The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Map-Generation and Identification of Multiple Sound Sources from Robot in Motion

Yoko Sasaki[1], Simon Thompson[1], Masahito Kaneyoshi[1], Satoshi Kagami[1]

Abstract— **The paper presents a multiple sound sources mapping system from a robot embedded microphone array. The robot localizes sound direction and recognizes what sound it is while the robot is in motion. Then the system estimates the positions of the sound sources using triangulation from a short time period of directional localization results. Three key components are denoted: 1) accurate directional localization and separation of multiple sound sources using a microphone array 2) separated sound recognition from a several tens of milliseconds input signal 3) sound position estimation using the RAndom SAmple Consensus (RANSAC) algorithm from a tracked sound stream. By combining these techniques, the proposed system provides surrounding sound information: "Where does the sound come from?" and "What is the sound?". It works with short term signal input, and is helpful to initially notice surrounding events.**

I. Introduction

Sound source mapping is a vital function for a robot that operates in a human environment, such as in the home, office room or factory environment. In such situations, there are many sounds/noises as well as the voices of people around a robot. Focusing attention on understanding the surrounding environment from robot sensors, bearing-only Simultaneous Localization and Mapping (SLAM) techniques have been actively investigated in the last several years, mainly applied to optical sensors (ex. [1]). Just like visual information, sound signals are useful to sense surroundings, especially to initially notice an interesting event, for example, a human calling or a slamming door.

However, audio signals are different from vision information in two ways: 1) environmentally susceptible signals for directional localization, 2) varied characteristics of the sound sources. Difficulties for directional localization of sound is caused by acoustic reverberation, diffraction, resonance, interference, and so on. Difficulties caused by the characteristics of the sound source are that the received signal is dynamically changing in time or even sometimes missing.

Accurate directional localization for different pressure sound sources and robust tracking of detected sound sources are important for developing a sound-event mapping system. While many studies have investigated artificial audition [2], [3], challenging problems remain for covering a large area in a varied real environment. For a mobile robot, observed sound sources have varied position relationship such as moving closer, crossing each other, and varied distance to the sound.

1: Digital Human Research Center, National Institute of Advanced Industrial Science and Technology, 2-3-26, Aomi, Koto-ku, Tokyo, 135-0064, Japan. {y-sasaki, s.thompson, m-kaneyoshi, s.kagami}@aist.go.jp

For a mobile robot, application focused on the initial notice of a sound event, sound position estimation in a short time period is an important ability, and such sound position information in global coordinates is helpful for integrating other sensors, enhancing sound-separation capability for moving sound sources.

II. Related Work

Generally, there are two types of sound source mapping. One can obtain sound pressure distribution and the other can detect sound position as a point source. Martinson et al. [4] proposed an auditory evidence grid that considers ambient noise sources in the static environment to filter poor localization results and cover a larger area. This system can generate a static sound distribution map by collecting directional localization data at a variety of robot positions over time. As the latter types of mapping, Nakadai et al. present 2D sound position localization [5] using a particle filter by integrating a room and a robot microphone array. The room microphone array increases the resolution of the localization procedure and its robustness against ambient noise, and it can track two sound sources simultaneously.

The authors have worked on the latter types of mapping [6], [7]. These show the sound mapping strategy in two ways: triangulation based short time estimation, and particle filter based long term estimation. Both systems estimate sound positions using beamforming based sound directional localization result, but have limitation on covered area, subject sound (assuming continuous signal), and do not correspond to crossing condition.

In this paper, we focused on sound position estimation in a short time period from robot embedded sensor and propose a multiple sound sources mapping system by using a microphone array installed on a mobile robot. Based on the problem in past work, the proposed system provides the location of surrounding sound in varied condition and can identify what sound it is by combining robust directional localization and sound recognition from short time signal input. The proposed system provides sound position information with sound label around the robot in global coordinates, and the information is continuously updated in response to situation changes such as disappeared sounds. The result show the possibility to add sound information on SLAM framework by using robot embedded microphones.

III. Sound Identification from a Mobile Robot

This section provides an overview of the sound localization, separation and recognition methods which are used in

978-1-4244-6676-4/10/$25.00 ©2010 IEEE

the multiple-sound-sources mapping system.

