2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Creation of radiated sound intensity maps using multi-modal measurements onboard an autonomous mobile platform*

Jani Even[1], Nagasrikanth Kallakuri[1], Yoichi Morales[1], Carlos Ishi[1] and Norihiro Hagita[1]

Abstract— **This paper presents a method for mapping the radiated sound intensity of an environment using an autonomous mobile platform. The sound intensities radiated by the objects are estimated by combining the sound intensity at the platform's position (estimated with a steered response power algorithm) and the distances to the objects (estimated using laser range finders). By combining the estimated sound intensity at the platform's position with the platform's pose obtained from a particle filter based localization algorithm, the sound intensity radiated from the objects is registered in the cells of a grid map covering the environment. This procedure creates a map of the radiated sound intensity that contains information about the sound directivity. To illustrate the effectiveness of the proposed method, a map of radiated sound intensity is created for a test environment. Then the position and the directivity of the sound sources in the test environment are estimated from this map.**

I. INTRODUCTION

The localization of sound sources is an important application of microphone arrays (see [1] and references herein). Generally, a sound source localization method estimates the bearing (azimuth and elevation) of a sound source in a referential frame attached to the microphone array (the estimation of the range is more inaccurate). Since the sound intensity radiated by a sound source is rapidly attenuated during propagation and drops below the background noise, the effective localization range is usually limited to a few meters.

To perform sound source localization in a large environment, nowadays it is possible to extend the effective range of localization by using a microphone array attached to a mobile platform and explore the environment. A natural framework for sound source localization from a mobile platform is to use a conventional sound localization algorithm at different locations and combine the results from all these different locations [2], [3], [4], [5], [6]. The precision of the sound source localization is greatly improved by combining observations of a same sound source from different locations (using triangulation [5] or accumulation on a grid map [3], [6]). However the mobile platform takes some time to move from one location to the next one. Consequently the sound source localization results at different locations are obtained for different times. Thus the combination of observations obtained from different locations is mainly beneficial for sound sources that have a fixed position and emit sound continuously. In the remainder of this paper, sound sources having such characteristics are referred to as *environmental sounds*. This term was chosen as air conditioning units, computers, air ducts and fans are good examples of sounds having these characteristics.

This paper presents a method for estimating the sound intensity radiated by the environmental noises and incorporate this information in a set of *audio maps*. These audio maps are a spatial representation of the information gathered about the sound scape. Previous works presenting audio maps built with a mobile platform [3], [7], [6] aimed at creating maps displaying the probability that a sound sources is present at a given location. The approach in this paper is fundamentally different, because the focus is on estimating a physical quantity, the radiated sound intensity. Another particularity of the proposed approach is that by exploiting the mobility of an autonomous mobile platform, it is possible to observe the environmental noises from different angles. Consequently, the estimated radiated sound intensity takes into account the directivity of the environmental noises.

The proposed method relies on a multi-modal approach to the problem: The data from the microphone array is combined with the data from the sensors of the mobile platform. The audio modality is used to estimate the received sound intensity as a function of the angle of arrival at the platform's position whereas the laser range finders (LRFs) onboard the mobile platform are used to estimate the distance to the geometric features of the environment (furniture, walls, objects, etc.). The use of LRFs is convenient as autonomous mobile platforms are commonly equipped with LRFs to localize themselves in the environment. The approach assumes that the objects emitting the sounds are detectable by the LRFs. This assumption is a bit restrictive when the LRFs scan a two dimensional horizontal plane, but an extension to a three dimensional scan is possible.

In practice, the audio map is a grid map covering the environment and the radiated sound intensity is estimated for each of the cells of this grid map while the platform is autonomously navigating through the environment. The two onboard LRFs (front and back) provide range scans and a steered response power (SRP) algorithm generates audio scans. The estimated received sound intensity (from the audio scans) and the range information (from range scans) are combined with the platform pose (position and orientation that are estimated with a particle filter) in order to update the estimate of the sound intensity radiated by the cells of the grid map. For each cell, the fusion of the information is performed by means of a set of Kalman filters. The result is a set of audio maps that represent the estimated radiated

*This research was funded by the Ministry of Internal Affairs and Communications of Japan under the Strategic Information and Communications R&D Promotion Programme (SCOPE).

