

An Acoustic Localization System Using Microphone Array for Mobile Robot

Peng Yang^{1,2,*} Hao Sun^{1,2} Linan Zu^{1,2}

¹ School of Electrical Engineering and Automation,
Hebei University of Technology,
Tianjin, 300130, China

² Hebei Engineering Research Center of Control,
Hebei University of Technology,
Tianjin, 300130, China

*Corresponding author's Email: yphebut@yahoo.com.cn

Abstract: Robot audition system can greatly improve the adaptability of mobile robots especially in the case of robot vision system can't work effectively. So an acoustic localization system using microphone array for mobile robot is proposed in this paper. In order to localize the sound source target, the microphone array used in the system consists of four microphones which were placed respectively on the vertexes of regular tetrahedron because of sound's omni-directional property. According to the plane wave model of sound's propagation and the spherical wave model of sound's propagation, two different localization algorithms based on time delay of arrival are developed in this paper. The hardware and software of the experiment platform are also introduced in this paper. The experimental results show that the acoustic localization system can make the mobile robot rapidly and accurately reach the location of the sound source target.

Keywords: Robot audition; Sound source localization; Time delay of arrival

1. Introduction

With the development of robot technology, the applications of mobile robot are more extensive and the application backgrounds of mobile robot are more complicated. According to this situation, the improvement of the mobile robot's adaptability is required. Comprehensive utilization of multiple informations is an effective mean to improving the mobile robot's adaptability and is also an important sign of robot intelligence. So robot audition which is an important mean of environment detection and man-machine interaction is increasingly concerned. From a biological point of view, audition is one of the most important senses used by humans and animals to recognize their environment and communicate with their own species [1].

An artificial audition system for a mobile robot can be used for three purposes: *i.* localizing sound sources; *ii.* separating sound sources in order to process only signals that are relevant to a particular event in the environment; *iii.* processing sound sources to extract useful information from the

environment [2]. This paper focuses on the first purpose.

Most of the mobile robot's localization systems are based on robot vision, and robot vision is also a hot spot in the research of robotics. Camera which is the most popular visual sensor is widely used for the localization of mobile robot just now [3].

However some difficulties occur because of the limitation of camera's visual field and the dependence on light condition. If the target is not in the visual field of camera or the lighting condition is poor, the visual localization system of the mobile robot cannot work effectively. In many situation, the target which is cannot detected by visual sensor is a sound source. For example, the survivors under the ruins after a earthquake, the pipeline leakage point surrounding the fog which is produced by the chemical emission, and so on. In these cases, the most useful information is given by audition. The research of robot audition for localization is of great value.

Although the spatial resolution is relatively low compared with that of vision, audition has several

unique properties that shown in the following Table 1 [4].

Table 1. Audition versus vision

	Audition	Vision
Directivity	Omni-directional	Directional
Time resolution	High	Low
In darkness	Works	Not works
Occlusion	Not occurs	Occurs
Computation	Little	Much

Because of these unique properties, an acoustic localization system using microphone array for mobile robot is proposed in this paper in order to compensate for the limitations of mobile robot vision system. Two different localization algorithms based on time delay of arrival are developed to guarantee the rapidity and accuracy of the system in this paper. The hardware and software of the experiment platform are also introduced in this paper. The experimental results show that the acoustic localization system can make the mobile robot rapidly and accurately reach the location of the sound source target.

2. System overview

The proposed localization system, as show in Fig.1, is composed of three parts: *i.* microphone array unit; *ii.* pretreatment unit; *iii.* localization unit.

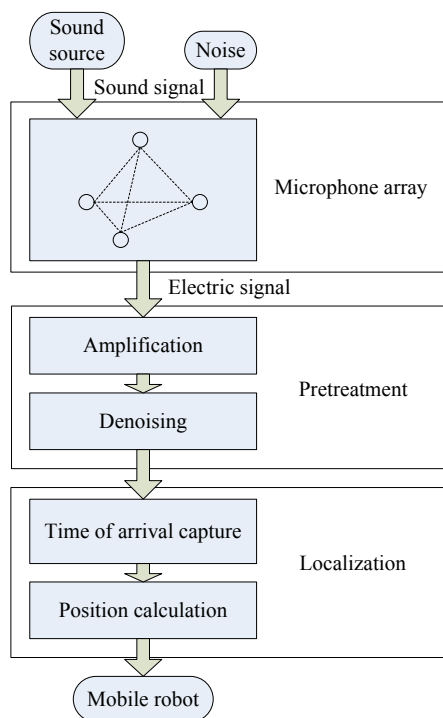


Figure.1 Model of the acoustic localization system.

