2D SOUND SOURCE MAPPING FROM MOBILE ROBOT USING BEAMFORMING AND PARTICLE FILTERING

S. Kagami[1,2,3], *S. Thompson*[1,3], *Y. Sasaki*[2,1], *H. Mizoguchi*[2,1], *T. Enomoto*[4]

1: Digital Human Research Center, Nat. Inst. of Advanced Industrial Science and Technology 2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan.

2: Dept. of Mechanical Engineering, Tokyo University of Science

3: JST, CREST

4: Kansai Electric Power Co., Inc.

ABSTRACT

This paper describes a particle filter based sound source mapping system that builds 2D sound source maps from directional sound readings taken from a mobile robot. The method uses a sound source localization model that is represented by gaussian distribution for both direction and distance. To do this, accurate directional localization of sound sources is required, and two key components have been developed to achieve this: 1) a 32ch low side-lobe microphone array that is designed by beam forming simulation to have a) an omnidirectional response, b) a narrower main-lobe, and c) lower side-lobes, in 700-2500[Hz] acoustic signals; 2) directional localization of different pressure sound sources by combining the Delay and Sum Beam Forming (DSBF) and the Frequency Band Selection (FBS) methods. Finally, experimental results show the proposed method can map sound sources in two dimesionals with high accuracy (less than 50[cm] error).

Index Terms— Array signal processing, sound source mapping, mobile robots, array design

1. INTRODUCTION

A sound source mapping function is vital for a robot that operates in a human environment. Bearing only Simultaneous Localization and Mapping (SLAM) technique has been actively investigated in the last several years, mainly applied to optical sensors(ex. [1]). However sound signals are significantly different in two ways: the audio signal used for directional localization and the characteristics of the sound source. Difficulties for directional localization for sound is caused by acoustic reverberation, diffraction, resonance, interference, and so on. On the other hand, difficulty caused by the characteristics of the sound source is that the content is usually unknown and always changing in time or even sometimes missing.

Particle filters are widely used in the perception area in robotics to handle noisy input to estimate surrounding map and/or robot location. Several methods have been proposed using particle filtering to achieve directional localization and in tracking in microphone centric coordinates (ex. [2, 3]). As for a mapping function, Nakadai et al.[4] presents a method to map sound source location using a particle filter from a microphone array attached both in a room and on a robot, citing the difficulty in performing sound source mapping from a robot mounted microphone array, especially when the robot is moving. In this paper, we propose a method to achieve 2D mapping by using only a microphone array mounted on a moving robot.

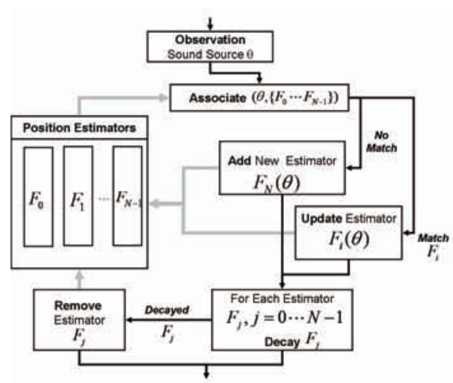


Fig. 1. Flow chart of estimator management for 2D sound source mapping with mulitple sound sources.

2. 2D SOUND SOURCE MAPPING

Two dimensional sound source position estimation is achieved by applying a bearing only state estimation technique. Individual particle filters are used to maintain position estimates of a particle sound source, with the set of particles representing a distribution over the x,y coordinate frame. An unknown number of sound sources can be present in a given environmnet, so the number of estimators must be managed, with sound source estimators being created and deleted as required. Figure 1 shows the flow chart of estimator management.

In addition to this, sound sources also emit signals intermittently, so the activity of sound sources over time needs to be monitored. The estimator monitors activity of a particular sound source by use of a decay mechanism. Signal detection causes 'growth' of the decay value, while an absence of signal causes decay. Once a sound source decays to a given value, the estimator is deleted.

Upon observation of new sound source Obs at time k in a particular direction θ_k, a new particle filter estimator F_N is created and initialized from the current robot location with its particles spread over a 2D Gaussian distribution over the direction estimate θ_k, at a default distance r_D. The variances associated with the distributions are determined by the error in the directional sound source estimate σ_θ^2, and a default large variance in distance σ_r^2, reflecting the absence of distance information in the bearing only observation. Each particle is an estimate as to 2D sound source position $s_i = [x, y]$.