[1]ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan even at atr.jp

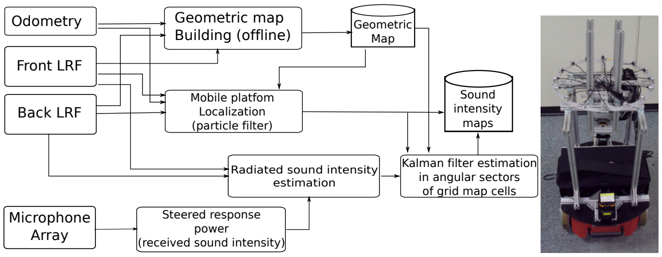


Fig. 1. Block diagram of the proposed method for creating map of radiated sound intensity and experimental platform with omni-directional microphone array and two laser range finders (front LRF clearly seen and back LRF partly occluded).

sound intensity and its directivity.

To illustrate the effectiveness of the proposed mapping technique, a map of radiated sound intensity was created, then sound source localization was performed by identifying the areas that radiate higher sound intensity and the directivity of these areas was estimated.

II. PRIOR WORK IN THE FIELD

Sound source localization using a mobile platform has been treated as a mapping problem in [3], [7], [6]. In these works, the audio map of the space explored by the mobile platform is based on a grid map. The grid map with cells of fixed size is overlaid on the environment and the probabilities of having a sound source in the cells are estimated during the exploration. In [6], audio ray casting is used and only the cells occupied by objects are considered whereas in [3], [7] all the cells are treated equally. These methods rely on a SRP with phase transform (SRP-PHAT [8]) to estimate the powers for a grid centered on the platform and these estimated powers are used to update the probabilities for the grid cells. In [7], after creating the audio map of probabilities, the platform is tele-operated to gather sound data in the vicinity of the sound sources at a fixed distance in order to estimate the directivity of these sources. The main difference between the proposed approach and [3], [7], [6] is that we estimate a radiated sound intensity at each cell and not a probability. Moreover, contrary to [7], in our approach the directivity is estimated at the same time as the localization is performed as the radiated sound intensity of each cell is estimated for different angular sectors and does not require the platform to be at fixed distance of the sound sources, see Sect.III-D.

The proposed approach is also different from works like [5] that proposes a triangulation based approach to sound source localization from a mobile platform or [9] that relies on a particle filtering rather than a direct triangulation. Both these works create maps of sound source locations that does not contain sound intensity or directivity information.

III. PROPOSED APPROACH

Figure 1 shows an overview of the proposed approach and the experimental platform. The different parts of the method are explained in the following sections.

Fig. 2. Geometric map of the environments (*corridor* on left and *room* on right)showing the position of the sound sources in red and there directivity in green. The white cells are not occupied cells (free space), the black cells are occupied cells (mainly walls) and the gray cells are unexplored cells. The cells are 5 cm by 5 cm.

A. Geometric map Building

The creation of the radiated sound intensity map requires the availability of a *geometric map* that describes the environment and enables the mobile platform to localize itself. Such a map represents the environment by an occupancy grid, namely a grid which cells are either empty (open space), occupied (walls and structures) or unexplored.

The geometric map is built in advance using the 3D Toolkit library framework [10], [11]. First the odometry and LRF data necessary to build the geometric map are gathered by driving the mobile platform through the environment with a remote controller. Then the LRF scans are aligned by correcting the trajectory of the platform using iterative closest point based simultaneous location and mapping (SLAM) [12] and the resulting aligned scans are combined to create the occupancy grid map [13], [14].

The geometric maps obtained for the test environments is shown in Fig.2. The free space (unoccupied cells) is in white, the walls and structures (occupied cells) are in black and the gray area shows the unknown space (cells that were not observed during the data gathering step).

B. Mobile platform Localization

The goal of the platform localization is to precisely estimate the pose (location and orientation) of the platform in the geometric map representing the environment. In this paper, this task is performed by a particle filter based localization algorithm that combines the information from the odometry (wheel encoders) and the LRFs (see [15] and references herein). A set of M particles approximates the probability density function of the platform's pose. The likelihood of the particles during the estimation step is based on the ray casting approach [15]. The platform pose $\{x(k), y(k), \theta(k)\}$ is given by the weighted average of the particles poses (using the most likely particle is another possibility).

