

# Incorporation of acoustic sensors in the regulation of a mobile robot

Christian A. Luna-Aguilar<sup>a</sup>, América Morales-Díaz<sup>a</sup>, Mario Castelán<sup>a\*</sup> and Climent Nadeu<sup>b</sup>

<sup>a</sup> Robotics and Advanced Manufacturing Group, Centro de Investigación y de Estudios Avanzados del IPN, Saltillo, México; <sup>b</sup> Center for Language and Speech Technologies and Applications, Technical University of Catalunya, Barcelona, Spain.

## ARTICLE HISTORY

Compiled October 24, 2018

## ABSTRACT

This article introduces the incorporation of acoustic sensors for the localization of a mobile robot. The robot is considered as a sound source and its position is located applying a Time Delay of Arrival (TDOA) method. Since the accuracy of this method varies with the microphone array, a navigation acoustic map that indicates the location errors is built. This map also provides the robot with navigation trajectories point-to-point and the control is capable to drive the robot through these trajectories to a desired configuration. The proposed localization method is thoroughly tested using both a 900 Hz square signal and the natural sound of the robot, which is driven near the desired point with an average error of 0.067 m.

## KEYWORDS

Acoustic localization; regulation control; navigation map.

## 1. Introduction

Nowadays, robotic systems are used to develop several tasks as mapping, localization, handling and transportation of pieces and object recognition, among others. To develop this tasks, robotic systems, and in particular mobile robots, need to be autonomous systems that can navigate in the desired environment.

Navigation requires to move the mobile robot along the desired path in a specific environment by means of a control strategy, and to robot localization is needed to this end. Therefore, robot localization and pose estimation in the desired environment is a problem that is commonly encountered in mobile robotics. Ideally, robot localization can be estimated by motor odometry [1,2]. However, this method is not accurate since errors accumulate due to wheel skid or even due to path irregularities. To this end, vision based localization of mobile robots is an area widely studied. For instance, an effective visual odometry method is proposed in [3,4]. A drawback of vision systems is the working area that is allowed by the sensor (camera) capabilities. One way to deal with this is to use a camera array that can be expensive and also spatially limited. Therefore, the development of sensor fusion strategies may enhance the robot localization area. Acoustic source localization is another active research field that is also capable of locating a robot within a working area, see for instance [5,6,10]. An

---

\*Corresponding author: email: mario.castelan@cinvestav.mx

advantage of acoustic sensors is their low cost and that they can be used also for other functionalities like speech [11] or speaker recognition [12]. Furthermore, they can be combined in an arrange to enhance localization, even without a rigid structure, with distributed microphones [13]. Therefore, one way to enlarge the localization area is to combine acoustic and visual localization. To this end, our paper defines a navigation map using acoustic localization, which allows to drive a mobile robot toward a safety zone where an external visual sensor can be used. This could be a potential application in intelligent warehouses, where the inclusion of sensors enriches the capabilities of the robots. In the case of microphones, for instance, there may be areas where cameras do not cover a required viewing area and specially designed sounds may help the robot re-localize if they are out of the sight of the cameras. Moreover, as reviewed in this paper, the natural noise of the robot may also be used for the purposes of localization.

### 1.1. *Contribution*

The contribution of our method relies in the combination of integrability constraints for generating an acoustic navigation map. These constraints are drawn from the field of surface analysis and gradient field integration, a sub-area of computer vision. The novelty here consists in integrating noisy acoustic gradient maps into smooth and consistent fields. This strategy is put into action together with a point to point regulation control based in the gradient of the localization map while the robot position is estimated a Time Delay of Arrival (TDOA) method with Phase Transform (PHAT)[14].

### 1.2. *State of the art*

Acoustic sensors represent an innovative technology for robot perception and localization. The human hearing system can be emulated by using an acoustic sensor array that receives sound signals to estimate the position of objects. The obtained sound signals by the acoustic sensors are manipulated by electronic devices in order to detect the sound source position and also to follow the position of the interest object.

Acoustic localization can be active, meaning that the sound creates an echo that will allow its further analysis to determine sound source position. It also can be passive, which involves sound or vibration detection of the source of interest.

