2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Mapping sound emitting structures in 3D

Jani Even, Yoichi Morales, Nagasrikanth Kallakuri[2], Jonas Furrer, Carlos Toshinori Ishi, Norihiro Hagita

Abstract— **This paper presents a framework for creating a 3D map of an environment that contains the probability of a geometric feature to emit a sound. The goal is to provide an automated tool for condition monitoring of plants. The map is created by a mobile platform equipped with a microphone array and laser range sensors. The microphone array is used to estimate the sound power received from different directions whereas the laser range sensors are used for estimating the platform pose in the environment. During navigation, a ray casting method projects the audio measurements made onboard the mobile platform to the map of the environment. Experimental results show that the created map is an efficient tool for sound source localization.**

I. INTRODUCTION

In acoustic signal processing, spatial filtering techniques using microphone arrays have been increasingly used [6]. To efficiently use spatial filtering methods it is necessary to know the precise locations from which the sounds to filter are emitted. Consequently, sound source localization has been extensively studied in the field of acoustic signal processing.

Microphone arrays and spatial filtering techniques have naturally made their way into robotics [16], [1], [13]. For some applications, like speech interfaces for example, the sound source localization algorithms can be used without robotic specific modifications [7]. This is especially true if the robot is not moving during speech acquisition.

But in the case of a mobile platform, it is possible to exploit the mobility of the platform for sound source localization. The natural approach is to perform sound source localization from different positions of the platform and combine the results. A straightforward approach is to use triangulation techniques to combine bearing estimates [14].

In [11], [12], this combination problem is treated in a probabilistic manner by estimating the probability of having a sound source for the cells of a grid covering the environment. This approach is very interesting as the results, the grid with assigned probability, can be seen as a probabilistic map of the sound source.

In [10], we proposed a method for building probabilistic maps of sound sources that reduces the complexity of the grid approach in [11] by considering only grid cells that are occupied by an object as a potential sound source candidate.

In this paper, we are interested in developing a method to automate the detection of sound sources in a known environment. An important goal of the proposed work is to establish the correspondence between the detected sound sources in the acoustic domain and geometric structures in the environment. The motivation for this aspiration is in condition monitoring of plants [9]. The purpose of condition monitoring is to assess the condition of the machinery components in order to schedule maintenance before a failure occurs. Thus, an important task is to detect the apparition of an unexpected sound and pinpoint the localization of the sound source on the machinery.

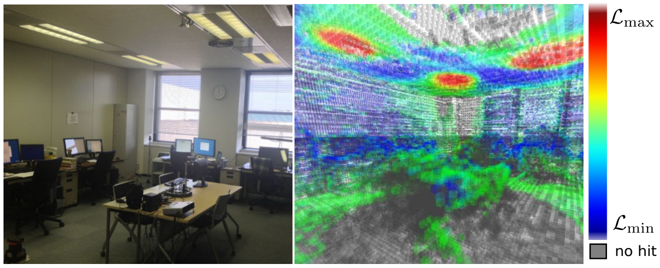

Fig. 1. View of the experimental room and the sound map. The scale $[\mathcal{L}_{\min}, \mathcal{L}_{\max}]$ denotes the probability of sound source presence. The active air conditioning grills in the ceiling appear in red.

The correspondence between a sound source and a structure in the environment is kept in a 3D map of the environment. In the remainder, this map is referred to as the *sound map* if it contains sound information and geometric information or *geometric map* if it contains only geometric information.

The sound map is build from the data collected by a mobile platform that navigates through the environment. The platform is equipped with wheel encoders and laser range finders (LRFs) to localize itself and a microphone array to perform sound source localization.

Sound source localization is performed continuously while navigating the environment and the sound source localization results obtained from the different positions are fused. Prior to the fusion process, the received sound power is transformed into a likelihood of sound source presence. Then this likelihood is accumulated to obtain the odds of having a sound source at a given position. Consequently, the output of the proposed approach is a 3D map of the environment that contains information on the presence of sound sources in a probabilistic form.