The sound source makes a sound in noisy

environment. The sound signal which consists of the target's sound and noise is converted into the electric signal with the regular tetrahedral microphone array unit. Through amplification and denoising in pretreatment unit, the electric signal of target's sound can be processed by localization unit. In localization unit, the system capture the arrival times of four channels. Then according to the two proposed localization algorithms and the calculated time delays of arrival, the azimuth of the sound source can be worked out and the mobile robot can find the direction to the target.

3. Localization algorithm

Considering the capabilities of omni-directional positioning, the real-time performance, the convenience of installation and other issues, the regular tetrahedral microphone array which consists of four microphones was used in the proposed system. This structure can effectively avoid the impact of sound speed [5].

3.1 Algorithm base on the plane wave model

When sound waves spread, sound source is considered as a core of spherical wave[6, 7]. If the distance from the microphone array to the sound source is far bigger than the microphone array's size, sound waves can also be seen as parallel wave to the calculation [8], and the complexity of computing will be greatly reduced while the accuracy requirements of the system are met. Only three microphones are used in this algorithm and shown in Fig.2.

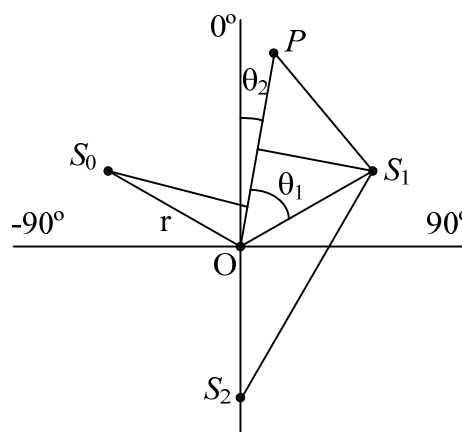


Figure.2 The equilateral triangle microphone array.

in Fig.2, P is the sound source target, S_i is the microphone, d_i is the distance from the target P to microphone S_i . θ_1 is the horizontal angle between the sound source and S_1 . θ_2 is the angle between the

sound source and the perpendicular bisector of microphone S_0 and microphone S_1 . Prescribe the perpendicular bisector of microphone S_0 and microphone S_1 is the front direction of robot which is far away from S_2 . r is the length of OS_i .

So time difference meets the following conditions:

$$\Delta t \leq \frac{D}{V} \quad (1)$$

Where, Δt is the time difference between two microphones. D is the distance between the two microphones, V for the speed of sound. d_0, d_1 and d_2 is the distances between the sound source and microphone. Assume d_j is larger than d_i . So

$$\Delta t = \frac{d_j - d_i}{V} \quad (2)$$

According to the geometric relationship, we have

$$r \cos \theta_1 - r \cos(120^\circ - \theta_1) = \Delta t \cdot V \quad (3)$$

$$\theta_2 = 60^\circ - \theta_1 \quad (4)$$

Substituting Eq.(4) into Eq.(3), we have

$$\theta_2 = \arcsin\left(\frac{\Delta t \cdot V}{\sqrt{3}r}\right) \quad (5)$$

While the sound first reach microphone S_1 , the sound source angle is θ_2 . If the sound first reach microphone S_0 , so the sound source angle will be $-\theta_2$. That means the sound source at the left of the robot.

3.2 Algorithm base on the spherical wave model

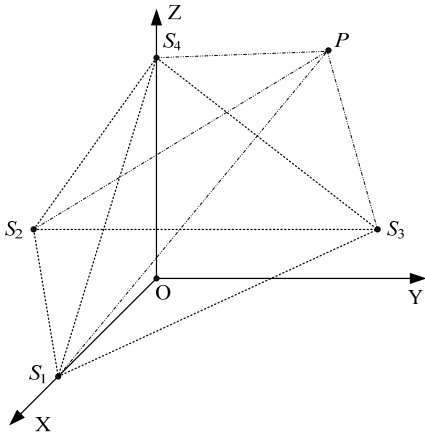


Figure.3 The regular tetrahedral microphone array in rectangular space coordinate system.