There are a number of methods used for spatial sound localization [15], which can be coarsely divided into three main groups: (1) Array beamforming based methods, in particular the Steered Response Power (SRP) approach, which relies on the power of the signal received when the microphone array steers each target location [16]. (2) Methods based on high spectral resolution [17], which use spectral decomposition of the covariance matrices, like the Multiple Signal Classification (MUSIC) algorithm; they can work with several sound sources but are considered less robust against reverberation than beamforming based methods. And (3) Time Delay of Arrival (TDOA) methods, which work with the estimated time delays between signals captured by two spatially separated (a pair of) microphones; the time delays are combined with the coordinates of the microphones to estimate the source location. Time delays are often obtained from the Generalized Cross-Correlation (GCC) [18], employing the Phase Transform (PHAT) [14]. PHAT filtering is also used in combination with SRP (SRP-PHAT). The last technique is widely used because of its robustness [12], provided that the number of microphones is high enough. That is not the case in our application, so we will use the TDOA approach, which has the additional advantage

of requiring a low computational load.

Actually, the TDOA approach has already been used in robotics. One of the first works that uses the TDOA method was done by Hyuo Jo et al. [5]. The authors used a mobile robot with an incorporated sonar that is capable of locating acoustic reflectors. The navigation is based on these acoustic reflectors that are the reference points, which also determine the robot position. In the work of Valin et al. [6], the authors present a robust method for acoustic localization in a 3D space based on TDOA with the robot achieving a real time sound source localization. Also, in [7], a set of sound sources is used as landmarks and a microphone array is mounted over a wheeled robot for the purposes of simultaneous localization of sound sources and robot.

Over the years, acoustic localization techniques have been used to develop robotic cooperative tasks. For instance, in the work of Yuanqing Lie et al. [8], the relative pose of the robot group is estimated using acoustic sensors. To accomplish this aim, the authors applied the TDOA method to estimate the distance and the relative angle between the robots, this information is combined with odometry to update the state that describes the position and the angle of each robot. In the work of DongJu Kim and JongSuk Choi [9], the authors proposed a fusion between acoustic and visual localization to overcome the limitations of visual based methods. In their work, acoustic localization is based on TDOA with a four microphone array in a quadrilateral formation shape. Guo et al. [10] provided a relative localization method relaying on acoustic localization techniques for robots. They estimated the relative position of nearest robots using the reception specific sound that each one generated. However, they do not obtain an absolute localization of each member of the group in the working area.

Unlike our approach, a feature that the papers mentioned above have in common is that, in order to do sound source localization, they are based on additional information or in extra signal post-processing, such as the robot position measured with odometry, or the sound delay estimation obtained with electronics devices such as the Zig-Bee.

Other approaches for sound source localization have explored different possibilities. In [23] the authors proposed a method for motion planning for a mobile robot with a microphone to reach the better listening spot in the speech recognition technique; to maximize the accuracy of the target Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm was used; the effectiveness of this approach was verified through real experiments. A modeling approach capable of controlling three DOF of a robot in a plane using two microphones based on the measurement of the time difference arrival (TDOA) was provided in [24]. To demonstrate the performance of their method the authors presented several simulation results. A novel active audition method that moves multiple robots to optimal positions by simulating delay-and-sum beamforming from a possible layout under the condition that the positions of sound sources are already known, was proposed in [25]. Each robot had its own microphone array and the method effectiveness was proved by both simulations and in experiments using three sound sources and two robots. A statistical method that jointly estimates the time offset between each pair of microphone arrays, the positions of moving sound sources and those of stationary robots in an online manner was proposed in [26]. The authors provided experimental results using two robots and three sound sources in an anechoic chamber.

Unlike previous work, we use the TDOA GCC-PHAT method for acoustic localization and, from its results, we build a navigation map in order drive a mobile robot to a designated area by means of a regulation control. Also, for us it is important to explore the possibility of localizing the robot within a non-isolated room through its

natural sound rather than imposing external sounds through speakers.

The paper is organized as follows: in Section 3, the TDOA acoustic localization method and the navigation map generation are described; in Section 4 the point to point regulation control based on the localization map is presented; in Section 5 experimental results are outlined and finally, in Section 6 some concluding remarks are discussed.

## 2. Acoustic localization

TDOA based sound source localization implies mutual estimation of the delays generated by the wave sound trajectory between acoustic sources and sensors. Once these are estimated, the sound source positions can be calculated using geometry constraints.

Considering a lineal microphone array together with a sound source located at  $(x_s, y_s)$ , the signal  $s_k(t)$  that is acquired by microphone  $k$  can be expressed as

$$s_k(t) = \alpha_k r(t - t_k) + n_k(t), \quad (1)$$

where  $r(t)$  represents the waveform generated by the source,  $\alpha_k$  is an attenuation factor due to the propagation effects,  $t_k$  is the time propagation of the wave front, and  $n_k(t)$  includes the undesirable components due to the noise and the reverberation effect, which can be correlated with  $r(t)$ .

