REAL-TIME SOUND LOCALIZATION USING TIME DIFFERENCE FOR HUMAN ROBOT INTERACTION

Ji-Yeoun Lee*, Su-young Chi**, Jae-Yeun Lee**, Minsoo Hahn*, Young-Jo Cho**

*Information and Communication University
{jyle278, mshahn}@icu.ac.kr
** Intelligent Robot Research Division
Electronics and Telecommunications Research Institute
{chisy, leejy, youngjo}@etri.re.kr

Abstract: This paper suggests an algorithm that can detect the direction of sound source at real time. The algorithm uses the time difference or phase difference in the sound source from the three microphones, which are located with difference of 120 degree between two microphones. In addition, sound source tracing algorithm having shorter execution time than that of an existing algorithm is proposed to fit the real-time service robot. Our algorithm has almost correct result within range of an error of $\pm$ 6 degree. In consideration of the states of reverberation, echo and the altitude of sound source, if the algorithm is improved, then better efficiency will be expected. *Copyright © 2005 IFAC*

Keywords: Sound localization, Human-machine interface, Robot vision, Interaction mechanism.

1. INTRODUCTION

The intelligent robot needs an ability that recognizes and foretells surrounding as information helper and public helper. Then, the important sense mediums are a sight and hearing, and above two senses are essential to perception functions of robot.

Sight and hearing of robot play a complementary role that helps each other. But, hearing information has the advantage that can extract sound source in the places where are dark or do not look directly by a visual field and long distance. Also, because the dangerous events accompany a roaring sound to human or robot, hearing plays an important role to confident conservation of robot or human. Therefore,

cognition and trace of sound source make robot move to a person or individual neighborhood that need help of robot. And these make possible to foretell surroundings and do suitable confrontation.

Human can percept theoretically the positions of sound source using only two ears with accuracy of ± 10 degree in case of the front in a plane, and with ± 15 degree in case of the back side or the side. So, in this paper, for equal performance improvement in case of the artificial human like robot, three microphones are arranged to difference of 120 degree between two microphones. And that the time difference or phase difference of the sound source that each microphone detects are analyzed and the direction of sound source are obtained.

This paper describes sound source recording system in chapter 2 and sound source tracing algorithm to implement a step that confirms position of sound source while continuously shortening execution time by real-time, high believability in chapter 3. Next, in chapter 4, experimental setup of 3-channel sound localization system and performance of azimuth estimation algorithm are explained. Finally in chapter 5, conclusion about sound source tracing algorithm is discussed.

2. SOUND SOURCE RECODING SYSTEM

2.1 3-channel sound localization system

Figure 1 is generally 3-channel sound localization system. Microphone sets are established over 100cm above the ground. Each microphone is located to difference of 120 degree between two microphones and distance of 30cm on the center of circle like the figure 2. Sound source was happened in about 5m distance of the microphone center.
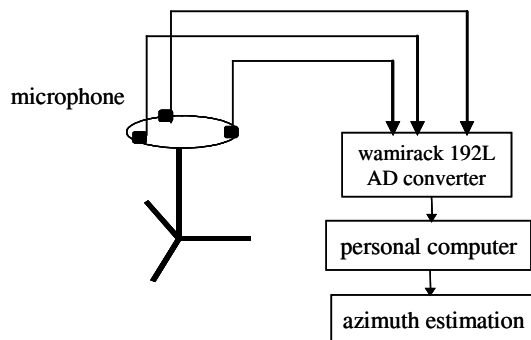

Figure 1. 3-channel sound localization system

The 3-channel AD (Analog to Digital) converter using Ego systems Inc's wamirack 192L digital audio recoding interface like the figure 2 is implemented. Wamirack 192L supports 24 Bit/ 96KHz AD converter and 24 Bit/ 192kHz DA (Digital to Analog) converter. And this system consists of the rack space unit with 4-in/4-out balanced XLR connectors and a PCI interface.

Figure 2. Wamirack 192L digital audio recoding Interface

Using wamirack 192L digital audio recoding interface, sound source is received through three microphones (Philips, SBC ME670) by real-time. Figure 3 shows the positions of the three microphones and the sound source. The three microphones are placed on the circumference of a circle in such a manner that the three microphones form a regular triangle on a horizontal plane, where the radius of the circle is set to a 120 degree.