C. Received directional sound intensity

The received directional sound intensity is estimated as a function of the angle of arrival by using a steered response power (SRP) algorithm (see [8] and references herein). In this paper, the SRP method is based on a delay and sum beamformer. The processing is done in the frequency domain after applying a short time Fourier transform (STFT) to the observed signals sampled at 48kHz (the analysis window is 25 ms long and the shift of the window is 10 ms).

After the STFT, at the sampled time t, the observed signals from the Q microphones at the frequency f are denoted by $U_1(f,t), \cdots, U_Q(f,t)$. Because the geometry of the microphone array is precisely known, it is possible to steer the array using spatial beam forming to estimate the sound from a spatial location (described by the spherical coordinates $\{\rho, \theta, \phi\}$ in the array's frame of reference). The beamforming output at the frequency f is denoted by

$$S(f,t,\{\rho,\theta,\phi\}) = \frac{1}{Q}\sum_{q=1}^{Q} e^{-2j\pi f \tau(\{\rho,\theta,\phi\},q)} U_q(f,t), \quad (1)$$

where $\tau(\{\rho,\theta,\phi\},q)$ is the delay at the qth microphone for the location $\{\rho,\theta,\phi\}$ (the microphone 1 is the reference $\tau(\rho,\theta,\phi,1) = 0$).

The SRP algorithm estimates the powers for a set of N locations. Contrary to [3], we assume the far field condition holds (ρ large compared to the array aperture). Moreover, the search space is limited to azimuth ($\phi = 0$). Thus a set of N locations is defined by the angles $\theta_n \in [0, 2\pi]$ with $n \in [1, N]$ (step $\Delta\theta$).

Then the power is computed in the frequency band $[1000, 5000]$ Hz corresponding to the discrete frequencies f_{1000} and f_{5000} by taking

$$P(t,\theta_n) = \frac{1}{\#F}\sum_{f=f_{1000}}^{f_{5000}} |S(f,t,\theta_n)|^2$$

where $\#F$ is the number of discrete frequencies in the band $[1000, 5000]$. Then time averaging is applied to combine $L = 10$ STFT frames

$$J_n(k) = \frac{1}{L}\sum_{t=k-L}^{k} P(t,\theta_n)$$

where k is the index corresponding to the blocks of L frames.

Namely the kth audio scan is a set of N angles θ_n with their associated power $J_n(k)$. The frequency of the audio scan is 10 Hz ($L = 10$ STFT window with a shift of 10 ms). A very important point is that contrary to [3], [7], [6], the phase transform Eq.(2) is not used. The phase transform is commonly applied to the frequency domain signals prior to processing the beamforming Eq.(1) by taking

$$U_{i,\mathrm{PHAT}}(f,t) = \frac{U_i(f,t)}{|U_i(f,t)|} \quad (2)$$

as it results in an SRP that is less sensitive to reverberation [8]. However, by discarding the amplitude of the signals of interest, the SRP with phase transform does not give any insight about the received sound intensity at the platform's position (it rather measures the coherence of the received signals). Consequently, not using the phase transform is a key point for estimating the radiated sound intensity as the SRP is then proportional to the received sound intensity.

D. Radiated sound intensity estimation

To create the audio map, it is necessary to estimate the radiated sound intensity corresponding to the received sound intensity and to position the sources of radiation in the map.

The audio scan $\{\theta_n, J_n(k)\}$ gives the angle of arrival and the received sound intensity in the platform coordinate frame. First, an estimated range $\rho_n(k)$ is associated to each of the directions θ_n by selecting the ray in the LRF scans that has the closest angle to θ_n. The LRF scans are denoted by $\{\theta_{i,F}, \rho_{i,F}(k)\}_{i\in[1,I]}$ for the front LRF and $\{\theta_{i,B}, \rho_{i,B}(k)\}_{i\in[1,I]}$ for the back LRF (There are $I = 1080$ rays per LRF scans). Namely the index of the range in the front or rear LRF scan is obtained by taking

$$i_0 = \mathrm{argmin}_i\ |\theta_{i,\bigcirc} - \theta_n| \quad , \quad \rho_n(k) = \rho_{i_0,\bigcirc}(k)$$

where $\bigcirc \in \{F, B\}$ denotes the front or back LRF. Then, by combining $\{\theta_n, \rho_n(k)\}$ with the platform pose $\{x(k), y(k), \theta(k)\}$ a position in the global referential is obtained. That position corresponds to a cell $\{i, j\}$ of the grid map covering the room. This transform is illustrated in Fig.3.