Fig. 1 shows the calculation flow of our robot audition system. In the DSBF phase, scanning the focus obtains the spatial spectrum which indicates the sound pressure distribution around the robot. MLF-based sound localization is calculated using the spatial spectrum. Then, the FBS phase separates localized sound sources using the DSBF-enhanced signals. Finally, separated sound sources are recognized using PCMs sound identification, and the sound directions with their sound labels (recognition result) are observed at each time step. Each functions are explained below.

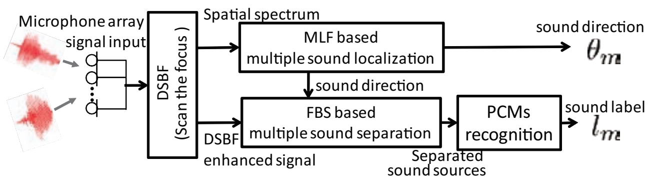


Fig. 1. Calculation flow of the auditory system at a time slot

A. Multiple Sound Localization and Separation

Sound localization and separation is based on Delay and Sum Beam Forming (DSBF) using a microphone array, to achieve a directional localization of multiple sound sources from a robot in motion,

Main-Lobe Fitting (MLF) [8] is used for multiple sound source localization. The method detects point sound sources using the main-lobe model of a microphone array. Selecting reliable peaks by MLF rejects deformed peaks caused by reflection or interference, or deformed peaks between two close sound sources. The main-lobe model is obtained from the microphone arrangement of the array, and MLF-based sound localization is suitable for mobile robot whose acoustic environment changes after the robot's motion. In addition, it can avoid erroneous localization when direction of two sounds are close together.

Frequency Band Selection (FBS) [9] is used for localized sound-source separation. The method assumes that the frequency components of each existing sound source are not overlapped, and separate the target sound sources by selecting larger frequency components from DSBF enhanced signals.

B. Pitch-Cluster-Maps Sound Identification

Pitch-Cluster-Maps (PCMs)-based sound identification [10] is applied to separated sound-source recognition at each time slot. It adopts the Vector Quantization (VQ) approach for real time processing, and the method uses only instant pitch data (not time sequence information). The PCMs sound identification is applicable for several tens of milliseconds of sound input, and is suitable for a mobile robot application in which conditions are dynamically changing.

A binarized frequency spectra of a target sound source are used to generate a PCMs model. These spectra are grouped into K-clusters by the k-means method, and the centered spectrum of the grouped data for each cluster is used as the PCMs model. Sound identification phase is performed as similarity-distance estimation which selects the model which has the minimum distance of binarized frequency spectrum. As a final decision, if the spectral distance is bigger than threshold defined by pitch variance of the model, the input signal is recognized as unknown signal.

IV. Sound Source Mapping

This section provides an overview of the robot motion function we used in experiments, and explains the proposed sound sources mapping system from a robot in motion.

A. Particle-Filter-Based Robot Position Localization

Localization is the process of determining where a robot is within a known map. The framework behind the localization system developed in this work is that of particle filter localization [11]. This widely used method for mobile robot localization uses sampling of the robot's position to approximate its location probability density function (PDF). A motion model is used to predict a sample's future position according to control or odometry information, and the predicted positions are evaluated according to how their expected sensory view matches the current sensor view. Resampling of the predicted positions based on this evaluation ensures the sampled positions remain distributed about the correct localization estimate. For 2D sensory data, the matching process can be done efficiently enough to allow a sufficient number of particles (or samples) to accurately approximate the PDF and ensure continuous localization.

B. Path Planning and Path Following

Path planning of the mobile robot is achieved using the method proposed in [12]. The method uses a 2D A* path planner [13] to generate the optimally shortest path through a 2D grid map to a goal location. A local subgoal is created along the 2D path (at about 5 m) and is used as the target for generating cubic and fourth-order curvature polynomials. These local trajectories are continuous in curvature and allow the robot to make smooth motions through the environment.

Continuous curvature trajectory planning in an environment containing obstacles can be formulated by augmenting the robot's posture with a cost term, reflecting the cost of traveling along the current trajectory.

The path planning is mainly designed to generate efficient motion in the known map and works independently of directional localization of sounds. For sound localization, smooth motions help reducing erroneous directional localization and robust tracking.