If a local planar propagation model is considered together with a temporal delay estimation technique applied to the pair of signals  $(s_{i_1}(t), s_{i_2}(t))$  in the microphones  $i_1$  and  $i_2$  gives an estimate of the mutual delay  $\tau_i = (\tau_{i_1} - \tau_{i_2})$ , named  $\hat{\tau}_i$ ; then, it is possible to estimate the angle  $\theta$  between the direction of the wave and the axis of the microphone array:

$$\hat{\theta} = \cos^{-1} \left( \frac{c\hat{\tau}_i}{d_i} \right), \quad (2)$$

where  $c$  is the sound velocity and  $d_i$  the distance between the two microphones.

We can use the position of the maximum of the GCC between the two signals  $s_{i_1}(t)$  and  $s_{i_2}(t)$ , as an estimator of the unknown delay  $\tau_i$  between them. To compute the GCC for each possible delay  $\tau$  the following expression is used:

$$\hat{R}_i(\tau) = \int_{-\infty}^{+\infty} \psi_g(f) G_i(f) e^{j2\pi f t \tau} df, \quad (3)$$

where  $G_i(f) = S_{i_1}(f) S_{i_2}^*(f)$  is the cross-spectrum, and  $\psi_g(f)$  is a frequency weighting function that in general depends on the noise, and that helps on to make the delay estimation process more robust.

Therefore, the TDOA estimate results from

$$\tau_i = \arg \max_{\tau} \hat{R}_i(\tau). \quad (4)$$

The widely-used PHAT weighting function is defined as

$$\psi_g(f) \equiv \frac{1}{G_i(f)} = \frac{1}{|S_{i_1}(f)||S_{i_2}^*(f)|}. \quad (5)$$

Hence, the generalized GCC is defined as follows:

$$\hat{R}_i(\tau) = \int_{-\infty}^{+\infty} \frac{G_i(f)}{|G_i^*(f)|} e^{j2\pi f t \tau} df. \quad (6)$$

### 2.1. Sound source localization

Once the TDOAs and the corresponding sound incidence angles ( $\hat{\theta}$ ) for each pair of microphones are calculated, a triangulation method can be applied by taking into consideration the knowledge of the microphone array geometry. The equation of the straight line that represents the acoustic path from the sound source to each pair of microphones is

$$y = m(x - x_p) + y_p, \quad (7)$$

where  $m$  is the line slope, which depends on  $\theta$ , and  $(x_p, y_p)$  is the mid-point  $p$  between the positions of the two microphones.

Equation (7) can be written in matrix form as:

$$\underbrace{\begin{bmatrix} -m & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\mathbf{x}} = \underbrace{-mx_p + y_p}_{b}. \quad (8)$$

There are as many equations as pairs of microphones in the array. In the case of two pairs, the estimated source position is the point where the two straight lines cross. If the number of pairs is higher than two, the line crossing points may not coincide, and a least-squares criterion is typically used. This way, the solution to the set of linear equations like the one in (8) is obtained from

$$\mathbf{x} = \mathbf{A}^+ \mathbf{b} \quad (9)$$

being  $\mathbf{A}^+$  the pseudo-inverse of matrix  $\mathbf{A}$ . In our experiments it is calculated by QR factorization with the Householder transformation.

### 2.2. Acoustic navigation map

In order to estimate an area where the acoustic source can be reliably localized using TDOA and to provide feasible trajectories to the mobile robot, an acoustic navigation map is built. To this end, a gradient map is generated through a matrix of acoustic localization errors. The gradient field of this matrix contains both discontinuities and

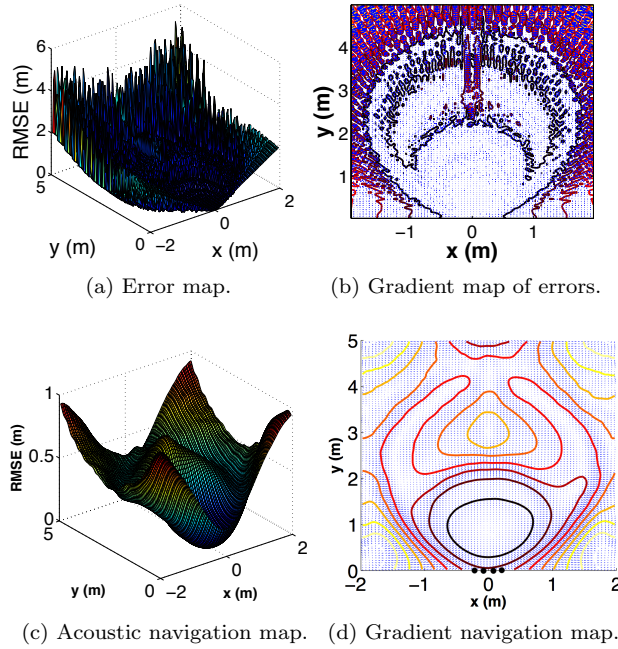


Figure 1.: **Acoustic navigation map.** A discontinuous localization error map (a) and its gradient field (b) are shown at the top row. Smoothed versions used as acoustic navigation maps and its gradient field are shown in (c) and (d), respectively, at the bottom row.