Initiatization, then occurs as follows

1. Sound source $Obs(k) = \theta_k$
2. From robot pose P_R = $[x, y, \theta)]$ initialize particles $S = \{s_0, \ldots, s_{N_P-1}\}$
   For all s_i in S do
   - $r_i = r_D + G(\sigma_r^2)$, $\alpha = \theta_k + G(\sigma_\theta^2)$, where $G(\sigma^2)$ is a function returning a Gaussian distributed random value with variance σ^2
   - $s_i = [(x_R + r_i)\cos(\alpha + \theta_R),$
     $(y_R + r_i)\cos(\alpha + \theta_R)]$

The filter then propagates particles representing the probability density function of the sound source location as follows:

1. Observe sound source $Obs(k) = \theta_k$ from $(x, y, \theta)_R$
2. Disperse S, $s_i(k) = s_i(k-1)+\omega$, where ω is a random, Gaussian motion
3. Measure S, such that $p(s_i(k)) = SM(s_i(k), \theta_k\, P_R)$
4. Resample S with replacement, based on $p(s_i(k))$

where $SM(s_i(k), \theta_k\, P_R)$ is a sensor model returning the probability of observing the sound source at position $s_i(k)$ at angle θ_k from the current robot position. Because of the dynamic content of sound signals only the direction of the sound source is used in the sensor model:

$$SM(s_i(k), \theta_k\, P_R) = G(\theta_k - exp(s_i(k), P_R), \sigma_\theta^2)$$

where $G_N(x, \sigma^2)$ is a normalised Gaussian probability function for deviance x and variance σ^2, and $exp(s_i(k), P_R)$ is the expected angle of observation from robot pose P_R to particle $s_i(k)$. Note as the robot pose changes due to ego motion, the expected angle to sound sources will change accordingly.

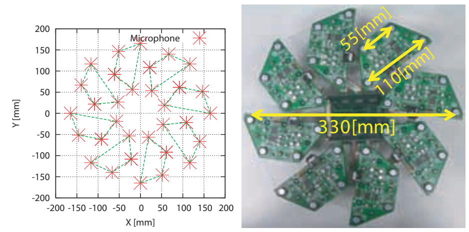


Fig. 2. 32ch microphone arrangement (left) and photo (right)

The dispersal of particles is similar to a 'motion model' and does allow for some motion of the sound source. However, in this work we assume static sound sources and the dispersal of particles is primarily intended to avoid filter convergence on incorrect estimates due to noisy observations.

3. DIRECTIONAL LOCALIZATION

The proposed sound source mapping method can handle noise in directional localization. However, noise should be stochastically small. Therefore, the directional localization system needs to be robust from false positive detection. For this purpose, we have been designing and developing a low side-lobe microphone array that is optimized for the Delay and Sum Beam Forming (DSBF) directional localisation method.

3.1. 32ch Low Side-lobe Microphone Array

In order to calculate sound source direction for audio input with an unknown frequency, we developed a microphone array and firewire interface board.

The diameter of the microphone array is limited to 33[cm] due to our mobile robot size. Through simulation of sound pressure distribution, we empirically decided the microphone arrangement to minimize side-lobes. Fig.2(left) shows the resulting microphone arrangement which consists of the octagonal arrangement of eight 4ch microphone boards that have an isosceles trapezoid shape. Fig.2(right) shows a picture of the array. The system has 16bits resolution for a simultaneous sampling rate of $16kHz$.

Fig.3 shows the beam forming simulation results at 1000, 1400, 2000 [Hz]. At each frequency, the focus direction gain compared to side-lobe is 12[dB] at minimum and 16[dB] in average (from 700-2500[Hz]).

Fig.4 shows simulated and measured directivity pattern of this microphone array. The horizontal axis is direction, and the array is focusing on 0[deg] direction. The vertical axis is signal gain in [dB] compared to the focused direction.