C. Triangulation based Sound Position Estimation

A robot equipped with a microphone array localizes sound direction while moving and estimates sound positions by triangulation. Fig. 2 shows an example of sound localization with a moving microphone array. If the robot measures direction data at two different positions, it can estimate the sound position by calculating the point at the intersection of two vectors. If multiple sounds exist, there will be undesired cross points, like point P in Figure 2.

The problem is eliminated by categorizing direction data before triangulation. When the system calculates the sound position using the direction data derived from the same sound source, it can estimate the sound position without mismatching points.

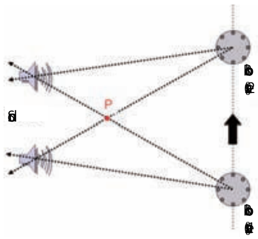


Fig. 2. Position estimation by triangulation from robot in motion

1) Tracking sound data: From the sound directional localization and identification system described in Section II, localized sound directions and their sound labels are obtained from the highest power intensity to the lowest at each time step. When the robot or the sound sources are assumed not to jump to another position, the data can be categorized using a continuously variable angle. The sound direction data is categorized using the displacement of sound-source directions and the recognition result, which is predictable from the robot trajectory.

$Obs(t)$ for a sound-source observation at time t is defined by equation (1)

$$\begin{aligned} Obs(t) &= \{\mathbf{s}_1(t), \mathbf{s}_2(t), \cdots, \mathbf{s}_M(t)\} \\ \mathbf{s}_m(t) &= [\theta_m, l_m]^T \quad (0 \leq m \leq M) \end{aligned} \tag{1}$$

where M is the number of localized sound sources at time t, and $\mathbf{s}$ denotes a sound observation vector including the localized direction θ and recognized sound label l.

From observed sound sources $Obs(t)$, each $\mathbf{s}_m(t)$ is connected to existing stream p or set as a new stream. For a connection judgement, the distance value J_n is defined by equation (2) and the stream number p for observation $\mathbf{s}_m(t)$ is determined by equation (3).

$$J_n = |\theta_m(t) - \hat{\theta}_n(t - \Delta t)|/\Delta t + \epsilon \times L(l_m, \hat{l}_n) \tag{2}$$

$$p = \underset{n}{\operatorname{argmin}}(J_n) \tag{3}$$

where ϵ is a weighting value of the recognition result and $L(a, b)$ is a model matching variable. $L = 1$ when the sound label of a and b is the same, otherwise $L = 0$. $\hat{l}_n$ is the sound label of stream n, which is obtained by the highest number of label including the stream.

When $\theta_m(t)$ satisfies $J_p < D_{thld}$, $\theta_m(t)$ is connected to $\hat{\theta}_p(t - \Delta t)$, otherwise, $\theta_m(t)$ is set as a new stream. The given parameter D_{thld} is the threshold of connection value. It is heuristically decided based on calcuration cycle, robot speed and sound localization accuracy.

For audio signals in a real environment, sound sources may appear or disappear at any moment. To correspond to that situation, we set a time limit on a tracked stream so that if the source Obs has not been observed for a certain amount of time, we consider that the sound no longer exists.

2) Motion Triangulation using RANSAC: From the observed sound directions on each tracked stream, the sound position is calculated using triangulation.

When observed direction has not changed enough, such as the robot does not move during the time that sound is generated or the robot's traveling direction is similar to the sound direction, the system cannot estimate the distance from the sound source by triangulation. On the other hand, the auditory information is helpful for a robot when directional localization and separation is achieved. For such situations, average direction in global coordinates and the voting result of sound labels from the tracked stream is used as detected sound information.

For triangulation, the past N direction data are used, so the number of data pairs used to measure cross points reaches ${}_NC_2 = N(N-1)/2$. Since the localization process sometimes fail to find the sound sources or some pairs of vectors have no cross points, the number of maximum cross points is less than ${}_NC_2$. From these cross points, the system estimates the sound source positions using the RANdom SAmple Consensus (RANSAC) [14].

The process is described as follows. First, i random cross points are selected and the root-mean-square error of i points is calculated. Next, the i points having the lowest mean squared error is searched to increase i to the maximum number of cross points. The estimation of the sound sources position is then achieved from the average of the i obtained data points.

This algorithm assumes that most cross points are distributed near the true position. Furthermore, in order to eliminate gross error, the algorithm eliminates cross points if the pair of vectors are nearly parallel or if the interval of the data pair is too close.