The algorithm base on the plane wave model can work well when the target is far away from the microphone array and when the elevation is zero or negligible. However, when the mobile robot come close to the target, the elevation often increases and the algorithm will become invalid. So we develop

the algorithm base on the spherical wave model.

The microphone array is shown in Fig.3. In Fig.3, S_1, S_2, S_3 and S_4 are the vertexes of regular tetrahedron and are also the positions of microphones. P is the position of the sound source. The edge length of the regular tetrahedron is a . The zero point O of the rectangular space coordinate system was placed at the center of triangle $S_1S_2S_3$. Triangle $S_1S_2S_3$ is on XOY plane. S_1 is on X -axis and S_4 is on Z -axis. The supposed coordinates of the sound source is $P(x, y, z)$ and the microphones' coordinates are as follows.

$$S_1\left(\frac{\sqrt{3}}{3}a, 0, 0\right); S_2\left(-\frac{\sqrt{3}}{6}a, -\frac{1}{2}a, 0\right);$$

$$S_3\left(-\frac{\sqrt{3}}{6}a, \frac{1}{2}a, 0\right); S_4\left(0, 0, \frac{\sqrt{6}}{3}a\right).$$

According to the sound spherical wave propagation model and the geometric relationship of the model, the equations can obtained as follows.

$$\left(x - \frac{\sqrt{3}}{3}a\right)^2 + y^2 + z^2 = (r_4 + d_{14})^2 \quad (6)$$

$$\left(x + \frac{\sqrt{3}}{6}a\right)^2 + \left(y + \frac{a}{2}\right)^2 + z^2 = (r_4 + d_{24})^2 \quad (7)$$

$$\left(x + \frac{\sqrt{3}}{6}a\right)^2 + \left(y - \frac{a}{2}\right)^2 + z^2 = (r_4 + d_{34})^2 \quad (8)$$

$$x^2 + y^2 + \left(z - \frac{\sqrt{6}}{3}a\right)^2 = r_4^2 \quad (9)$$

In these equations, r_i is the distance between the sound source P and the microphone S_i . And d_{ij} is equal to r_i minus r_j .

Eq.(7) minus Eq.(8), we have

$$y = \frac{2r_4(d_{24} - d_{34}) + d_{24}^2 - d_{34}^2}{2a} \quad (10)$$

Eq.(7) minus Eq.(6), we have

$$\sqrt{3}ax + ay = 2r_4(d_{24} - d_{14}) + d_{24}^2 - d_{14}^2 \quad (11)$$

Substituting Eq.(10) into Eq.(11), we have

$$x = \frac{2r_4(d_{24} + d_{34} - 2d_{14}) + d_{24}^2 + d_{34}^2 - 2d_{14}^2}{2\sqrt{3}a} \quad (12)$$

Suppose φ is the azimuth of the sound source, then we have

$$\begin{aligned} \operatorname{tg} \varphi &= \frac{y}{x} \\ &= \sqrt{3} \frac{2r_4(d_{24} - d_{34}) + (d_{24}^2 - d_{34}^2)}{2r_4(d_{24} + d_{34} - 2d_{14}) + (d_{24}^2 + d_{34}^2 - 2d_{14}^2)} \\ &= \frac{\sqrt{3}(d_{24} - d_{34})(2r_4 + d_{24} + d_{34})}{(d_{24} - d_{14})(2r_4 + d_{24} + d_{14}) + (d_{34} - d_{14})(2r_4 + d_{34} + d_{14})} \end{aligned}$$

Suppose the distance between the sound source

and the microphone array is far greater than the size of the microphone array, then we have

$$\varphi \approx \arctg\left(\sqrt{3} \frac{d_{24} - d_{34}}{d_{24} + d_{34} - 2d_{14}}\right) \quad (13)$$

According to the range of \arctg and the relations of r_1 , r_2 and r_3 , we have the final result. When $\varphi < 0$, $r_2 < r_1$ and $r_2 < r_3$, the result is $\varphi - \pi$. When $\varphi > 0$, $r_3 < r_1$ and $r_3 < r_2$, the result is $\varphi + \pi$. In rest conditions, the result is φ .