The angle $\beta_n(k)$ denotes the direction of radiation for the cell $\{i, j\}$ corresponding to the received sound intensity $\{\theta_n, J_n(k)\}$ at the mobile platform. Namely it is the angle (in the global referential) at which the mobile platform is seen from the cell $\{i, j\}$ when the kth measurement is performed.

Taking into account this angle makes it possible to estimate the directivity of the sound intensity radiated by the cells. This is an important point of the proposed approach that differentiate it from the approaches in [3], [6] that do not consider the directivity or from the approach in [7] that requires an exploration of the sound source at a fixed distance.

In practice, to account for the directivity of the cells, the

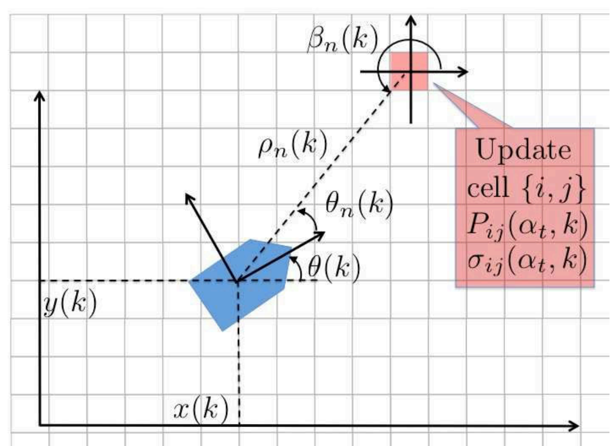


Fig. 3. Update of the cell $\{i, j\}$ by transforming the scans in platform referential coordinate frame to global coordinate frame using the platform pose $\{x(k), y(k), \theta(k)\}$, the scan angle θ_n and its associated range $\rho_n(k)$. The angle $\beta_n(k)$ (global coordinate frame) is exploited to estimate the cell directivity.

radiated sound intensity is estimated for T angular sectors

$$\alpha_t = \left[\frac{2\pi}{T}(t-1), \frac{2\pi}{T}t\right[_{t\in[1,T]},$$

where the notation $[a, b[$ denotes the interval $[a, b]$ excluding b. For the cell $\{i, j\}$, the sound intensity radiated at a reference distance ρ_0 in the sector α_t is denoted by $P_{ij}(\alpha_t, k)$. Then considering the propagation of the sound wave at a distance $\rho_{ij}(k)$ the sound intensity is

$$S_{ij}(\alpha_t, k) = \left(\frac{\rho_0}{\rho_{ij}(k)}\right)^2 P_{ij}(\alpha_t, k). \tag{3}$$

By taking the estimated received sound intensity $J_n(k)$ for $\beta_n(k) \in \alpha_t$ as an estimate of $S_{ij}(\alpha_t, k)$ made at the platform's position from the distance $\rho_n(k)$, Eq.(3) is approximated by

$$J_n(k) \approx \left(\frac{\rho_0}{\rho_n(k)}\right)^2 P_{ij}(\alpha_t, k). \tag{4}$$

The fusion of the observations for a given sector α_t of a cell $\{i, j\}$ is performed by a Kalman filter. Assuming that the radiated sound intensity is relatively stationary for the observation time, the process model is

$$P_{ij}(\alpha_t, k) = P_{ij}(\alpha_t, k-1) + w(k-1)$$

where w is the Gaussian noise $\mathcal{N}(0, q)$. The observation equation is

$$S_{ij}(\alpha_t, k) = \left(\frac{\rho_0}{\rho_{ij}(k)}\right)^2 P_{ij}(\alpha_t, k) + v(k)$$

where v is the Gaussian noise $\mathcal{N}(0, r)$. Using the approximation Eq.(4), it becomes