A summary of the sound position mapping method is presented in Table I. The method can be used for online data processing. The mapping system runs at every time step using detected sound streams from the last time limitted direction data. For each stream, if observed direction in global coordinates has not changed enough, it does not calcurate cross points and outputs average sound direction and label. Otherwise, it calculates cross points of directional localization vector and estimates sound position using RANSAC. For both case, the sound label is decided as the voting result in each stream.

V. Ssytem

A. Mobile Robot : Penguin II

Fig. 3 shows the mobile robot equipped with a SICK laser sensor and a 32 channel microphone array. There are two drive wheels attached in front of the body, and two passive caster wheels attached in the rear of the body. The computer has a 2GHz PentiumM CPU and 1GB of main memory. The wheel odometry and laser sensor data are used for robot position estimation.

Robot pose estimation is calculated on a client computer outside of the robot. It receives wheel odometry and laser data, and estimate the robot pose $(x, y, \phi)_R$ in a known

TABLE I

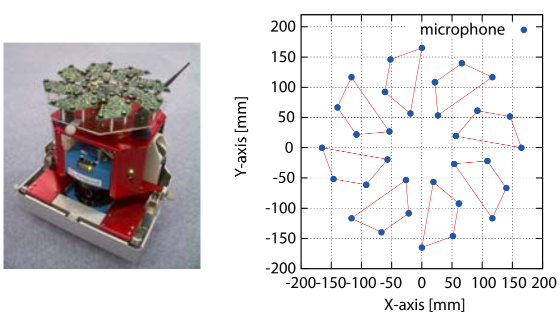
SOUND POSITION MAPPING ALGORITHM

Process to estimate the sound sources position
1. Add observed sound direction data to the tracked stream
for $s = 0$ to detected number of sound stream **do**
if directions in global coordinates are nearly parallel
2. Calculate average sound direction in global coordinates
3. Estimate the recognized sound label as a maximum voting label
else
4. Calculate the cross points by triangulation
(Using past local data, the number of points is up to ${}_NC_2$)
5. Pick up i random points
and calculate the root-mean-square error for each point
6. Search i, which has the least mean square error
7. Estimate the sound position as average of the i points in step 6.
endif
endfor

map. The client computer used for the experiment has 3GHz Pentium4 CPU and 1GB main memory. The calculation cycle of pose estimation is about 6 Hz for a 190×115 px size map (Fig. 5 in the next section).

B. The Robot Embedded Microphone Array

The proposed sound localization, separation and recognition is tested using a 32 channel microphone array attached on the mobile robot. The mobile robot embedded microphone array and its microphone arrangement are shown in Fig. 3. Through beam forming simulation, we decided the microphone arrangement to minimize side-lobes [15]. At each frequency from 700 to 3000 Hz, the focus direction gain compared to the highest side-lobe has 16 dB in average.



a) Mobile Robot b) Microphone arrangement

Fig. 3. 32 channel microphone array attached on a mobile robot: pen2r

The microphone array has 32 omni-directional electlet condenser microphones and can sample 32 data channels simultaneously. Sampling frequency is 16 kHz and resolution is 16 bit. For calculation in experiment section, data length is set to 1024 points and shift length is 512 points. Each 64 msec of sampled data is recognized every 32 msec.

VI. EXPERIMENT ON THE MOBILE ROBOT

A. Experimental condition

Fig. 4 shows a room used in this work. The room size is 10 × 18 m, and it has a reverberation time (RT_{20}) of 150 msec and background noise level of 35 dBA. 11 loudspeakers shown in Fig. 4 are used for sound sources.

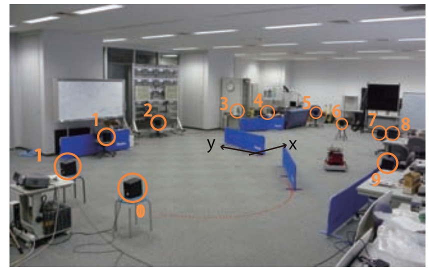


Fig. 4. The room and layout of speakers

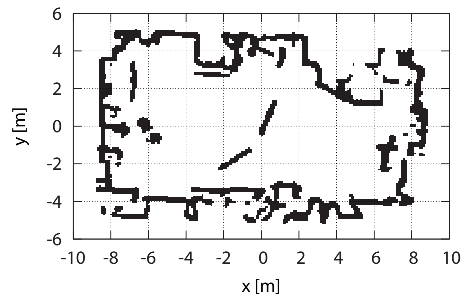


Fig. 5. Grid map used for robot position estimation

Fig. 5 shows the room map with the x-y coordinates used for robot pose estimation. The map is generated from laser scan data with wheel odometory by using ICP based scan matching [16]. The map is comprised of binary data (white and black), and 1 px indicates 0.1 m square area. The map size is 190×115 px.