local minima. For this reason, we propose to smooth the gradient field of the error matrix so as to obtain an integrable surface. We call this surface the acoustic navigation map. An example of the error map and its gradient are shown in Figures 1a and 1b, respectively. The working area in the example is of  $5 \times 4$  meters.

The following procedure describes how the localization error map is generated:

- (1) Divide the working area in  $5 \text{ cm} \times 5 \text{ cm}$  cells to generate a grid. We considered this area as representative of a reasonable error for a stopping condition.
- (2) Randomly select 20 virtual positions from the sound source per each cell.
- (3) Emit a 900 Hz square signal in each selected sound source .
- (4) Estimate the sound source using TDOA.
- (5) Estimate the localization error using the root mean square error as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\|\vec{x}_i^{est}\| - \|\vec{x}_i^{ref}\|)^2}, \quad (10)$$

where  $\vec{x}_i^{est} = (x_i^{est}, y_i^{est})^T \in \mathbb{R}^2$  is the estimated sound source position vector obtained from TDOA,  $\vec{x}_i^{ref} = (x_i^{ref}, y_i^{ref})^T \in \mathbb{R}^2$  is the reference sound source position vector,  $n$  is the total number of estimations during the trial and  $\|\cdot\|$  denotes the  $L_2$  norm.

A lineal four microphone array is used to acquire the position of the sound source. The origin of the world is located at the physical center of this array. The separation between microphones is 6.2 cm, which is the distance between the microphones of the

PS3<sup>®</sup> Eye sensor, also used in the real experiments. Five pairs of microphones are used in our experiments: all combinations except the one corresponding to the two end microphones.

From the generated sound source localization error map and its gradient map, we would like to have a minimum error zone where the mobile robot can be located. Since the gradient map possess several minima and is also discontinuous, it is desirable to have smoother and continuous version of it. To this end, we applied a gradient field integration technique based on solutions for the *Shape-from-shading* [20] problem, which is a problem of computer vision.

It is possible to integrate a surface using the vector field obtained from their partial derivatives. The anti-derivatives group from a vectorial field defines a definite integral, which corresponds to the characteristic surface from this field. Let  $Z(x, y)$  be a scalar function defined in  $\mathfrak{R}^3$ , moreover,  $Z$  is a surface function if there exist their partial derivatives  $\left(\frac{\partial Z(x, y)}{\partial x}, \frac{\partial Z(x, y)}{\partial y}\right)$  and the integrability condition is fulfilled as follows:

$$\frac{\partial^2 Z(x, y)}{\partial x \partial y} = \frac{\partial^2 Z(x, y)}{\partial y \partial x}, \quad (11)$$

then its gradient field is,

$$\nabla Z(x, y) = \left[ \frac{\partial Z(x, y)}{\partial x}, \frac{\partial Z(x, y)}{\partial y} \right]. \quad (12)$$

Moreover, the gradient function points in the function maximum growth direction too.

Basically, there are two main methods known to approximate the gradient field, these are the local integration along a given trajectory and the global ones. Through basis functions, global methods can obtain the best approximation from non integrable gradient fields as error is not accumulated. Frankot and Chellappa in [20] assumed that if  $Z$  is unknown and satisfies the integrability condition (11), it is possible to represent it using Fourier coefficients as

$$Z(x, y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Z^{(F)}(u, v) e^{-j(ux+vy)} du dv, \quad (13)$$

where

$$Z^{(F)}(u, v) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Z(x, y) e^{j(ux+vy)} dx dy. \quad (14)$$

In (14),  $Z^{(F)}$  represents the Fourier coefficients of  $Z$ . Frankot and Chellappa determined that it is possible to reconstruct the  $Z$  function in the Fourier space by orthogonally projecting the non-integrable gradient field onto its nearest integrable version. Once this gradient field is projected, the inverse Fourier transform is used to reconstruct the newly integrated surface. This global minimization framework may include curvature and area constraints [21]. Also, it is required to first filter out high gradient values in order to keep the stability of the reconstruction free from energy bias.