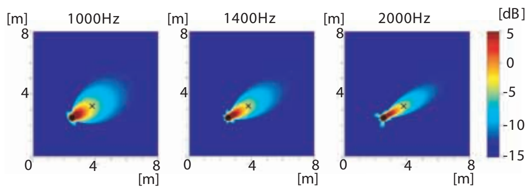


Fig. 3. Beamforming simulation result at 1000,1400,2000[Hz]

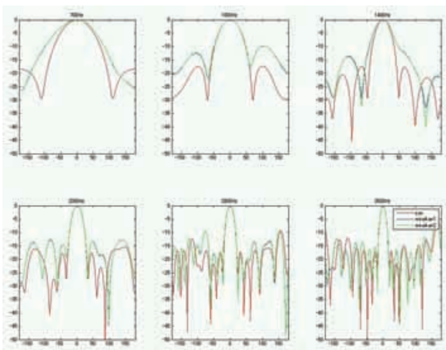


Fig. 4. Simulated (red) and measured (green&blue) directivity pattern at our office

3.2. Frequency Band Selection Method

The DSBF method has limited performance, especially the method does not remove other signals perfectly (just reduces). Thus, we apply the FBS method[5] after DSBF for the detection of multiple sound sources. FBS is a kind of binary mask and segregates objective sound sources from mixed sound by selecting the frequency components judged to be from a common objective sound source.

The process is as follows. Let $X_a(\omega_j)$ and $X_b(\omega_j)$ be the frequency components of DSBF-enhanced signals for objective and noise sources, respectively. The selected frequency component $X_{as}(\omega_j)$ is expressed as Equation(1):

$$X_{as}(\omega_j) = \begin{cases} X_a(\omega_j) & \text{if } |X_a(\omega_j)| \geq |X_b(\omega_j)| \\ 0 & \text{else} \end{cases} \quad (1)$$

This process rejects the attenuated noise signal from the DSBF-enhanced signal. The segregated waveform is obtained by the inverse Fourier transform of $X_{as}(\omega)$.

When the frequency components of each signal are independent, FBS can separate the desired sound source. This assumption is usually effective for human voice or every day sound within a short time period.

Fig.5 shows the FBS procedure for multiple sound sources. The first step filters out the average signal of each microphone (no delayed signal) input by FBS and finds the loudest sound from the spatial spectrum. When the frequency component of the average signal is higher than any DSBF-enhanced signal

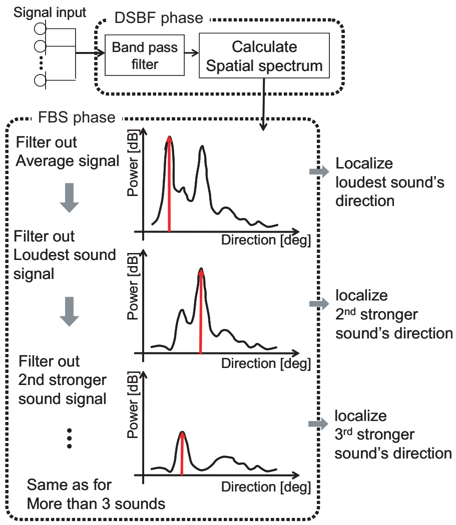


Fig. 5. FBS sound localization process

from each direction, the system filters out the spectrum of that frequency. This process rejects omni-directional noise sounds.

The second step filters out the 1st sound signal by FBS, and finds the second strongest sound from the spectrum. When the frequency component of the DSBF-enhanced signal of the 1st sound's direction is higher than that of any other direction, the system filters out the spectrum at each frequency.

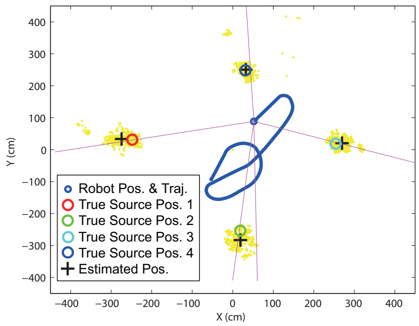
If there are more than two sounds, the system finds the third strongest sound, and so on, after filtering out the second strongest sound signal. The method localizes multiple sounds from the highest power intensity to the lowest at each time step. Then the system can continuously localize multiple sound sources and separate each sound source during movement.

4. EXPERIMENTS

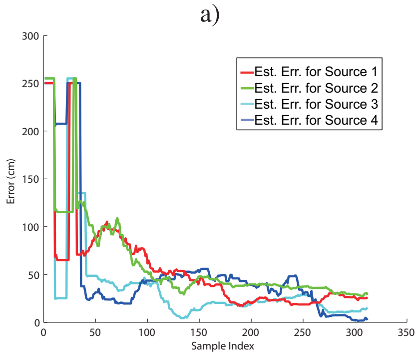
We conducted two experiments using our mobile robot "Pen2". A commercial motion capture system (Motion Analysis Eagle) with 12 cameras measures robot position in 240[Hz] as a ground truth. Standard deviation of robot position measured by this MOCAP system is 0.042[mm] in translation and 1.09e-5[deg] in rotation.