For experiments in this section, the robot is controlled by the path planning result explained in Section IV. Goals of the path are manually selected. The average velocity is 0.4 m/s (1.1 m/s maximum) and maximum angular velocity is 1.3 rad/s.

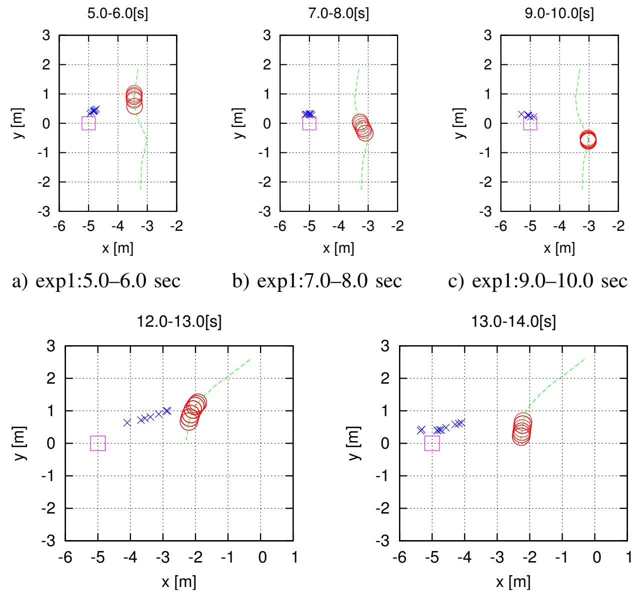
For tracking and estimating sound positions, the time limit is set based on the assumption of sound-source movement or to detect sounds going quiet, while a large samount of data provides high accuracy. In this paper, sound observations in the last 2 sec are used for mapping.

For PCMs recognition, the number of models is 16, including classical music, female/male voice, daily sounds and chirping sounds of living creatures. And cluster number $K = 15$ for each model.

B. Triangulation

Classical music from speaker11 (the speaker numbers are described in Fig. 4) is used as the sound source to evaluate position mapping for one sound source.

Fig. 6 shows the time progress of sound position estimation. The red circles with the green line indicate robot path, the magenta square is the speaker position and the blue points are estimated sound positions when the robot is at the red circles. Robot position interval of sound localization data pairs used in triangulation is 330 mm on average. When the robot is crossing in front of the sound (in (a) – (c)), the distance error is less than 400 mm. On the other hand, when sound direction is similar to the robot's traveling direction (in (d), (e)), the distance error becomes large.

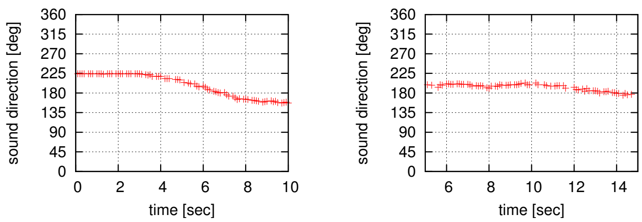


a) exp1:5.0–6.0 sec b) exp1:7.0–8.0 sec c) exp1:9.0–10.0 sec

d) exp3:12.0–13.0 sec e) exp3:13.0–14.0 sec

Fig. 6. Sound position estimation for 1 sound source

Fig. 7 shows the tracked sound direction data while the robot is in motion. The direction is in robot-centric coordinates, with 0 deg for the frontal direction and 90 deg for the left side.



a) exp1 b) exp3

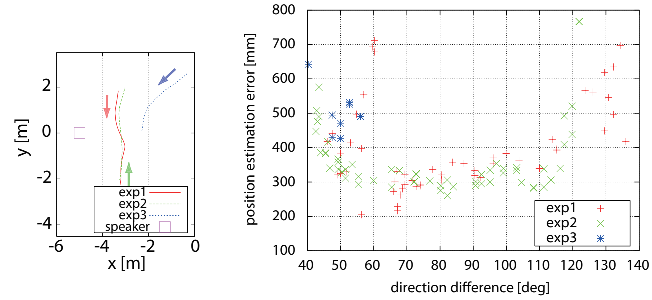
Fig. 7. Tracked sound direction in robot coordinates