3.3 Simulation results of the second algorithm

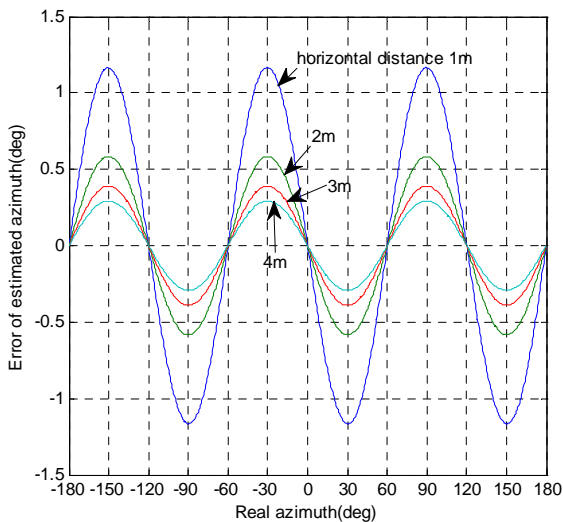


Figure.4 The error of azimuth at elevation zero.

Fig.4 shows the error similar to the sine function, and the max error appears at $k \cdot 60^\circ + 30^\circ$ ($k=0, \pm 1, \pm 2, \dots$). And with the increasing of the distance from the target to the microphone array, the error is becoming little.

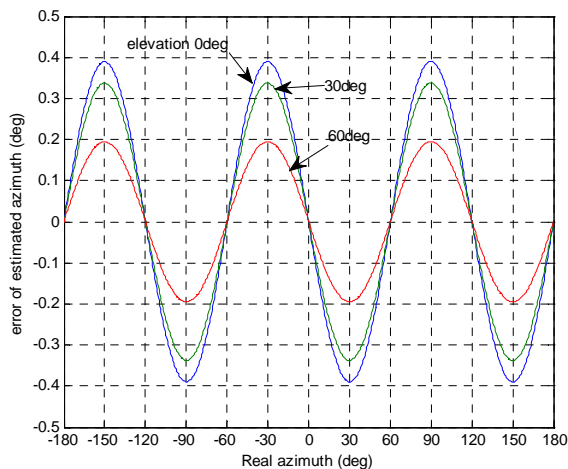


Figure.5 The error at the horizontal distance 3m.

Fig.5 shows the error is reduced while the elevation is increasing under the condition of

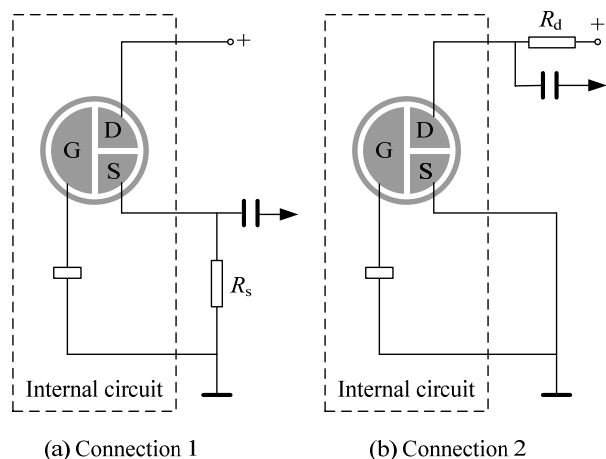
constant distant. And the accuracy of estimation is very high.

4. Experiment platform

In this section, the design of the experiment platform is introduced.

4.1 Microphone array unit

Sound signal collection is a key part of the auditory system. Sound signal collection as the start of the measurement, requires not only a better sensitivity but also has better anti-interference ability. Electret microphone is a micro-electric sound transducer, belongs to the most commonly used condenser microphone. This microphone has a high sensitivity, wide range of frequency response and the advantages of small size. The system selected electret microphone as sound sensor, using it to convert voice signal into electrical signal.