Fig.1 represents the sound map (right) created for a room (left). The colors in the 3D map shows the accumulated log-odds of having a sound source at that location (The gray color indicates that no information is available). Using this sound map, it is possible to detect that the air conditioning

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).

The authors are with ATR Intelligent Robotics and Communication Laboratories, Kyoto, Japan. even at atr.jp

[2] Now with the Electrical and Computer Engineering Dept, Carnegie Mellon University.

978-1-4799-3684-7/14/$31.00 ©2014 IEEE

grills in the ceiling are pulsing air at the time of mapping. As illustrated by this example, the created sound map can be a powerful condition monitoring tool for detecting structures emitting sound. Moreover, it is important to stress that for detection of sound sources in real structures, the proposed 3D sound map approach is an improvement over the maps obtained with the methods in [11], [10].

The presentation of the method is composed of five parts. First the creation of the geometric map is discussed. Then the localization of the platform in this geometric map is presented. The two next parts describes the onboard audio sensing and the creation of the likelihood from the audio power. Finally the last part explains how the likelihood is fused to create the sound map.

An experimental section illustrates the creation of the sound map and evaluates the influence of the different parameters.

II. GEOMETRIC MAP BUILDING

The creation of the 3D sound map requires the availability of a 3D geometric map that describes the environment.

The geometric map is built in advance using the 3D Toolkit library framework [4], [15]. In this framework, to build the geometric maps, odometry data and 3D scan data are necessary.

These data were acquired by driving a mobile platform through the environments to model. The mobile platform is equipped with wheel encoders, to provide odometry data, and a 3D Lidar, to provide scan data. Fig.2 shows one of the platforms we used (a Pioneer P3).

The scans are aligned by correcting the trajectory of the platform using iterative closest point based simultaneous location and mapping (SLAM) [3] and an octree representation of the environment is created [8].

In this paper, at the finest level of decomposition, the edge length of the voxels is 0.05 m and the voxels centered at the position $\{x, y, z\}$ is denoted by c_{xyz}. The geometric map refers to the voxels at the lowest level that are occupied.

In Fig.3, a view of an indoors environment and the corresponding view in the associated geometric map are juxtaposed. Note that the voxels that compose the octree are clearly visible.

III. PLATFORM LOCALIZATION

To create the sound map, it is necessary to precisely determine the position of the mobile platform in the geometric map that describes the environment.

In this paper, it is assumed that the ground is flat and that the platform's pitch and roll are negligible. Consequently, the pose of the platform is composed of its 2D location $\{x_r(t), y_r(t)\}$ and its orientation $\theta_r(t)$. The altitude is assumed constant $z_r(t) = z_0$ and the pitch and roll null $\{\phi_r(t) = 0, \gamma_r(t) = 0\}$.

Since the localization is reduced to a 2D problem, laser range finders (LRFs) scanning in the horizontal plane at a height h_{LRF} are used to localize the mobile platform in a 2D map (The platform in Fig.2 is equipped with two LRFs scanning the horizontal plane, one in front and one in the back). The 2D map is created by taking an horizontal slice of the geometric map at the height $\{h_{\text{LRF}} - \epsilon, h_{\text{LRF}} + \epsilon\}$ and flattening it. Then the referential in the 2D map coincide perfectly with the one in the geometric map. Fig.4 gives the naming conventions for the pose $\{x_r(t), y_r(t), \theta_r(t)\}$ in the referential of the 2D map. The green arrows shows the orientation of the mobile platform. This map correspond to the environment depicted in Fig.(3).

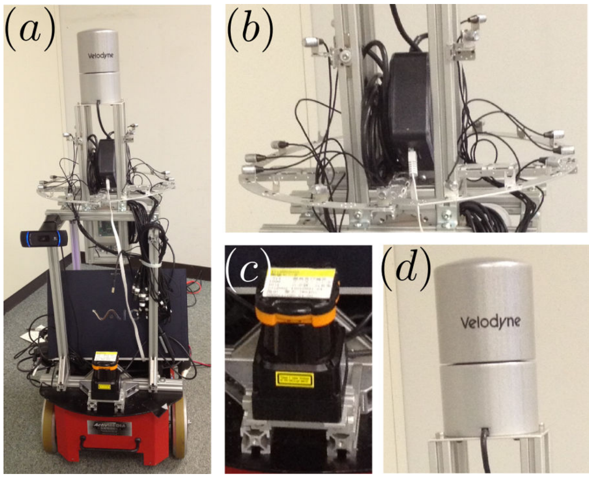

Fig. 2. Mobile platform example with sensors placement (a) and close views of the sensors: 16 Sony ECM-C10 microphone array (b), Hokuyo UTM-30LX LRF (c) and Velodyne HDL-32E 3D LIDAR (d).