$$J_n(k) = \left(\frac{\rho_0}{\rho_n(k)}\right)^2 P_{ij}(\alpha_t, k) + v(k).$$

The Kalman filter provides an estimate $\widehat{P_{ij}(\alpha_t, k)}$ of $P_{ij}(\alpha_t, k)$ and its associated variance $\sigma_{ij}(\alpha_t, k)$. Each of the cells is assigned one Kalman filter for each to the T angular sectors. Each time the cell $\{i, j\}$ is *seen* from the mobile platform, i.e. when $\beta_n(k) \in \alpha_t$, the corresponding Kalman filter estimate is updated. The number of times for which the sector α_t of the cell $\{i, j\}$ is updated is denoted by $K_{ij}(\alpha_t, k)$.

By performing this estimation task while exploring the environment, the mobile platform essentially creates a grid map having cells that contain the information about the radiated sound intensity for T directions (the angular sectors). An important advantage of the proposed method is that the observation equation used in the Kalman filters takes into account the fact that the cells are often seen from different distances at same angle and the method does not require an explicit exploration of the sound sources at a fixed distance as in [7]. Note that this property comes from the fact that the phase transform Eq.(2) usually performed prior to SRP is not used.

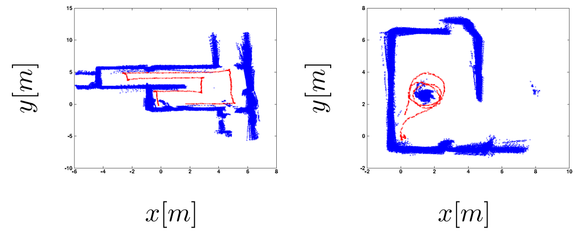


Fig. 4. Trajectory of the platform (red) and LRF scans (blue) for the corridor (left) and the room (right).

In practice, a small neighborhood of the cell $\{i, j\}$ is selected and updated. Namely the observation $J_n(k)$ for the cell $\{i, j\}$ is used to update the cells $\{i_*, j_*\}$ for which

$$\text{distance}\{\text{cell}_{ij}, \text{cell}_{i_* j_*}\} \leq \delta.$$

Moreover only the estimated ranges in $[\rho_{\text{min}}, \rho_{\text{max}}]$ are considered (i.e. too close and too far scans are discarded).

IV. EXPERIMENTS

For the experimental validation we used a pioneer platform equipped with two motor encoders and two laser range finders (UTM-30LX from Hokuyo, maximal range 30 m) and a microphone array composed of 16 Sony ECM-C10 microphones mounted on a circular frame (diameter 31 cm). The audio capture interface is a Tokyo Electron Device Limited TDBD16AD-USB that samples the signals at 48kHz. The experimental platform can be observed in Fig.1. The experimental evaluation of the approach was conducted in two environments: a corridor and a room. Figure 2 depicts the geometric maps of the environment (The dimension of the cells in this map is 5 cm x 5 cm). The platform navigates autonomously in the corridor using a set of way points that defined a loop covering all parts of the corridor (the velocity is 0.8 ms^{-1}). The red line in Fig.4 (left) shows the estimate of the trajectory given by the particle filter during one loop in the corridor. In the remainder, a *run* corresponds to the platform performing one loop in the corridor. Three such runs were performed. For the room, the platform performs several loops around the sound source, see the red line in Fig.4 (right).

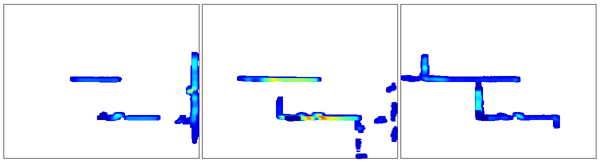
Several (up to three) sound sources were placed in the corridors (these locations are in the scan plane of the LRFs). The positions of the sound sources appear as $S1$, $S2$ and $S3$ in Fig.2 (left). These sound sources are loudspeakers playing recorded sounds: the sound of an air conditioning unit ($S1$ with a sound pressure of 78.5 dBA measured at 5 cm), the sound of a desktop computer fan ($S2$ at 77.5 dBA) and the sound of a server rack ($S3$ at 77 dBA). The sound pressure in the quiet corridor was around 42 dBA. The activation pattern of the sound sources for the runs can be observed in Table II and their positions are referenced in Fig.2(left). For the room, only the sound source $S1$ was place in the middle of the room (the server rack sound at 77 dBA). The sound pressure without sound source was 45 dBA.