(a) Connection 1

(b) Connection 2

Figure.6 The electret microphone circuit connections.

Electret microphone in the circuit has two kinds of connections: source output and drain output. The impedance of source output is 2k less than the output, circuit more stable, dynamic range, but the output signal is smaller than the drain output. Drain output has voltage gain, so the microphone sensitivity is higher than the source output, but dynamic range of circuit is slightly smaller. Circuit connection is shown in Fig.6. R_s and R_d are determined by the size of the supply voltage. Generally these value can be chosen between 2.2k-5.1k, such as power supplied voltage of 6V, R_s for 4.7k, R_d for 2.2k. In output circuit, if positive ground power is supplied, you could also get a source and drain output by only change D and S for each other. Regardless source output or drain output, the work voltage of the electret microphone must be DC, as it

equipped with its internal FET.

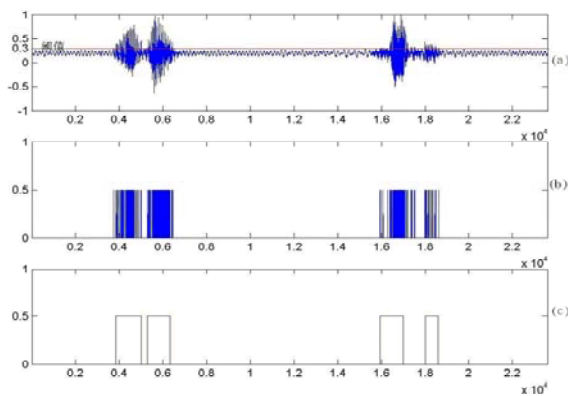


Figure.7 The signal processing in pretreatment unit.

Actually, the sensitivity of different microphone can not be exactly same, and there are differences in magnification of the same transistor and the resistance, it makes the difference of four channels and then reduces the accuracy of the system. In order to ensure the accuracy of the voice endpoint detection, when you select components [9], you should ensure the original device of each road consistent as far as possible. At the same time, use white noise as a sound source and place each microphone in the same location, so measure the average response rate separately. By measuring, choose the original group whose basic rate of response is basically the same. After such calibration, the accuracy of sound source position was improved.

4.2 Pretreatment unit

Pretreatment unit include two parts. They are voice signal amplification and denosing [10-13].

4.2.1 Signal Amplification

Because of the limited range of the microphone's detection, when the sound source is far away from the microphone, the signal microphone collected is small, and is not conducive to algorithm treatment. So there is a requirement to enlarge analog signal which microphone captured [5,14]. The voice detection circuit, transistor are selected here with the structure of two level enlarge. A reasonable selection of supply voltage, as well as client base and the collector resistance value, so that the changes in signal amplitude of the voice is obvious in order to facilitate follow-up treatment. One capacitance is used to isolate DC. Two resistors compose pressure circuit in order to set threshold voltage. The voice signal amplified output at the A-point, which and the threshold voltage B point are the two inputs of the comparator.

4.2.2 Denosing

That the target sound exists whether or not is the standards of the robot auditory system mentioned in this paper. The deference between the sound source signal and background sound marks the presence of the target sound. So the amplified voice signal should be compared to the pre-set background sound intensity threshold in order to determine whether there is the target sound or not. If the judge is that there is no sound currently, the auditory system does nothing to deal with, which will improve the efficiency of the system. Signal that before comparing as shown in Fig.7 (a). The voice signal which is after comparing is shown in Fig.7 (b). The signal after waveform shaping is a continuous high-low level signal as shown in Fig.7 (b). Since the position unit only need to detect voice start time. Use flip-latch to do maintain treatment to facilitate the following time capture as shown in Fig.7 (c).