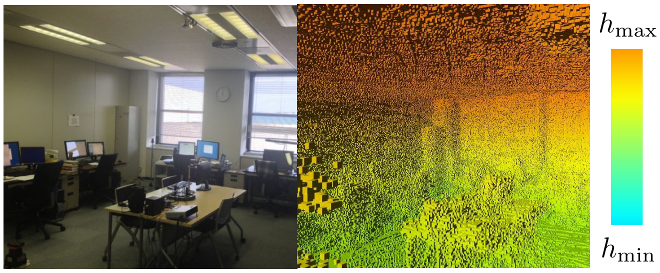

Fig. 3. Photo of the indoor environment and the corresponding view in the geometric map. The scale $[h_{\min} = -0.1, h_{\max} = 2.8]$ denotes the height in meter.

The localization algorithm is a particle filter (see [17] and references herein). In the prediction step, the particles are propagated accordingly to the odometry data. In the correction step, the likelihood of the particle is computed by using the ray casting approach to match the LRFs scan to the 2D map. Resampling is performed when the number of effective particles is too low.

The number of particles is 200 and correction is performed when the platform moved by 0.1 m or rotated by 5 degrees.

IV. STEERED RESPONSE POWER

In the field of sound source localization, steered response power (SRP) algorithms are regarded as efficient approaches (see [6] and references herein). In particular, the steered response power algorithms that make use of the phase transform (SRP-PHAT) [5] are considered a good match for

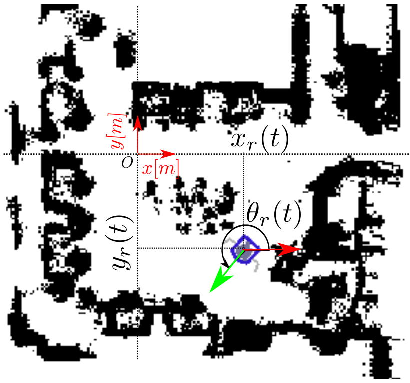

Fig. 4. Mobile platform localization in the $2D$ map created from the geometric map (the unit vectors $\{x[m], y[m]\}$ show the axes of the room coordinate frame and their units).

robotic applications [2]. Consequently, we use an SRP-PHAT algorithm to process the signals from the microphone array mounted on the mobile platform.

Let us first present the general idea of SRP algorithms. The *response* of the microphone array is the output signal obtained by applying a filter to combine the N signals received by the microphone into one. In SRP algorithm, this filter is a spatial filter than ideally lets only pass through the signals coming form a given direction $\{\theta_i, \phi_i\}$, see conventions in Fig.5. The *steered response power* is the power of this spatial filter's output signal. Namely, it is an estimate of the power coming from the direction $\{\theta_i, \phi_i\}$. Then sound source localization is performed by steering the array to a set of candidate directions, estimating the power for each of them, and finding candidate directions with higher power. The directions with higher power point to the location of the sound sources.

Now let us describe the SRP-PHAT algorithm used onboard the mobile platform in more details. The mobile platform is equipped with a microphone array depicted in Fig.2. The microphone array is composed of $N = 16$ omnidirectional microphones. The audio signals are synchronously sampled at 16 kHz by the capture interface. The audio processing is done in the frequency domain. The frequency domain signals are denoted by $X_n(f, t)$ where n is the microphone index, f the frequency bin index and t the frame index. They are obtained by applying a short time Fourier transform (STFT) to the audio signals. The analysis window is W points long and the shift of the window is $W/2$.