Fig. 8 shows estimated sound position error in relation to the angle difference between the robot's traveling direction and the sound direction. The robot's path during the experiment is shown in Fig. 8 (a), and the estimated position error in relation to the angle difference is shown in (b). The performance depends on the distance from the sound sources and the robot position interval for triangulation, but in this condition, the errors are less than 400 mm for an angle difference of 60 to 120 deg (± 30 deg from viewed edge-on).

C. Tracking

To evaluate tracking two sound sources, we used two loudspeakers playing male and female speech in the following two conditions:

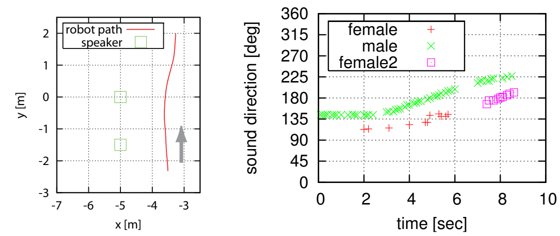
exp4: female at (-5.0, 0.0) m and male at (-5, -1.5) m
exp5: female at (-3.0, 0.0) m and male at (-1.5, -3.5) m



a) robot path b) distance error

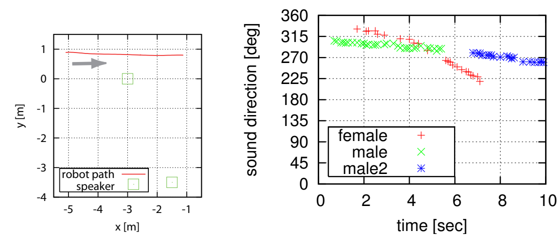
Fig. 8. Estimated position error related to robot traveling direction

The robot path and tracking results are shown in Fig. 9 and Fig. 10. In exp5, the two sound sources are crossing together at around 5 sec, but the tracking result (Fig. 10 (b)) shows both sound sources are tracked correctly.



a) robot path b) tracked sound direction

Fig. 9. Two sound sources tracking result : exp4



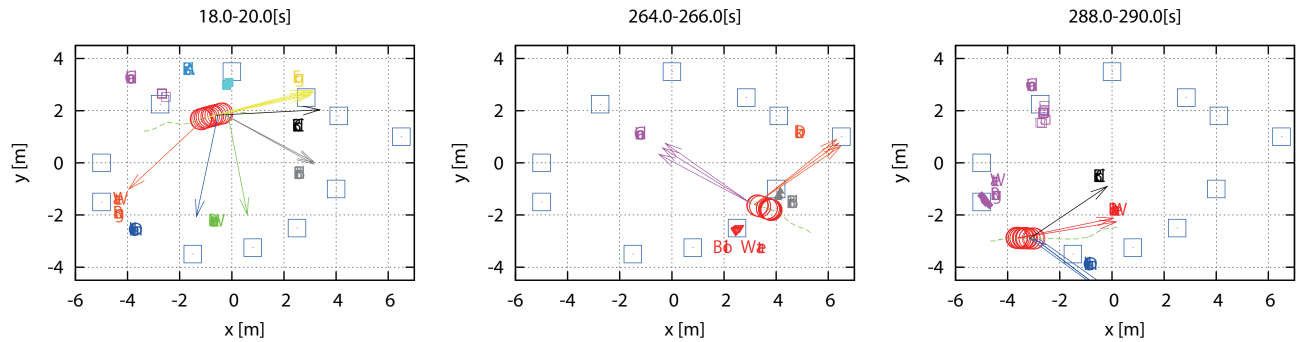
a) robot path b) tracked sound direction

Fig. 10. Two sound sources tracking result : exp5

D. Mapping

We tested sound source mapping for multiple sound sources in a large area using 11 loudspeakers playing a variety of sounds. The signal to noise ratio was 15 dBA from background noise level. The speaker arrangement and each sound source are shown in Fig. 11. All sound sources are included in PCMs codebook. From the original sound sources, half were used for PCMs codebook generation, and the remaining half were used as test sound sources.