Also the direction of the microphone #0 is a standard and the azimuth of the sound source is represented by θ (the clockwise direction is positive). R is the distance from the center of the microphone set to each microphone.
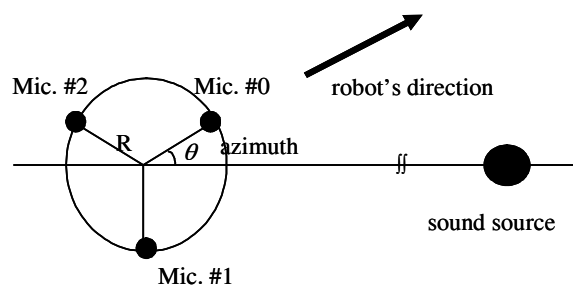

Figure 3. Arrangement of the microphone set

3. SOUND SOURCE TRACING ALGORITHM

In real situation, the execution time of the robot must be considered in such a manner that robot listens to sound source and sets up the direction and moves the position. Thus, an algorithm like the figure 4 is suggested to improve the problem of the existing algorithm. As only calculating time delay in a section of the sound source to be occurred, an algorithm reduces in halves of execution time from point of the suitable application of the real world.
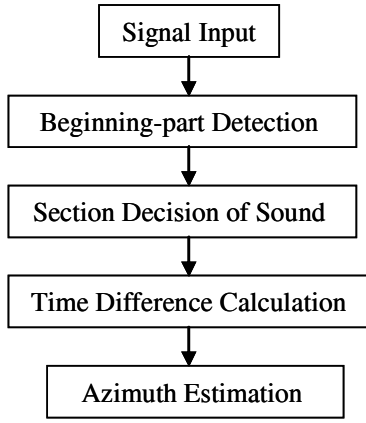
```
Signal Input
    ↓
Beginning-part Detection
    ↓
Section Decision of Sound
    ↓
Time Difference Calculation
    ↓
Azimuth Estimation
```

Figure 4. Procedure of sound localization

3.1 beginning-part detection and section decision of the sound source.

Using short-time energy of the equation (1) and zero crossing rate (ZCR) of the equation (2), beginning-part of the sound source is extracted. To extract correct beginning-part of the sound source, standard value of the energy and ZCR is fixed to experiment value through numerous experiments.

$$E_n = \sum_{m=n-N+1}^{n} x^2(m) \qquad (1)$$

$$ZCR = \frac{1}{2}\sum_{i=0}^{N-1} |\operatorname{sgn}(x(i)) - \operatorname{sgn}(x(i+1))| \qquad (2)$$

$$\operatorname{sgn}(x) = \begin{cases} 1, & x \ge 0 \\ -1, & x < 0 \end{cases}$$

Through the above equation (1) and (2), with beginning-part of the extracted three sound sources, in accordance with the fastest order, two beginning points is set up by the present section of the sound sources to be occurred.

3.2 time difference calculation

The estimation of time delays has been extensively researched in the past. The most common is to use the generalized cross correlation (CC) method. The CC based sound localization method is computationally simple and efficient in comparison with other techniques.

In the decided section of the sound source, the arrival time differences are calculated from the sound wave of two microphones by the cross-correlation method like equation (3).

$$\bar{\tau} = \arg\max\{\bar{\phi}_{ij}(d)\} \qquad (3)$$
$$(i,j) = (0,1),(1,2),(2,0)$$

$$\bar{\phi}_{ij} = \lim_{i \to \infty} \frac{\sum_{k=-i}^{l} x_1(k)x_2(k+d)}{\sqrt{\sum_{k=-l}^{l} x_1^2(k) \sum_{k=-i}^{l} x_2^2(k)}}$$

Also, the arrival time difference for each pair of microphones must satisfy the condition equation (4).

$$|\bar{\tau}| \le D/C \qquad (4)$$

where D is the distance between two microphones, and C is the velocity of sound.