First the PHAT transform is applied to the frequency components

$$V_n(f, t) \quad = \quad \frac{X_n(f, t)}{|X_n(f, t)|}. \tag{1}$$

Then the power of the received sound is estimated for a set of candidate directions $\{\theta_i, \phi_i\}_{i \in [1,I]}$. The green dots in Fig.5 represent a set of candidate directions.

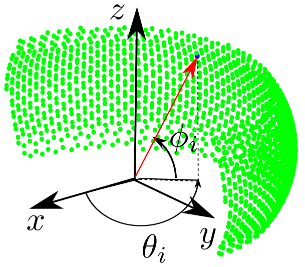

Fig. 5. Set of candidate directions (green dots) and conventions for the angles $\{\theta, \phi\}$ in the array referential.

For each of the candidate directions, the frequency domain processing is decomposed in 3 stages. First the response is steered in the candidate direction $\{\theta_i, \phi_i\}$ by applying a delay and sum spatial filter

$$Y(\theta_i, \phi_i, f, t) \quad = \quad H(\theta_i, \phi_i, f) \begin{bmatrix} V_1(f, t) \\ \vdots \\ V_N(f, t) \end{bmatrix}, \tag{2}$$

with

$$H(\theta_i, \phi_i, f) = \frac{1}{N} \left[e^{-j\tau_1(\theta_i, \phi_i, f)}, \cdots, e^{-j\tau_N(\theta_i, \phi_i, f)} \right], \tag{3}$$

where $\tau_n(\theta_i, \phi_i, f)$ is the phase delay at the microphone n in the frequency bin f for a signal coming from the direction $\{\theta_i, \phi_i\}$. Assuming that the sound sources are in the far field, the filter is entirely characterized by the angles $\{\theta_i, \phi_i\}$ and the microphone positions.

Then the power of the beamformer output is estimated by a K frame averaging

$$S(\theta_i, \phi_i, f, k) \quad = \quad \frac{1}{K} \sum_{t=0}^{K-1} |Y(\theta_i, \phi_i, f, k - t)|^2. \tag{4}$$

Note the introduction of the index k to show that the power has a different rate (the period is $KW/2$ samples).

Finally, the steered response power in the direction $\{\theta_i, \phi_i\}$ is obtained by selecting a limited band of frequencies

$$S(\theta_i, \phi_i, k) \quad = \quad \sum_{f=f_{\text{min}}}^{f_{\text{max}}} S(\theta_i, \phi_i, f, k). \tag{5}$$

In the remainder, the term *audio scan* refers to the set of candidate directions $\{\theta_i, \phi_i\}_{i \in [1,I]}$ (where I denotes the number of directions) and their associated power $S(\theta_i, \phi_i, k)$ computed at a given frame k. The kth audio scan is denoted by $\mathcal{S}(k) = \{S(\theta_1, \phi_1, k), \cdots, S(\theta_I, \phi_I, k)\}$.

An audio scan is represented as a colored portion of a sphere in Fig.6 (left). The color is function of the power for each of the candidate directions. This audio scan clearly exhibits an area of higher power on the top left side indicating the presence of a sound source.

V. AUDIO LIKELIHOOD

In order to fuse the sound source localization results of different audio scans together, the power $S(\theta_i, \phi_i, k)$ is first transformed in a likelihood $L(\theta_i, \phi_i, k)$. This likelihood $L(\theta_i, \phi_i, k)$ expresses the belief of having a sound source in the candidate direction $\{\theta_i, \phi_i\}$.

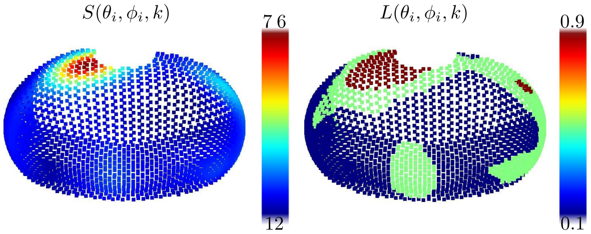

Fig. 6. Transformation by the thresholding function of the power (left) into a likelihood (right).