Fig. 12 shows the time progress of the sound position estimation with recognized sound labels. In Fig. 12, vectors from the robot position shows detected sound direction when the robot fails to measure the position of a sound source because the localized angle did not change much while continuing to localize the direction. In Fig. 12 (a), the position of the nearby two sound sources was determined and



a) detect two positions and five directions b) detect two positions and two directions c) detect two positions and three directions

Fig. 12. Map-generation result at each time step

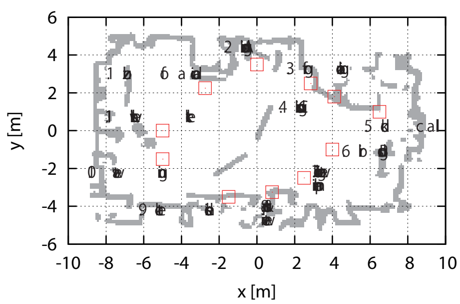


Fig. 11. Layout of the loudspeakers and sound source

the five other sound sources were detected with localization angles and recognized sound labels. In Fig. 12 (b), speaker6 (BirdsongB) and speaker7 (Boiling water) were detected with their position even when the sound direction from speaker 5 and 6 were crossing each other. And speaker 5 (Duck calling) behind the BirdsongB was correctly recognized with its direction. This result shows that observed sound tracking using sound recognition works correctly. In Fig. 12 (c), two sound sources were detected with 2D position and three other sound sources were recognized with their directions. The buzz of a cicada was detected at 5 m distant from the robot.

TABLE II
MAPPING RESULT: DISTANCE ERROR AND RECOGNITION PERFORMANCE

speaker	distance error mm	recognition performance % correct	wrong	unknown
1	235	83.0	0.9	16.1
2	285	76.5	0.0	23.5
3	348	74.6	11.9	13.6
4	326	84.8	0.0	15.2
5	465	79.5	0.0	20.5
6	245	68.4	15.8	15.8
7	241	59.5	23.8	16.7
8	221	73.3	0.0	26.7
9	161	66.7	0.0	33.3
10	274	31.7	45.0	23.3
11	302	80.0	0.0	20.0
average	282	70.7	8.8	20.4

The performance on each sound source is shown in Table II. The first column is the speaker number shown in Fig. 4. Fig. 13 summarizes the total estimated sound positions during the experiment (340 sec movement in total). The distance error of the estimated sound position was 282 mm on average, and 70.7 % had the correct sound label. For speaker7 (Boiling water), 23 % of the detected sources were recognized as "pour and stir water", and speaker10 (Running water) was mistaken for "pour and stir water" and "Boiling water".

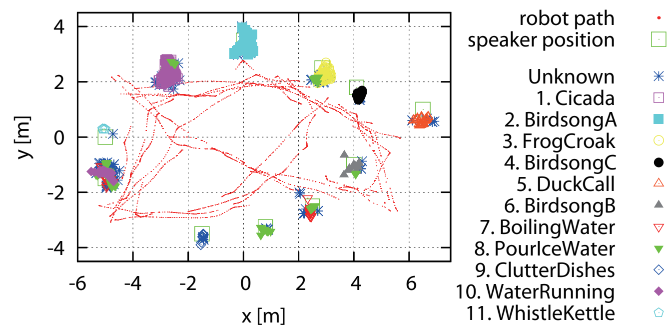


Fig. 13. All result of sound source mapping for 11 sound sources

VII. Conclusions and Future Works

The paper proposes multiple-sound-source mapping from a robot in motion. The proposed sound source mapping system is achieved by the following key components: 1) accurate directional localization and separation of multiple sound sources from a robot in motion 2) recognition of the separated sound sources from short term input 3) sound position estimation from the observed sound direction and the recognized sound label 4) robot position localization from a known map.

The sound directional localization using DSBF and MLF provides accurate directional localization for multiple different pressure sound sources on the 32 channel microphone array attached to a mobile robot. It works in sound crossing condition and covers larger area by avoiding erroneous localization of reflection or deformed peak when two sources are close together. PCMs-based sound identification works well for recognizing separated sound sources, and the sound label (recognition result) is useful for tracking.