The concept of acoustic navigation map is inspired by the acoustic map obtained through the SRP-PHAT technique. We have explained how the Frankot and Chellappa

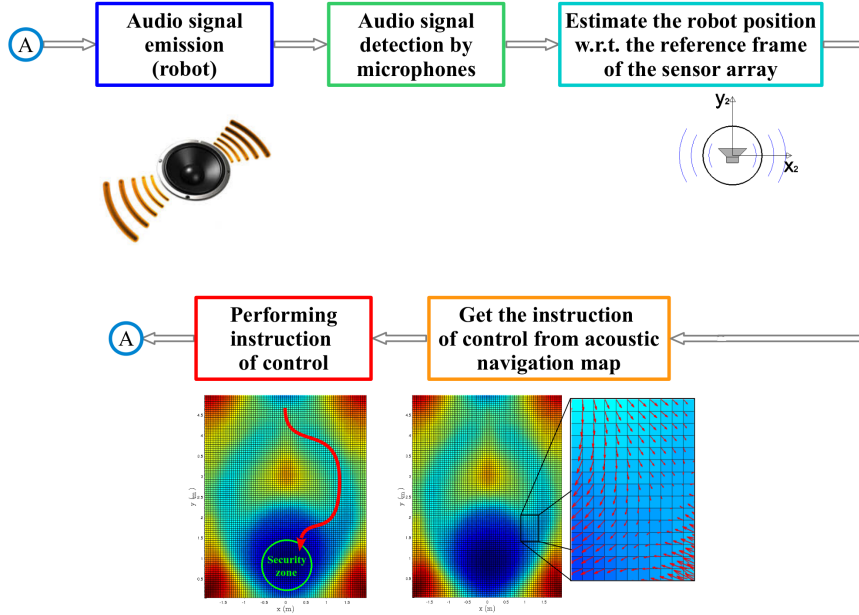


Figure 2.: **Methodology for our point regulation control.** We use an acoustic navigation map whose gradient allows the robot reach a final position where there is a set of microphones.

integrability scheme can be applied onto highly discontinuous acoustic localization error maps. An example of imposing integrability constraints for this task is presented in Figure 1c and 1d, where it is clear how a global integration method benefits the continuity of the produced map and also leads to the existence of a global minimum. This is a necessary condition for the point to point regulation control described in the next section.

### 3. Point regulation control

In this section we explain a method that combines acoustic data with an acoustic localization map to regulate the robot position to drive it to a safe localization area, i.e., where another sensor such as a camera can be used to locate the robot. [The general idea is depicted in Figure 2.](#)

We use a differential type unicycle robot. Its Cartesian configuration is described by the  $(x, y, \theta)$  coordinates, where  $(x, y)$  are the center of mass planar position of the robot and  $\theta$  is its orientation. A caster wheel only provides stabilization to the robot. The unicycle robot possesses a non-holonomic constraint that avoids lateral motion described by  $\dot{x} \sin(\theta) - \dot{y} \cos(\theta) = 0$  and its kinematic model is  $\dot{x} = v \cos(\theta)$ ,  $\dot{y} = v \sin(\theta)$  and  $\dot{\theta} = \omega$ , where  $v$  and  $\omega$  are the linear and angular velocities, respectively, which are the control inputs. Moreover, they are bounded as  $0 < |v| \leq v_{\max}$  and  $0 \leq |\omega| \leq \omega_{\max}$ .

The robot is equipped with a speaker to emit signals. A four microphone array with 192  $KHz$  sampling frequency  $F_s$  and a separation of 6.2 cm is fixed in a position within the working area. The audio signals emitted by the robot are detected by each microphone and processed using the TDOA method presented in Section 2. This allows to estimate the robot position, i.e., the Cartesian  $(x, y)$  coordinates while its orientation is calculated using the kinematic model of the robot.