From the sound source localization literature and the idea behind the SRP approach, it is expected that a large power should correspond to a strong belief. For example in [12], [10] a scale version of the power was used as likelihood. In audio source tracking, creating a likelihood by scaling the power is also common [18].

In this paper, rather than using a scaled power, a nonlinear function is applied in order to create the likelihood. The selected nonlinear function is a double thresholding function

$$F(x) = \begin{cases} p_{\text{min}}, & \text{if } x < T_1 \\ p_{\text{max}}, & \text{if } x > T_2 \\ p_{\text{med}}, & \text{else} \end{cases} \tag{6}$$

Fig.6 shows the transformation of an audio scan with the nonlinear function (the parameters are set to $T_1 = 20$, $T_2 = 30$, $p_{\text{min}} = 0.1$, $p_{\text{med}} = 0.5$, $p_{\text{max}} = 0.9$).

Consequently while the mobile platform is navigating the environment for each audio scan $\mathcal{S}(k) = \{S(\theta_1, \phi_1, k), \cdots, S(\theta_I, \phi_I, k)\}$ a *likelihood scan* $L(k) = \{L(\theta_1, \phi_1, k), \cdots, L(\theta_I, \phi_I, k)\}$ is create by applying the nonlinear function. Each of these likelihood scans contains the likelihood of having a sound source for one of the candidate directions $\{\theta_i, \phi_i\}$.

VI. AUDIO MAP BUILDING

To understand the creation of the sound map, let us first discuss about the structure used to store the audio information.

The sound map is an octree having the same resolution as the geometric map. A voxel at the lowest level in the sound map is considered occupied if the corresponding voxel in the geometric map is occupied. The voxels of the sound map have some additional fields to store the audio information.

$\mathcal{L}(c_{xyz})$ denotes the log-odds of having a sound source within the voxel c_{xyz}. $\mathcal{M}(c_{xyz})$ counts the number of times the voxel c_{xyz} was updated during sound map creation. $\mathcal{U}(c_{xyz})$ contains the last time the voxel c_{xyz} was updated.

The candidate directions $\{\theta_i, \phi_i\}$ are defined in the referential centered at the microphone array depicted in Fig.5. This referential is rigidly attached to the mobile platform. The axis coincide with the platform's ones but the array origin is at a height $z_a = 1.6$ m. Thus to relate the likelihood $L(\theta_i, \phi_i, k)$ to a geometric structure in the environment, the candidate direction has to be combined with the estimated pose of the platform.

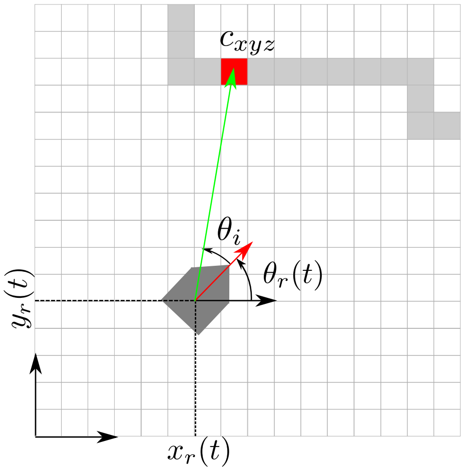

Fig. 7. Ray casting from the mobile platform pose $\{x_r(t), y_r(t), \theta_r(t)\}$ in the direction $\{\theta_i, \phi_i\}$ hitting the voxel c_{xyz}.

This combination is illustrated in Fig.7. For simplicity, a top view is presented and the elevation angle ϕ_i is omitted. The grayed squares represent the voxels of the geometric map.

For the likelihood scan $L(k)$, the pose $\{x_r(t), y_r(t), \theta_r(t)\}$ of the platform with t the closest to k is considered.

For each of the candidate direction, a ray is casted from the pose of the robot in the referential of the geometric map. Namely a ray is casted from the point $\{x_r(t), y_r(t), z_a\}$ in the direction $\{\theta_r(t) + \theta_i, \phi_i\}$. The rationale behind the use of ray casting is to trace back the sound until its sources as in [10]. However, by using a 3D representation of the environment, the sources are not restricted to be in a plane as in [10].