4.3 Localization unit

4.3.1 Time of arrival capture

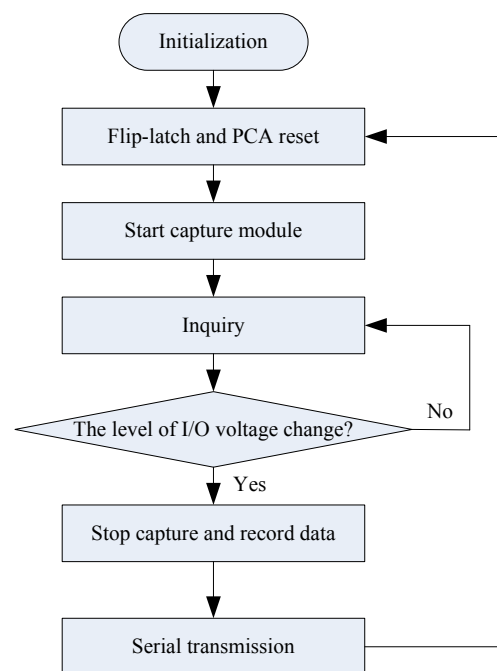


Figure.8 The program flow chart.

In order to accomplish the sound capture and transmission, single-chip 8-bit microcontroller is used here. The programmable counter array of the single-chip 8-bit microcontroller has five 16-bit capture or compare modules special timer. Each module can be programmed to achieve the following four kinds of model: capture mode, software timer mode, high-speed PWM output mode and PWM mode. Here four capture or compare modules are used, and be

set to capture mode. The clock is set a quarter to single-chip CPU clock frequency 11.0592 MHz which can fully meet the experiment's requirements. When sound arrives, the flip-latch produces a pulse, and maintains the level. Capture module to capture the arrival time of the pulse and storage. When get all the four channel times of the microphone array, the data reached PC using serial communication of single-chip for position calculation [15]. The flow chart of Single-chip program is as shown in Fig.8.

4.3.2 Position calculation



Figure.9 The program Interface of sound source localization and Robot control program Interface.

In order to verify the validity and correctness of algorithms, this module used c++ language to do simulation at pioneer3-mobilesim simulation platform. The sound source localization and robot control program interface was shown in Fig.9.

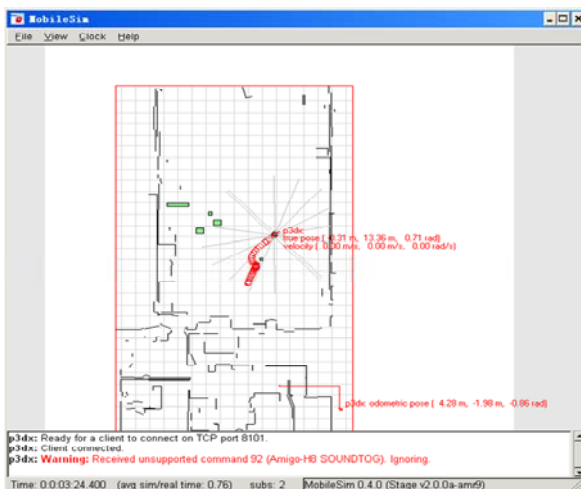


Figure.10 Simulation interface of robot toward the target sound source.

The simulation interface of robot toward the

target sound source was shown in Fig.10. Pioneer3-DX robot before and after has a total of 16 sonar sensors, can be used to obtain obstacle information in all directions. In the map, around the robot is the robot-ray sonar which can detect the obstacle distance. Grid is robot path. Position the robot to stop is the robot's sound source goal to arrive. The other side of the grid is the starting point for the robot.

In the experimental procedure, the serial port for receive and position calculation module, the robot toward the target module, are three separate threads. First of all, the serial to accept data and calculation the sound source position, and then robot toward the target sound source position thread will begin go function to toward the target. In experimental procedures collision avoidance module and toward the target module have different priority. If there are obstacles on the course toward the sound source target, robot will execute the high priority of collision avoidance module to avoid obstacles first, then continuous go to the sound source.

5. Results

Table 2. Azimuth results with three microphones positioning system at elevation zero

Real azimuth (deg)	Estimated azimuth (deg)
0	0.583143
45	47.280518
60	60.619927
90	94.849625
105	107.871594
135	137.590667
180	180.438402
-120	-120.35564
-90	-93.992458

Here we give two sets of data got from experiment. They are got from the serial port for receive and position calculation module. First, we used the localization algorithm based on the plane wave model of sound's propagation and obtain the first set of data. Then we replace the localization algorithm based on the plane wave model of sound's propagation with the localization algorithm based on

the spherical wave model of sound's propagation. So we obtain the second set of data. In these experiments, the azimuth of sound sources covers a +180 range of azimuth and the distance from the center of the microphone array to the sound source is 1m. The elevation is zero. The system is tested in fifteen azimuths on the same horizontal plane as the microphone array. The azimuth is estimated five times and is averaged. There shows the estimation of nine typical azimuths in Table 2.