In order to drive the robot toward the safety area while considering the navigation map, the following regulation control is defined for the linear and angular velocities:

$$\begin{aligned} v &= \gamma \|\nabla F(x, y)\|, \\ \omega &= -\delta \theta_e \left| \frac{\sin(\theta_e)}{\theta_e} \right|, \end{aligned} \quad (15)$$

where  $\gamma, \delta > 0$  are the control gains,  $F(x, y)$  is the function that describes the navigation map behavior,  $\nabla F(x, y) = \left[ \frac{\partial F(x, y)}{\partial x}, \frac{\partial F(x, y)}{\partial y} \right]$  is the gradient of  $F$  evaluated in coordinates  $(x, y)$ , having its minimum value within the security zone, and since  $\frac{\sin(\theta_e)}{\theta_e} = \int_0^1 \cos(s\theta_e) ds$  is a smooth function in  $\theta_e$ , and given that  $\lim_{\theta_e \rightarrow 0} \frac{\sin(\theta_e)}{\theta_e} = 1$ , this term is added to smooth the unicycle robot angular velocity; the symbols  $|\cdot|, \|\cdot\|$  denote the absolute value and norm  $L_2$ , respectively; finally,  $\theta_e$  is the orientation error defined as follows:

$$\theta_e = \theta - \theta_r = \theta - \arctan \left( \frac{\frac{\partial F(x, y)}{\partial y}}{\frac{\partial F(x, y)}{\partial x}} \right), \quad (16)$$

here, the unicycle kinematic model non-holonomic constraint  $\frac{\dot{x}_r}{\cos(\theta_r)} = \frac{\dot{y}_r}{\sin(\theta_r)}$  provides  $\theta_r = \arctan(\frac{\dot{y}_r}{\dot{x}_r})$ , and from the navigation map  $\dot{y}_r \approx \frac{\partial F(x, y)}{\partial y}$  and  $\dot{x}_r \approx \frac{\partial F(x, y)}{\partial x}$ . The angle  $\theta = \int_0^t \omega_p dt$  is also calculated from the robot kinematic model where  $\omega_p$  keeps the previously calculated angular velocity and it is compared against the gradient direction that is measured with respect to the positive axis  $x$ . The values of  $\nabla F(x, y)$  are obtained from a dataset created during the acoustic map generation stage and at the beginning  $\theta$  is considered known.

#### 4. Experimental results

To demonstrate the performance of the point regulation control using a navigation map shown in Section 3, a series of experiments are carried out in which a mobile robot is equipped with a speaker that emits acoustic signals. In addition, there is a fixed array of four microphones fixed at the origin of the working area. The position of the robot is estimated by processing the audio signals with the TDOA acoustic localization method, since the sound source is located in the center of mass of the robot. Two alternatives are presented in the acoustic signal used to estimate the position of the robot, the first consists of a square signal of 900 Hz emitted by the speaker mounted on the robot and as a second option we used the natural sound emitted by the motors and the tires of the robot. All results were obtained in a non-isolated laboratory room of size 7m  $\times$  5m whose T60 reverberation characteristic is about 0.7 seconds. The acoustic environment was rather quiet during the experiments, and the mean signal-to-noise ratio of the acquired signals was relatively high. In this section we present experimental results of the point regulation control based on the acoustic navigation map, which (unlike that shown in Section 3) has an integral action. However, the control methodology continues as presented above.

The experimental platform developed to test the point regulation control with a mobile differential robot is made up of a control system in charge of determining the control speeds of the robot and a vision system through which the trajectory performed

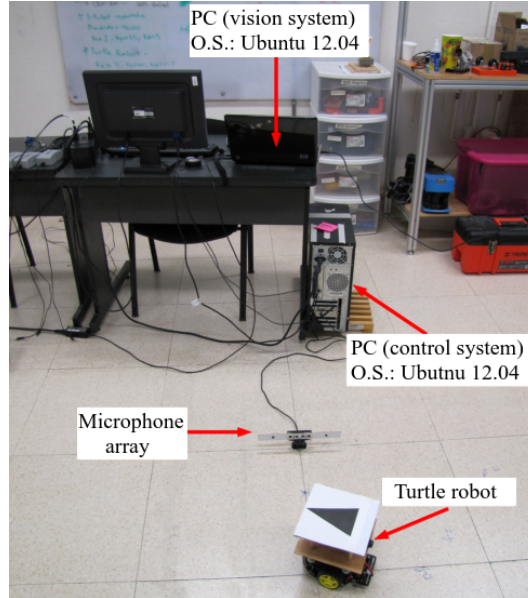


Figure 3.: **Experimental scenario.** The different elements of our experimental setting are shown. The turtle robot is tracked with a vision system and tries to reach a safety zone near the microphone array.

by the robot is calculated. The elements conforming the control system are: a PC with Ubuntu 12.04 operating system, a linear array of four microphones ( $F_s = 16$  KHz), a turtle robot and a speaker, see Figure 3. On the other hand, the vision system is made up of a PC with Ubuntu 12.04 operating system, an uEye camera model UI-1220SE-M-GL and a Navitar NMV-5M23 lens. The camera is placed in the upper part of the laboratory. The operating frequency of the control system is approximately 10 Hz. To carry out the control and to determine the trajectory of the robot we used the programming language C++. For obtaining the audio signal, we used the open source platform PortAudio on the PC of the control system.