The ray casting is limited to a maximum range R_{max}. If the ray hit a voxel c_{xyz} of the geometric map then the likelihood $L(\theta_i, \phi_i, k)$ is used to update the log-odds of having a sound source in the corresponding voxel of the sound map.

The audio related fields of the voxel are updated as follows

$$\begin{aligned} \mathcal{L}(c_{xyz}) &= \mathcal{L}(c_{xyz}) + \log \frac{L(\theta_i, \phi_i, k)}{1 - L(\theta_i, \phi_i, k)} \\ \mathcal{M}(c_{xyz}) &= \mathcal{M}(c_{xyz}) + 1 \\ \mathcal{U}(c_{xyz}) &= t_k, \end{aligned}$$

where t_k is the time corresponding to the frame k. At initialization $\mathcal{L}(c_{xyz}) = 0$, $\mathcal{M}(c_{xyz}) = 0$ and $\mathcal{U}(c_{xyz})$ is undetermined. The choice $\mathcal{L}(c_{xyz}) = 0$ means that a voxel has equal chance to emit or not sound.

The log-odds $\mathcal{L}(c_{xyz})$ is no longer updated when it goes out of the interval $[\mathcal{L}_{\text{min}}, \mathcal{L}_{\text{max}}]$. Meaning that the odds of having a sound source at the voxel c_{xyz} is considered high or low enough to stop updating it.

VII. EXPERIMENTAL RESULTS

The experimental setting corresponds to the indoors environment depicted in Fig.3. At the time of the experiments, the sound sources in this environment are the grills of the

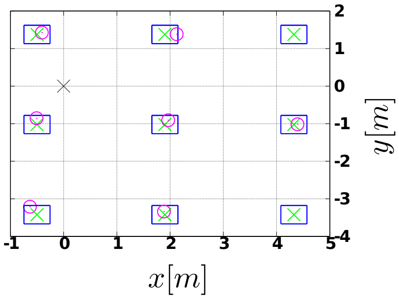

Fig. 8. Top view of the grills placement in blue, their center in green and the estimated positions in magenta.

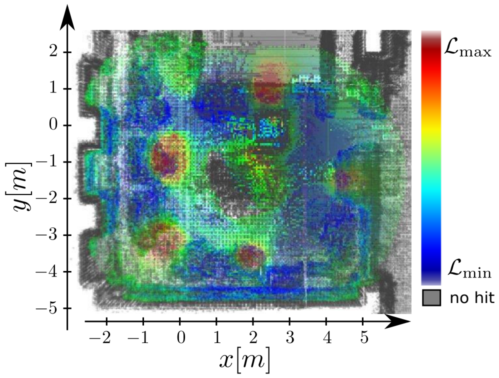

Fig. 9. Top view of the probabilistic 3D sound map, the color represents the probability of sound source presence.

air conditioning system. The grills are in the ceiling of the room and have a square shape (0.5 m edge). Fig.8 shows a top view of the room with the grills in blue.

To build the sound map, the mobile platform was driven three times around the table in the center of the room in a clockwise manner (see the 2D map in Fig.4). During the driving, the sensor data were all logged. Then in addition to the actual driving, the logged sensor data were played back in order to obtain results for other parameter settings. Moreover, to take into account the probabilistic nature of the particle filter based navigation, for each of the experimental conditions, 10 repetitions of the experiment (referred to as runs) are generated. It is important to stress that these runs are not simulated data but equivalent to the actual driving as all the sensor data necessary for the processing are logged.

Fig.9 shows the top view of one of the sound maps generated. Areas of high log-odds are visible around the location of the grills. The localization of the sound sources is estimated by clustering the voxels with positive log-odds (probability of having a sound source larger than 0.5). The clustering method is a $k$means method seeded with the positions of the local maxima of the log-odds. Each obtained cluster is assigned to the closest grill, then that grill is marked as detected and the distance to this grill is computed. In Figure 8, the magenta circles indicates the position of the detected sound sources (note that two sources are not detected).