The parameters are given in Table I.

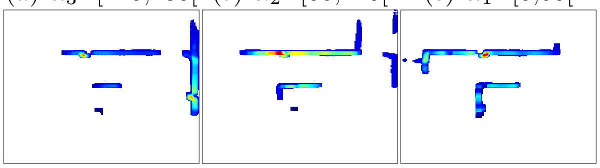
TABLE I

PARAMETERS USED DURING ALL RUNS.

$\Delta\theta$	T	$d_{\min}$	$d_{\max}$	δ	ρ_0	q	r	$\widehat{P_{ij}(\alpha_t, 0)}$	$\sigma_{ij}(\alpha_t, 0)$
1^o	6	0.5m	6m	0.05m	0.05m	10	0.05	20	10



(a) α_3=[120,180[(b) α_2=[60,120[(c) α_1=[0,60[



(d) α_4=[180,240[(e) α_5=[240,300[(f) α_6=[300,360[

Fig. 5. Directional radiated sound intensity maps $\widehat{P_{ij}(\alpha_t, k)}$ for the run 1 in the corridor. Each of the maps (a)-(f) corresponds to one of the angular sectors $\alpha_t = [60(t-1), 60t[$.

A. Results

The directional radiated sound intensity maps estimated during the run 1 are given in Fig.5(a)-(f). The maps for the different sectors are plotted counter clockwise from α_1 in Fig.5(c) to α_6 in Fig.5(f). Only the cells for which the number of visits $K_{ij}(\alpha_t, k)$ is greater than 100 are considered. In all figures, the color scale goes from blue for lower sound intensities to red for higher sound intensities (loudspeaker orientations are given in Fig.2).

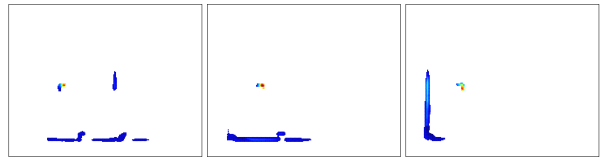
Figures 6 (a)-(f) show the directional radiated sound intensity maps estimated during the run in the room.

Figures 7 (left) and (right) show the maps obtained by taking $\max_t \widehat{P_{ij}(\alpha_t, k)}$ for the the run 1 and the run in the room. Namely, these maps shows the largest radiated intensity for each cell. A local maxima search was performed on the map Fig.7 (left) and the locations of the local maxima appear as black circles (the numbers are the ranks of the local maxima by decreasing intensity). The ground truth, i.e. the real positions of the sound sources are given as black crosses. For each of the sources the position errors are given in Table II. The results for the other runs are also in the table.

Figures 8 (a)-(f) and 9 (a)-(f) show the directivity maps in the neighborhood of $S3$ for the corridor and $S1$ for the room (the white arrows indicate the loudspeaker orientations).

B. Discussion

The sets of $T = 6$ grid maps in Fig.5(a)-(f) Fig.6(a)-(f) represent the sound intensity radiated by the environment. The cells have a radiation pattern that is function of the geometry of the environment and the path of the mobile platform. In particular, some of the angular sectors were never updated as these cells were not *seen* from the platform. However, the cells close to the sound sources and some other cells exhibit a clear directivity pattern that is function of the



(a) α_3=[120,180[(b) α_2=[60,120[(c) α_1=[0,60[

(d) α_4=[180,240[(e) α_5=[240,300[(f) α_6=[300,360[

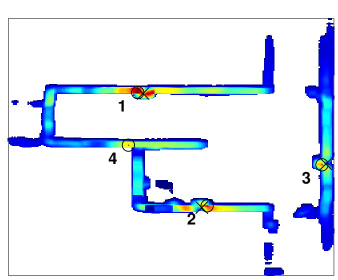
Fig. 6. Directional radiated sound intensity maps $\widehat{P_{ij}(\alpha_t, k)}$ for the run in the room. Each of the maps (a)-(f) corresponds to one of the angular sectors $\alpha_t = [60(t-1), 60t[$.