The Turtle robot is equipped with a wooden base on which the speaker is placed. In addition, an identification mark used by the vision system to determine the position of the robot is placed on the top. The communication between this type of robot and the PC is carried out through Bluetooth. The information sent by the PC is interpreted as wheel speeds by means of an Arduino UNO processor. The friction between the wheels of the robot and the floor is not considered in the control system.

In order to obtain the position and orientation of the robot using vision, a triangular black mark placed on the top of the robot is used. In the image obtained by the camera, the edges of the robot mark are detected, thus allowing the determination of the centroid of the robot and consequently the position of the robot. The orientation of the robot is obtained as a consequence of determining the angle of inclination of the vertex farthest from the centroid of the mark. This procedure, used in [22], is carried out in an iterative way in order to count the position of the robot at each instant of time.

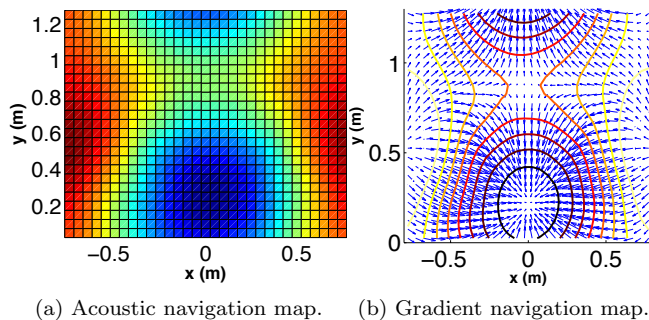


Figure 4.: **Navigation map and its gradient.** This example corresponds to the working area of our experiments.

#### 4.1. Adjusting control parameters

In order to improve the performance of the point regulation control using the navigation map shown in Section 3, an integral action is added to the control law given by Equation 15 so as to avoid wheel speeds below 50 mm/s. This is because the Turtle robot does not break the frictional force between the wheels and the floor when its speed is in that operating range. This behavior does not occur in simulation since it does not consider the frictional force.

The new control law for point regulation with a differential robot using a navigation map is

$$\begin{aligned}
 v &= \gamma \|\nabla F(x, y)\| + \frac{1}{\gamma_1} \int_0^t \|\nabla F(x, y)\| dt, \\
 \omega &= -\delta \theta_e \left| \frac{\sin(\theta_e)}{\theta_e} \right|,
 \end{aligned} \tag{17}$$

where  $\gamma, \gamma_1, \delta > 0$  are the control gains and the new term considers the integral action of the gradients of function  $F$  norm evaluated over the point  $(x, y)$ , while the remaining terms are the same as in Equations 15 and 16.

It should be noted that, as the sampling frequency of the microphone arrangement is  $F_s = 16$  KHz, the area of acoustic detection is rather limited, so that the dimensions of the area of work used in the experiments are  $1.5\text{m} \times 1.25\text{m}$ . For this working area, the acoustic navigation map and its gradient field are shown in Figure 4. **Although the area was bound to about one square meter, our methodology based on gradient map integration is not constrained to this space, as the map is adaptable to a wider area and by increasing the number of microphones. Also, the resolution of the grid could be modified in accordance with the localization purposes. In the case of our experiments, and based on the size of the robot, we decided that a convenient resolution was  $5\text{cm} \times 5\text{cm}$ , which is not limited by any means to alternative configurations.**

In order to show the control capability described in Equation 17, eleven tests have been developed. Let  $p_d = (0.0 \text{ m}, 0.2 \text{ m})$  be the desired position, which is within the safety zone and is also the lowest value element in the acoustic navigation map. The point  $p_d$  is also considered a stopping condition, which depends on whether the RMSE between the current position and the desired position is less than a defined tolerance of 0.04 meters.