From Table 2 we can calculate the errors of all the localization results and get some conclusions as below. To the localization of the same distance, when the sound source lies on each midline of the microphone array shown, the error is minimal. The reason is as below. When calculating the time difference, we adopt the actual distance from the sound source to each microphone. But, to facilitate the calculation, when constructing the calculation model of the azimuth, we assume that the sound source is at the same plane with the microphones and the sound wave is looked as a plane wave. Then, we use a uniform model to approximate the relationship of different angles. Moreover, this model accords with the calculation model when the sound source lies on the midline of the microphone array, so the error is minimal. Another is that to the localization of the same distance, when the angle of the sound source approaches the angle of the midline of the microphone array (such as 30°, 90°, 150°, -150°, -90°, -30°), the error increases gradually instead of reducing gradually.



Figure.11 The mobile robot with the proposed acoustic localization system.

As following, the second set of data is given. This set of data has two tables. They are obtained by

the localization algorithm based on the plane wave model of sound's propagation. In Fig.11, there is a mobile robot with auditory navigation system. It is three-dimensional space positioning system, and has four microphones.

Table 3. Azimuth estimation results with four microphones positioning system at elevation zero

Real azimuth (deg)	Estimated azimuth (deg) when distance is 1m	Estimated azimuth (deg) when distance is 4m
0	0.504561	1.875546
30	30.587858	30.846923
60	59.707523	60.502375
90	89.858735	88.357609
120	121.012795	120.418715
150	150.105342	150.974111
180	-175.767628	-177.317044
-120	-120.664863	-120.873342
-90	-91.069087	-92.576985

Table 4. Azimuth estimation results with four microphones positioning system at elevation thirty

Real azimuth (deg)	Estimated azimuth (deg) when distance is 1m	Estimated azimuth (deg) when distance is 4m
0	-0.695941	-1.224041
30	27.874824	30.514932
60	60.633709	59.625084
90	87.635379	88.041292
120	121.335417	123.637362
150	144.036173	151.211440
180	-179.389380	177.368776
-120	-120.877765	-121.669024
-90	-92.365543	-91.364346

In this experiment, the azimuth of sound sources also covers a +180 range of azimuth and the distance from microphone to the sound source is 1m and 4m. And the elevation is not only zero but also 30 degrees. Each elevation has a table of data. The system is tested in fifteen azimuths on the same elevation as the microphone array. Each azimuth is estimated five times and is averaged. There shows the estimation of nine typical azimuths in Table 3

and Table 4.

From Table 3 and 4, we can calculate the errors of all the localization results and get some conclusions as below. To the localization of the different distance at the same elevation, error is not much change. To the localization of the same distance but at different elevation, error is not much change yet. So we can see that this model for angle of elevation is relatively stable. It had a strong adaptability. We can see from the data in the table that in addition to the error in one or two azimuths is greater, other errors are less than three degrees. Taking the error generated by sound source location into account, as well as the errors generated by microphone installed, this result is very satisfactory. Compared with the data in Table 2, we can see that four microphones positioning system is more accurate than three microphone positioning system. Even when there is elevation, the accuracy of the data can be met. And it is more stable. And all above have demonstrated and proved that the localization algorithm based on the spherical wave model of sound's propagation has a more powerful advantage than the localization algorithm based on the plane wave model of sound's propagation in sound source positioning.

The programs of the two algorithms all develop in C language. By using the function GetTickCount(), we get the ticks of running the each program one million times. The algorithm base on the plane wave model costs 24 ticks and the algorithm base on the spherical wave model costs 30 ticks. Because 1 tick equals to 55ms, we can get that to run the program of the first algorithm once needs 1.32 μ s and to run the program of the second algorithm once needs 1.65 μ s. So the rapidity of the algorithms is good and the first one is better. The experiment condition is that the frequency of CPU in the robot is 1GHz and the operation system is Windows 2000 Professional.