TABLE II

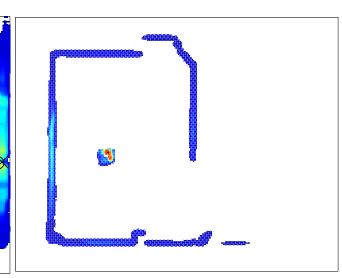
SOUND SOURCE LOCALIZATION RESULTS.

Run	Active Sources	Detected Sources	Error(m)
1	$S1$	$S1$	0.22
	$S2$	$S2$	0.16
	$S3$	$S3$	0.21
2	$S1$	$S1$	0.22
	$S2$	$S2$	0.15
	$S3$	$S3$	0.21
3	$S1$	$S1$	0.22
	$S3$	$S3$	0.3

sound they emit. Consequently the maps in Fig.7(left) and (right) that can be interpreted as a kind of superposition of the set of maps clearly shows the locations of the sound sources. The local maxima search in Fig.7(left) successfully estimated the positions of the sound sources (the three local maxima close to the true sources' location are the one with higher values). The average localization error was 0.21 m and the maximum error 0.3 m (See Table II). Considering that the loudspeakers are not point sources but may span several cells and that the localization was performed while moving, this localization error indicates a good performance. As comparison, the average localization error for the audio maps in [6] is 0.15 m and in [3] it is 0.24 m. But these

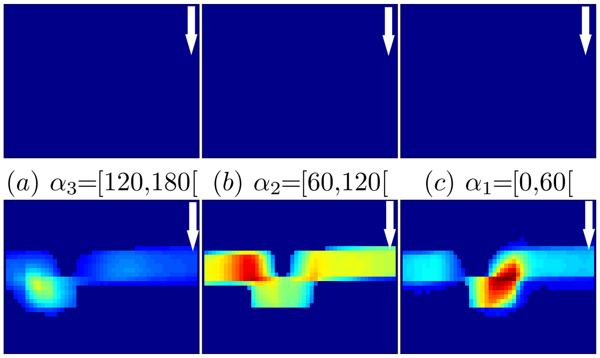


(a)



(b)

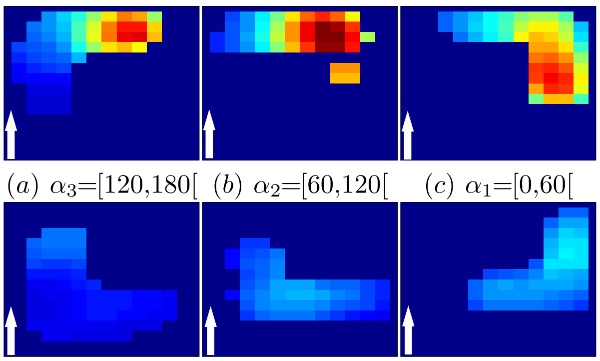
Fig. 7. (a) Map $\max_t \widehat{P_{ij}(\alpha_t, k)}$ for the corridor (run 1) and local maxima search results ; the crosses (x) are the ground truth and the circles (o) are the local maxima. (b) Map $\max_t \widehat{P_{ij}(\alpha_t, k)}$ for the room.



(a) α_3=[120,180[(b) α_2=[60,120[(c) α_1=[0,60[

(d) α_4=[180,240[(e) α_5=[240,300[(f) α_6=[300,360[

Fig. 8. Directivity of the sound source $S3$ in the corridor (run 1).



(a) α_3=[120,180[(b) α_2=[60,120[(c) α_1=[0,60[

(d) α_4=[180,240[(e) α_5=[240,300[(f) α_6=[300,360[

Fig. 9. Directivity of the sound source $S1$ in the room.

two approaches do not estimate the radiated power nor the directivity.

In Fig.7 (left), the fourth local maxima does not correspond to a sound source. It seem to be due to the superposition of the reflexion from the sound source $S3$ with the reflection of the mobile platform noise when the platform is moving close to the wall (the platform's noise reflects back to the onboard array when the platform is close to the wall).