3.3 azimuth estimation

In the figure 3, when the distance of the sound source is much greater than the distance R, the sound wave at the microphone sets can be considered a plane wave, and can be approximated by the following equation.

$$\hat{\theta} = \arccos(\frac{\bar{\tau}}{\Delta T}) - 30 \qquad (5)$$
$$\Delta T = \sqrt{3}R/C$$

The azimuths calculated in the above equation decide to final azimuth according to the present section of the sound source. That is, in the figure 3, microphone section between 0 and 1 represents from 0 to 120 degrees and microphone section between 1 and 2 represents from 120 to 240 degrees and microphone section between 2 and 0 represents from 240 to 360 degrees. Within the above azimuth range, the final azimuth is decided to according to position of the present sound source.

4. RESULTS

4.1 experimental setup

Sound source database tested in this paper was conducted in silent 2 − dimensional surrounding environment by wamirack 192L digital audio recording interface like the figure 2. An empty room

was used to conduct experiments on sound source localization only. And the database was recorded into 3 - channel at the real time in about 5m. Also, sound source is the hand clapping sound, which has 44.1 kHz sampling rate and is quantized by 16bit. Each sound had a period of about 8 seconds. In this experiment, the positions of sound sources covered a 360-degree range of azimuth and hand clapping sound is occurred in twelve positions on the same horizontal plane as the microphone sets. The number of the sound source is 60 items and these are tested by the five times at ten positions.

4.2 results

In this experiment, table 1 shows the results of sound localization using 3 microphone pairs with minimum time differences in each position. The results of azimuth estimation were satisfactory. For the hand clapping sound, the maximum error of azimuth estimation was only 6 degrees. We conducted experiment where the sound source was used to as beacons for robot navigation. But this experiment could be the first step in archiving the ultimate goal in which a robot can recognize the environment with sound.

Table 1. Azimuth estimation result

Original degree	Average degree (in 5m)	Average error
0	0.00	0.00
30	34.09	4.09
60	59.27	0.73
90	88.66	1.34
120	124.66	4.66
150	144.06	5.94
180	183.69	3.69
210	216.24	6.24
240	244.66	4.66
270	272.45	2.45
300	297.05	2.95
330	325.91	4.09

5. CONCLUSION

Necessity of sound source tracing technology is argued in chapter 1. And sound source tracing algorithm is proposed to the shorter execution time than that of an existing algorithm in focusing on the performance improvement of real-time service robot in chapter 2.

Sound source tracing algorithm showed almost correct result. Its error range was within ± 2 degree. And if it is improved in consideration of the altitude and reverberation or echo environment of the sound source, then better efficiency will be expected.

The actual scenario of sound source tracing algorithm researched by this paper is as follows: user calls the robot by voicing or clapping. If robot recognizes sound such as opening door, footstep, or breaking object, including above sounds, and so on, then moves to the position of the sound generation, next it brings some information to public service control room by using network.

Therefore, this paper is implemented as a basic research of the intelligent robot's technical development based on network. It will act important roles in functional reinforcement of the service robot such as information helper and public helper which have abilities that recognize and foretell the surroundings. Also, sound source tracing technology of the robot such as housekeeper robot, military robot, and entertainment robot will be applied to technology transfer and it will contribute on work of robot and protection of human and robot.

REFERENCES

Jie Huang; Supaongprapa, T.; Terakura, I.; Ohnishi, N.; Sugie, N;(1997), *Mobile Robot and Sound Localization*, IROS '97, Proceedings of the 1997 IEEE/RSJ International Conference on , Volume: 2 , 7-11 Sept. 1997.

Jie Huang; Ohnishi, N.; Sugie, N.;(1998), *Spatial Localization of Sound Sources : Azimuth and Elevation Estimation*, IMTC/98. Conference Proceedings. IEEE , Volume: 1 , 18-21 May 1998.

Valin, J.-M.; Michaud, F.; Hadjou, B.; Rouat, J.;(2004), *Localization of Simultaneous Moving Sound Sources for Mobile Robot using a Frequency-Domain Steered Beamformer Approach*, ICRA '04. 2004 IEEE International Conference on , Volume: 1 , April 26-May 1, 2004.

Guodong Guo and Stan Z. Li (2003), *Content-Based Audio Classification and Retrieval by Support Vector Machines*, IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 14, NO. 1, JANUARY. 2003.