|X_r(k)X_r(k)| \quad (4)$

From the previous mentioned method, cross correlation of two microphones is calculated and then the sound direction is estimated using that. Minimal three microphones are needed to estimate the sound direction, which is an azimuth angle, using TDOA method in a plane without the front back confusion. Also, to estimate the elevation angle, another microphone which locates the different plane with three microphones to estimate the azimuth angle is needed. Therefore we need minimal four microphones to estimate the sound direction in 3D space.

First, azimuth angle is estimated from three microphones in a plane using the combination of time delays which are calculated from each microphone pair. And then, considering the efficiency, elevation is estimated using the imaginary microphone method. Fig. 1 describes this method briefly, the position of an imaginary microphone is determined with the previous estimated azimuth angle and configuration of microphones, and signals of imaginary one are generated by the nearby two microphones.

![Fig. 1 Position and generation of an imaginary microphone](image1.png)

As elevation angle is estimated using an imaginary microphone method it may include the estimation error of azimuth angle. But the effect of this is a little unless microphones are far away. The general assumption of TDOA method is that microphone is located in free-field. From this assumption, signals of an imaginary microphone which is located at the same direction of source wave propagation vector are generated from a measured data of real microphone. To reduce an error, median value of the generated signals from them of nearby two real microphones is used (Fig. 1 (a)). Although full signals are generally used to calculate the cross correlation for time delay, this paper proposes that time delay is repeatedly calculated from signals which are segmented to the specified length to consider the robustness. The number of sample of a frame is $2^N$ (2048) which is good for fast Fourier transform. After the signals and position of an imaginary microphone are determined, elevation is estimated from time delay between an imaginary microphone and mic.T which is in the different plane in Fig. 1 (a). So it reduces a complexity of dimension from 3D to 2D like Fig. 1 (b). A direction of sound source in 3D space is estimated from the mentioned methods up to now. However, we have to consider the position of installed microphone on a real robot platform. Because the assumption of free-field condition is violated when microphones are installed, there occur lots of estimation errors. The estimation error for an azimuth angle affects the elevation angle error. Therefore we need to apply a different method as an installed microphone position.

### III. Proposed Approaches

The cases of the microphone placement are separated as following.

- **Case 1 Opened microphone position**

If there exists the direct path between a sound source and a microphone, the sound direction is estimated as following method, and that is called the opened microphone position. First, azimuth angle is estimated from three microphones in a plane. And then, as described in Fig. 2, the cross correlation of a microphone pair is converted into a sound direction, that is angle, to consider the configuration of microphones. Finally, the sound direction is determined from the summation of all converted cross correlations. The peak value of Fig. 2 indicates the real position of a sound source. Right side fig-
ures (1, 2, 3) of Fig. 2 which represent the cross correlation of each microphone pair have two peaks. These are the reason why there is the front back confusion due to use of two microphones only. This method is robust at the noise environment because cross correlations of other microphone pairs compensate one which is included a little error.

- **Case 2 Closed microphone position**

Next, in case there does not exist the direct path between a sound and a microphone because head or body of robot interrupts the path (Fig. 3(a)), sound direction is estimated as following method and that is the closed microphone position. First, using the time delay of each microphone pair, we determine the region of source location roughly as like Fig. 3 (b). After the determination of the region, a sound direction is estimated from only one microphone pair in that region.

If the microphones are located at the edge of triangular configuration as described Fig. 3 (b), one microphone pair can cover about 120°. Generally, as sound direction is estimated with TDOA method, if the source leans to the one side a sample delay brings out a large error. Therefore the way to install microphones on a platform is too important. Because this case is used only one microphone pair in a sound source region, time delays are repeatedly calculated using frames which are separated from full signals.

**IV. Experimental Results**

Experimental data are obtained from database of [7] which is hold by our laboratory. We apply the proposed methods to this database, which was measured in anechoic chamber, to consider only the effect of robot platform for sound source localization. Database specification is as following. The distance between a robot platform and a sound source was 1.5m, and a source position is varied 0°~ 180° for azimuth angle, 0°~ 30° for elevation angle as each 10 degrees (Fig. 4). The turn table controlled the angular position of the robot platform precisely. Elevation of source was changed using a specified structure. Source signal was random white noise, and sampling frequency of data acquisition board was 44.1 kHz. Considering the computation time and real applications, we analyze 20 data sets, and each data set has 6000 samples.

There are shown experimental results in Fig. 5 and Fig. 6. The former shows the estimated azimuth angle by change of azimuth one in a specified elevation one, and the later dose the estimated elevation angle by change of azimuth one. Each figure represents all set of data to the size of a circle and square in one specified elevation. As an elevation angle is higher, source is more interrupted by robot head. This effect is reflected in results. An estimation of a sound di-
consider the interruption of a direct path, some errors are one. Even though the closed microphone case is applied to closed microphone position is more exact than the opened position which are directly affected are different. The case of are similar. But the results ((c) ~ (h) of Fig. 7) at 10°~ 30° anechoic chamber without the reflection and reverberation aren’t affected a head or body of robot for two methods in an experimental space, which are reflection, reverberation, and so on. In this experiment, we considered only an effect of robot platform as estimating the sound source. So there is experimented in an anechoic chamber to eliminate environmental factors.

Results are rearranged to Fig. 7. These represent the mean and standard deviation of an estimation error of azimuth and elevation angle for each method. In each figure, transverse lines (dot and solid) represent error mean value and longi-

dinal line dose one standard deviation at each azimuth angle. The results (a), (b) of Fig. 7) at 0° positions which aren’t affected a head or body of robot for two methods in an anechoic chamber without the reflection and reverberation are similar. But the results (c) ~ (h) of Fig. 7) at 10°~ 30° position which are directly affected are different. The case of closed microphone position is more exact than the opened one. Even though the closed microphone case is applied to consider the interruption of a direct path, some errors are caused at 30° elevation because there doesn’t exist direct path perfectly in this case. In addition, these results represent that an estimation error of an elevation is smaller than those for azimuth. The reason why, first, distance between two microphones on platform is less than 30cm, therefore an estimation error for azimuth doesn’t affect an estimation of elevation, and microphones on a robot platform are installed not to affect direct path as change of elevation one.

V. CONCLUSIONS

In this paper, we present two cases to estimate the sound direction of 3D space using TDOA method and experiment each case with an IRP-2004 robot platform in an anechoic chamber and then compare with estimation errors of them. These results are only considered the effect of robot platform for sound source localization, therefore we can not mention the other effects, for example environmental and noise effects. The results of two cases are very similar at positions which satisfy the free-field condition without the interference of head or body. In case of interrupting a direct path between a microphone and a source, however, the case of closed microphone position has better performance than another. Researches for sound source localization keep going on being progressed with consideration of obstacles as like robot platforms.

Finally, even though this paper is experimented in an anechoic chamber, we are going to experiment these two cases in household environment with speech signal, and are going to compare with two cases about an effect of obstacles, environmental factors and input signals.

VI. ACKNOWLEDGMENTS

This work was partly supported by the Korea Science and Engineering Foundation (KOSEF) through the National Research Lab. Program funded by the Ministry of Science and Technology (M1050000112-05J0000-11210), the Brain Korea 21 Project, and the IT R&D program of MIC/IITA [2007-S001-01, “Audio-visual Signal Processing SoC for Intelligent Robot”]

VII. REFERENCES