The directivity patterns in 8 (a)-(f) shows the orientation of the sound source $S3$ (note that some of the angular sectors were not visited as they are on the wall side). For the sound source $S1$in the room that was observed from all direction during the run, all the cells have a value in 9 (a)-(f) and the source directivity is correctly estimated.

The results presented in this paper are obtained from a single run in the environment. But the location and directivity of the sound sources are already reasonably estimated. By using a trajectory that explores the environment more precisely and increasing the resolution of the directivity (using a larger T) it is possible to create more precise maps.

The scaling of the estimated sound intensity has not being fully investigated yet. But by exploiting Eq.(4), it may be possible to find a relation between the estimated sound intensity and the true value.

Note that the method proposed in this paper is creating the audio maps in real time while the mobile platform explores the environment.

V. CONCLUSIONS

This paper presented a framework for creating maps of radiated sound intensity with an autonomous mobile platform. The sound source localization results obtained in the experiments had an average distance error of 0.21 m using local maxima search, showing that the proposed framework is capable of localizing sound sources. Up to 3 sources within an 8m x 8m space were localized. These kind of mapping approach will aid the platform in attaining a better knowledge about environmental noise. In particular, knowing the directivity of the sound sources is an important information for a mobile platform. With the available framework, it is possible to extend the work to 3-Dimensional sound source localization that is more informative.

REFERENCES

[1] H. DiBiase, J. nad Silverman and M. Brandstein, *Microphone arrays : Signal Processing Techniques and Applications*, Springer-Verlag, 2007.

[2] J.-M. Valin, J. Rouat, and F. Michaud, "Enhanced robot audition based on microphone array source separation with post-filter," in *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, sept.-2 oct. 2004, vol. 3, pp. 2123 – 2128 vol.3.

[3] Eric Martinson and Alan C. Schultz, "Auditory evidence grids.," in *IROS*. 2006, pp. 1139–1144, IEEE.

[4] K. Nakadai, H.G. Okuno, H. Nakajima, Y. Hasegawa, and H. Tsujino, "An open source software system for robot audition hark and its evalation," in *IEEE-RAS International Conference on Humanoid Robots*, 2008, pp. 561–566.

[5] Y. Sasaki, S. Thompson, M. Kaneyoshi, and S. Kagami, "Map-generation and identification of multiple sound sources from robot in motion," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2010*, 2010, pp. 437–443.

[6] N. Kallakuri, J. Even, Y. Morales, C. Ishi, and N. Hagita, "Probabilistic approach for building auditory maps with a mobile microphone array," in *To appear in Proceedings of 2013 IEEE International Conference on Robotics and Automation, ICRA 2013*, 2013, pp. –.

[7] Eric Martinson and Alan C. Schultz, "Robotic discovery of the auditory scene," in *ICRA*, 2007, pp. 435–440.

[8] M.S. Brandstein and H.F. Silverman, "A robust method for speech signal time-delay estimation in reverberant rooms," in *IEEE Conference on Acoustics, Speech, and Signal Processing, ICASSP 1997*, 1997, pp. 375–378.

[9] S. Kagami, S. Thompson, Y. Sasaki, H. Mizoguchi, and T. Enomoto, "2d sound source mapping from mobile robot using beamforming and particle filtering," in *IEEE Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009*, 2009, pp. 3689–3692.

[10] D. Borrmann, J. Elseberg, K. Lingemann, Andreas Nüchter, and J. Hertzberg, "The Efficient Extension of Globally Consistent Scan Matching to 6 DoF," in *Proceedings of the 4th International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT '08)*, Atlanta, USA, June 2008, pp. 29–36.

[11] slam6d, "Slam6d - simultaneous localization and mapping with 6 dof," Retrieved December May, 20 2011 from `http://www.openslam.org/slam6d.html`, 2011.

[12] P.J. Besl and H.D. McKay, "A method for registration of 3-d shapes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[13] Hans Moravec and A. E. Elfes, "High resolution maps from wide angle sonar," in *Proceedings of the 1985 IEEE International Conference on Robotics and Automation*, March 1985, pp. 116–121.

[14] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, June 1989.

[15] Sebastian Thrun, Wolfram Burgard, and Dieter Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*, The MIT Press, 2005.