According to the different tasks, we can select the proper localization algorithm. If the sound source target we need to localize and the microphone array are all at the same horizontal plane, and the distance between them is far, we should choose the algorithm base on the plane wave model because of the less number of microphone and the less computation. Otherwise, the algorithm base on the spherical wave model should be chosen because of the better accuracy and applicability.

6. Conclusion

The target localization system based on auditory

designed by this paper can acquire the sound information effectively and calculate the position of target sound accurately, which can make up the limited field of view sensor of other robots and light tight barrier limitations. Simulate biology auditory system, use perceptual information around robot efficient, implement target localization based on auditory, and do experiments on Pioneer3 Mobilesim simulation platform, obtained a good experimental result.

Acknowledgments

This work was supported by Hebei Province Science and Technology Research and Development Guide Plan (072135179). And the Research Institute of Prosthetics & Orthotics of the Ministry of Civil Affairs of P. R. China provided great aids on the experiments.

References

- [1] S. Yamamoto, K. Nakadai, H. Tsujino, T. Yokoyama, and H. Okuno, "Improvement of robot audition by interfacing sound source separation and automatic speech recognition with missing feature theory", In: *Proceedings IEEE International Conf. on Robotics and Automation*, 2004, pp.1517-1523.
- [2] J. Valin, F. Michaud, J. Rouat, "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering", *Robotics and Autonomous Systems*, 2007, (55):216-228.
- [3] K. Nakadai, H. G. Okuno, H. Kitano, "Real-time sound source localization and separation for robot audition", In: *Proceedings IEEE International Conference on Spoken Language Processing*, 2002, pp.193-196.
- [4] J. Huang, T. Supaongrappa, I. Terakura, F. Wang, N. Ohnishi, N. Sugie, "A model-based sound localization system and its application to robot navigation", *Robotics and Autonomous Systems*, 1999, (27):199-209.
- [5] Y. Jin, R. Yang, "Research status and prospect of the acoustic localization techniques", *Elementary Electroacoustics*, 2007, (31):4-8.
- [6] Z. Lin, B. Xu, "Sound source localization based on microphone array", *Electro-acoustic technology*, 2004, (5):19-24.
- [7] M. S. Brandstein, *A Framework for Speech Source Localization Using Sensor Arrays*, Brown University, doctor thesis, 1995.
- [8] R. Irie, *Robust sound localization: An application of an auditory perception system for a humanoid robot*, MIT Department of Electrical Engineering and Computer Science, Master thesis, 1995.
- [9] S. Yamamoto, K. Nakadai, H. Tsujino, H. Okuno, "Assessment of general applicability of robot

audition system by recognizing three simultaneous speeches”, In: *Proceedings IEEE International Conf. on Intelligent Robots and Systems*, pp.2111-2116, 2004.

- [10]P. Aarabi, S. Zaky, “Robust sound localization using multi-source audiovisual information fusion”, *Infotmation Fusion*, 2001, (2):209-223.
- [11]Hiroshi G. Okuno, T. Ogata, K. Komatani, K. Nakadai, “Computational auditory scene analysis and its application to robot audition”, *Informatics Research for Development of Knowledge Society Infrastructure*, 2007, (29): 69-76.
- [12]M. Fujita, Y. Kuroki, T. Ishida, T. Doi, “Autonomous behavior control architecture of entertainment humanoid robot SDR-4X”, In: *Proceedings IEEE/RSJ International Conf.* , pp.960-967, 2003.
- [13]P. Stoica, K.C. Sharman, “Maximum likelihood methodfor direction-of-arrival estimation, In: *IEEE Trans.Acoust: Speech Signal Processing*, pp.1131-1143, 1990.
- [14]J. Huang, N. Ohnishi, X. Guo, N. Sugie, “Echo avoidancein a computational model of the precedence effect”, *Speech Communication*, 1999, (27) : 223-233.
- [15]J. M. Valin, F. Michaud, “Robust sound source localization using a microphone array on a mobile robot”, In: *Pceedings IEEE/RSJ*, 2003, pp.23-30.