The microphone array locates sound directions at around 12[Hz]. Reverberation time RT_{20} was 167[msec], and back ground noise level was 50[dBA] (mainly fan noise). Signal noise ratio was 20[dBA] for experiment 1, and 15[dBA] for experiment 2. Sound sources were music, male and female voices.

Fig.6 shows the results of localising four sound sources. In this experiment, all the loud speakers are placed on microphone array level. Fig.6(b) shows the convergence of the lo-

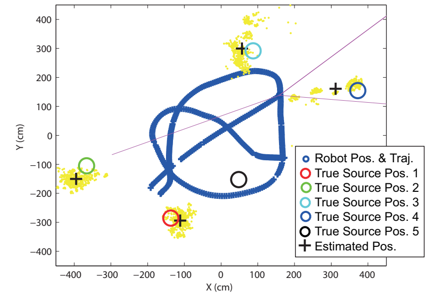


a)

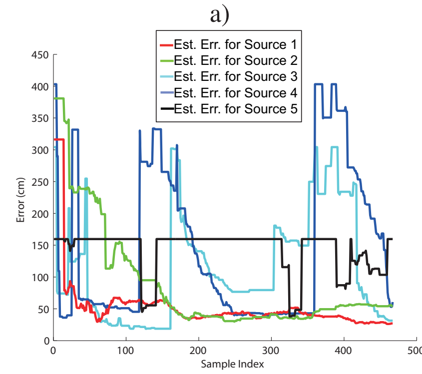


b)

Fig. 6. Experiment 1. a) 2D mapping of four sound sources. b) Sound source mapping error.



a)



b)

Fig. 7. Experiment 2. a) 2D mapping of five sound sources. b) Sound source mapping error

calisation process and the remaining error. After 100 samples (about 8[s]), the system achieves 2D mapping with around 50[cm] remaining error.

Fig.7 shows an experiment with five sound sources. In this case, sources are placed in different height (from 57 to 204[cm]). One source at (67, -145, 204)[cm] is not found at all. It may be placed too high up and only directional localization around yaw axis was conducted. Also the speaker provides a directional sound source, and the robot passes close to, and even behind the sound source, increasing the effects of the high position of the speaker. The remaining four sources are found. Fig.7(b) shows basically the same kind of convergence performance as the previous experiment. Interestingly, some sources are lost and refound as the robot moves throughout the enviornment. At each time a sound source is found, convergence occurs as like before.

5. CONCLUSION

This paper proposed a 2D sound source mapping method while robot is in motion, by applying a particle filter technique. The method is general for any directional localization. Combined with our 32ch low side-lobe microphone array and with Delay and Sum Beam Forming (DSBF) + Frequency Band Selection (FBS) methods, the system can map 2D arrangement of sound sources. Experimental results show after 100 samples, detected sound source locations converge in less than 50[cm].

Since one sound source that is located high above the array height is not mapped well in experiment 2, in the future, we would like to 1) extend our mapping function into 3D, 2) optimize microphone array design for two directional localization, 3) develop a more robust and two directional sound source detection method.

6. REFERENCES

[1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*, The MIT Press, September 2005.

[2] H. Asoh, I. Hara, and H. Asano, "Tracking human speech events using a particle filter," in *In Proceedings International Conference on Audio, Speech and Signal Processing*, 2005, pp. II/1153–1156.

[3] J. Valin, F. Michaud, and J. Rouat, "Robust 3d localization and tracking of sound sources using beamforming and particle filtering," in *In Proceedings International Conference on Audio, Speech and Signal Processing*, 2006, pp. IV/224–227.

[4] K. Nakadai, H. Nakajima, M. Murase, S. Kaijiri, K. Yamada, T. Nakamura, Y. Hasegawa, H. G. Okuno, and H. Tsujino, "Robust tracking of multiple sound sources by spatial integration of room and robot microphone arrays," in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing 2006*, Toulouse, France, May 2006, pp. IV/929–932.

[5] M. Aoki, M. Okamoto, S. Aoki, H. Matsui, T. Sakurai, and Y. Kaneda, "Sound source segregation based on estimating incident angle of each frequency component of input signals acquired by multiple microphones," *Acoustical Science and Technology*, vol. 22, no. 2, pp. 149–157, 2001.