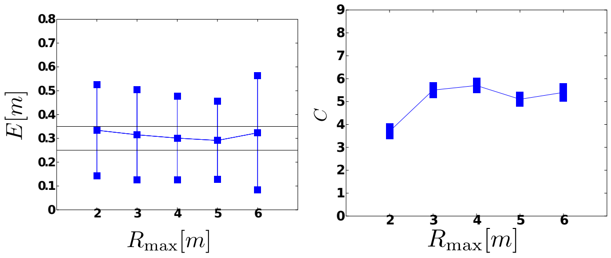

Fig. 10. Average error (left) and average number of detected sources (right) versus ray casting range limit and for the small aperture. The error bars show the standard deviations.

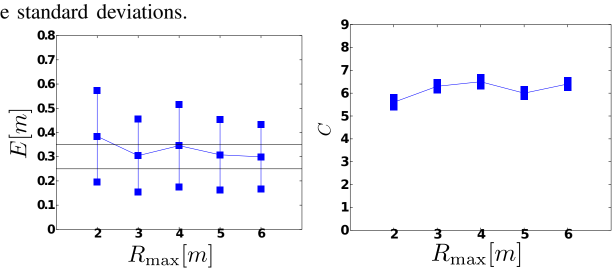

Fig. 11. Average error (left) and average number of detected sources (right) versus ray casting range limit and for the medium aperture. The error bars show the standard deviations.

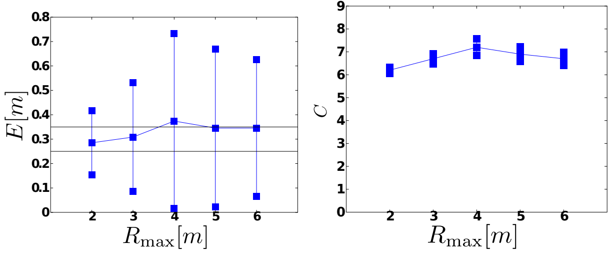

Fig. 12. Average error (left) and average number of detected sources (right) versus ray casting range limit and for the large aperture. The error bars show the standard deviations.

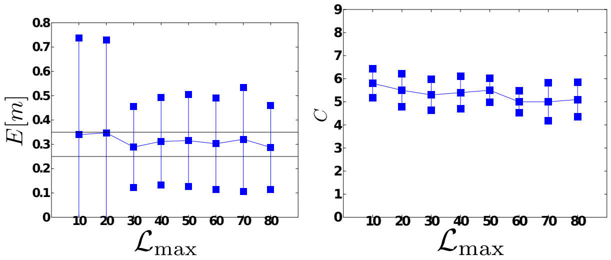

Fig. 13. Average error (left) and average number of detected sources (right) versus log-odds limit.

The sound source detection is evaluated in term of average number of detected sources C and average localization error E for these detected sources. For a given set of parameters, C is an average on the 10 generated runs. But E is an average on all the detected sources for the 10 runs.

The influence of the *aperture* of the scan grid was studied by using a small aperture $\theta_i \in [-15, 15]$, a middle aperture $\theta_i \in [-30, 30]$ and a large aperture $\theta_i \in [-60, 60]$. All with $\theta_i = 0$ indicating the front the platform and $\phi_i \in [-30, 75]$.

The influence of aperture was evaluated jointly with the ray casting range $R_{\max}$.

The other parameters of the methods are set to $W = 400$, $K = 5$, $f_{\min} = 1000$ Hz, $f_{\max} = 3000$ Hz, $\mathcal{L}_{\min} = -50$

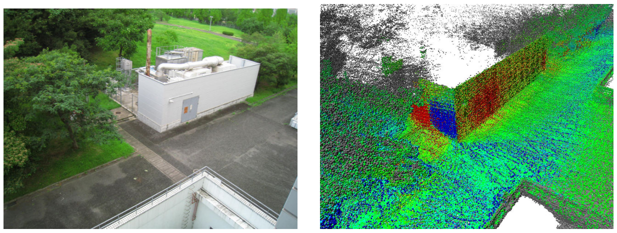

Fig. 14. Photo of an outdoor environment and the corresponding view in the sound map showing sound leakage from the door.

and $\mathcal{L}_{\max} = 50$. The parameters of the nonlinear function are the ones used in Fig.6.