The mapping result for 11 sound sources shows that the proposed system can calculate the 2D positions of multiple sound sources accurately, and recognize each sound source in a short period of time. The result show the possibility and limitation of surrounding sound event mapping by combining sound localization, separation, recognition and robot position localization from robot embedded sensors. Robust sound localization for a distant source or close two sounds are important to detect varied surrounding sound in global coordinates.

In this paper, the system measured the sound positions using triangulation and RANSAC estimation, and does not calculate the distance from the sound when the localized directions are nearly parallel. Reliability prediction for distance estimation, which depends on robot motion and sound position, is needed. Then selecting an interesting event and changing the robot's motion to investigate it are the next challenge to apply this auditory function.

REFERENCES

[1] Sebastian Thrun. Robotic mapping : A survey. Technical report, Technical Report CMU-CS-02-111, Carnegie Mellon University, 2002.

[2] Jean-Marc Valin, Francois Michaud, and Jean Rouat. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems Journal*, Vol. 55, No. 3, pp. 216–228, 2007.

[3] Jean-Marc Valin, Shunichi Yamamoto, Jean Rouat, Francois Michaud, Kazuhiro Nakadai, and Hirochi G Okuno. Robust recognition of simultaneous speech by a mobile robot. *IEEE Transactions on Robotics*, Vol. 23, No. 4, pp. 742–752, 2007.

[4] Eric Martinson and Alan Xchultz. Auditory evidence grids. In *Proceedings of 2006 IEEE/RSJ International Conference on Intelligent Robot and Systems(IROS2006)*, pp. 1140–1145, Beijing, China, Ovtober 2006.

[5] Kazuhiro Nakadai, Hirofumi Nakajima, Masamitsu Murase, Satoshi Kaijiri, Kentaro Yamada, Takahiro Nakamura, Yuji Hasegawa, Hiroshi G. Okuno, and Hiroshi 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*, pp. IV–929–932, Toulouse, France, May 2006.

[6] Yoko Sasaki, Satoshi Kagami, and Hiroshi Mizoguchi. Online short-term multiple sound source mapping for a mobile robot by robust motion triangulation. *Advanced Robotics*, Vol. 23, No. 1-2, pp. 145–164, Febrary 2009.

[7] Satoshi Kagami, Simon Thompson, Yoko Sasaki, Hiroshi Mizoguchi, and Takashi Enomoto. 2d sound source mapping from mobile robot using beamforming and particle filtering. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 3689–3692, Taipei, Taiwan, April 2009.

[8] Yoko Sasaki, Satoshi Kagami, and Hiroshi Mizoguchi. Simple sound source detection using main-lobe model of microphone array (in japanese). In *Proceedings of the 25th annual conference of the Robotics Society of Japan*, p. 1N13, Chiba, Japan, September 2007.

[9] Yuki Tamai, Yoko Sasaki, Satoshi Kagami, and Hiroshi Mizoguchi. Three ring microphone array for 3d sound localization and separation for mobile robot audition. In *Proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005)*, pp. 903–908, Edmonton, Canada, August 2005.

[10] Yoko Sasaki, Masahito Kaneyoshi, Satoshi Kagami, Hiroshi Mizoguchi, and Tadashi Enomoto. Daily sound recognition using pitch-cluster-maps for mobile robot audition. In *Proceedings of 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2009)*, pp. 2724–2729, St. Louis, Mi, USA, October 2009.

[11] Genshiro Kitagawa. Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, Vol. 5, No. 1, pp. 1–25, 3 1996.

[12] Simon Thompson and Satoshi Kagami. Continuous curvature trajectory generation with obstacle avoidance for car-like robots. In *Proceedings of International Conference on Computational Intelligence for Modeling Control and Automation(CIMCA2005)*, Vienna, 2005.

[13] J.J. Kuffner. Efficient optimal search of euclidean-cost grids and lattices. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1946–1951, Sendai, Japan, September 2004.

[14] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, Vol. 24, No. 6, pp. 381–395, 1981.

[15] Tomoaki Fujihara, Yoko Sasaki, Satoshi Kagami, and Hiroshi Mizoguchi. Arrangement optimization for narrow directivity and high s/n ratio beam forming microphone array. In *Proceedings of the 7th Annual IEEE Conference on SENSORS (IEEE SENSORS 2008)*, pp. 450–453, Lecce, Italy, October 2008.

[16] Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Transaction on Pattern Analysis and Machine Inteligence*, Vol. 14, No. 2, pp. 239–256, 2 1992.