Figures 10-12 illustrate the influence of two parameters: the ray casting range limit $R_{\max}$ and the the aperture of the scan grid (the range of θ_i). For each of the points the standard deviation is also indicated by error bars. For the plots of the average error, the horizontal black lines correspond to the radii of the circle circumscribed to the grill (≈ 0.35 m) and the circle inscribed in the grills (0.25 m).

As expected a larger aperture and a longer range increase the number of detected sound sources. However the accuracy of the detection is better with the small and medium apertures. A good trade off is the medium aperture with a range of 6 m as the error is among the small ones and the number of detected sources quite high. But the performance of the method is somehow robust to the choice of these parameters as the worst result for the large aperture with a range of 4 m is no catastrophic considering the size of the grills. As a comparison, the maps presented in [11], [10] also exhibit localization errors in the $0.2{\sim}0.3$ m range for the 2D case.

Fig.13 illustrates the effect of the log-odds limitation $\mathcal{L}_{\max}$ (the lower limit is set to $\mathcal{L}_{\min} = -\mathcal{L}_{\max}$) in the case of the small aperture with a range of $R_{\max} = 3$. The number of detected grills slightly decreases for larger values but the distance error is larger for smaller values. Thus taking a value in the range $30{\sim}70$ seem a good trade-off.

The undetected grills are the ones in the corners because the platform made three loops around the table in the center of the room. Moreover, the directivity of the grills may also make them easier to detect for larger values of ϕ_i. Driving the platform closer to the corners should reduce the variability of the count C.

VIII. Conclusion

This paper introduced a framework for creating a 3D description of the environment that contains the probability that a structure to emit sound. The aim of the method is to provide an automatic tool to sound source detection and associating them with geometric structures. Experimental results in an indoors environment showed that using the proposed approach it is possible to efficiently detect air conditioning grills and associate them with geometric feature in the environment. The method is not limited to indoors environment as illustrated by Fig.14, where the sound map shows the sound leakage from the door of a machinery building and from behind the building (higher intensities in red). The future work is to experiment in more diverse environments in order to determine the best parameter settings for different situations.

References

[1] S. Argentieri and P. Danès, "Broadband variations of the music high-resolution method for sound source localization in robotics," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS-2007, San Diego, USA*, pp. 2009–2014, 2007.

[2] A. Badali, J.-M. Valin, F. Michaud, and P. Aarabi, "Evaluating real-time audio localization algorithms for artificial audition on mobile robots," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009*, 2009, pp. 2033–2038.

[3] P. Besl and H. 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.

[4] D. Borrmann, J. Elseberg, K. Lingemann, A. 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.

[5] M. Brandstein and H. 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.

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

[7] J. Even, H. Saruwatari, and K. Shikano, "An improved permutation solver for blind signal separation based front-ends in robot audition," *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France*, pp. 2172–2177, 2008.

[8] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, 2013, software available at http://octomap.github.com. [Online]. Available: http://octomap.github.com

[9] A. Jar dine and L. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, p. 14831510, 2006.

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

[11] E. Martinson and A. C. Schultz, "Auditory evidence grids." in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2006*. IEEE, 2006, pp. 1139–1144.

[12] ——, "Robotic discovery of the auditory scene," in *Proceedings of 2013 IEEE International Conference on Robotics and Automation, ICRA 2007*, 2007, pp. 435–440.

[13] K. Nakadai, H. 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.

[14] 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.

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

[16] R. Takeda, K. Nakadai, K. Komatani, T. Ogata, and H. Okuno, "Exploiting known sound sources to improve ica-based robot audition in speech separation and recognition," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems IROS-2007*, pp. 1757–1762, 2007.

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

[18] D. B. Ward, E. A. Lehmann, and R. C. Williamsin, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Trans. Speech and Audio Processing*, vol. 11, pp. 826–836, Nov. 2003.