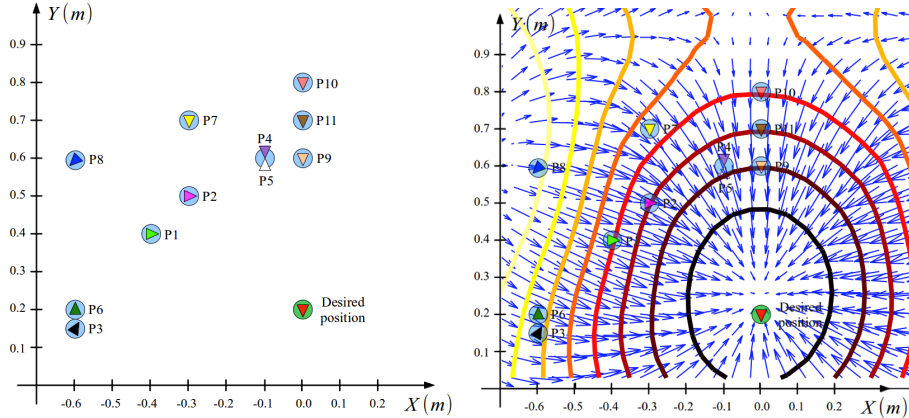


Figure 5.: **Robot initial conditions for all tests.** The final position (0,0.2) is shown in green. The right image shows the configuration of the eleven experiments overlapped over the gradient field of the acoustic navigation map.

Test No.	Initial configuration			Gains		
	$x_0$ (m)	$y_0$ (m)	$\theta_0$ (rad)	$\gamma$	$\delta$	$\gamma_1$
1	-0.4	0.4	0	50	2	1.5
2	-0.3	0.5	0	50	2	1.5
3	-0.6	0.15	$\pi/3$	25	2.5	1.5
4	-0.1	0.6	$-\pi/2$	50	1	1.5
5	-0.1	0.6	$\pi/2$	35	3	1.5
6	-0.6	0.2	$\pi/2$	25	3	1.5
7	-0.3	0.7	$-\pi/2$	50	1	1.5
8	-0.6	0.6	$-3\pi/4$	25	2	1.5
9	0.0	0.6	$-\pi/2$	30	2	1.5
10	0.0	0.8	$-\pi/2$	50	2	1.5
11	0.0	0.7	$-\pi/2$	50	2	1.5

Table 1.: Experimental initial conditions and control gains.

Table 1 shows the initial settings and control gains for each of the tests. These initial configurations of the robot, as well as the desired configuration of the robot, are shown in Figure 5. The control gains were adjusted heuristically using the system response time of the robot kinematics. For linear velocity, the proportional gain is in the range of 10 and the integral gain in the range  $1/\gamma$ . On the other hand, for the angular velocity the proportional gain is in the range of  $10^0$ .

As it is shown in Figure 5, all the initial conditions of the robot are located in the second quadrant of the plane and on the  $y$ -axis. Although no initial conditions are proposed in the first quadrant, due to the symmetry of the acoustic navigation map it is possible to carry out this kind of experimental tests, for brevity of space these will not be presented in this paper. The tests described in the following sections will be referred to as P1 for test 1, P2 for test 2, and so on for a total of 11 tests.

It is important to recall that the original gradient map (Fig. 1.a) is extremely noisy and not useful for any control purposes. In our approach, the smoothed Fourier-based map appears to have a monotonic decrease behavior. However, some areas of the map, specially near the borders, i.e. red zones in Fig. 4.a, exhibit gradients that depart from the flow toward the desired position. For this reason, not all the initial positions and

angles will lead to a monotonic decrease trajectory.

We now present the results obtained for the experimental tests proposed to determine the performance of the point regulation control using the navigation map. These are classified in accordance with the acoustic signal used to estimate the position of the robot by means of the TDOA method.

#### 4.2. *Acoustic signal: 900 Hz square signal*

Using the estimated TDOA position of a sound source emitting a square signal of 900 Hz frequency and placed in the center of mass of a mobile differential robot, experimental tests of the control presented in subsection 4.1 are performed.

Individual of these tests are individually shown in Figure 6. The figure presents the real robot trajectory (in red) measured with an aerial camera and the trajectory estimated by the TDOA (in blue). The first column of the figures shows results related with the test using the 900 Hz square signal, while the second column depicts results of the experiments using the natural sound of the robot.

We commence our analysis by commenting on the 900 Hz square signal tests (left column). It is noticeable in test P3 that due to an erroneous estimated position of the robot, the control considers that the robot has reached the desired position taking into account the tolerance of the error. However, the actual trajectory of the robot shows that the robot does not manage to place itself inside the safety zone even when the control tries to direct it there.

On the other hand, tests P5, P7 and P8, are considered successful, because in these tests the robot is directed by the control to a position near the desired point. This occurs in such a way that when it is estimated that the robot is within the security region of the point, it stops and it is considered that the control task has finished.

In particular, test P5, represents a challenge for control because most of the time the robot stays in an area where its position can not be accurately estimated. However, it manages to steer it because the control speeds cause the robot to perform a curve path originated by the behavior of the gradient of the acoustic navigation map. It is also noticeable how the robot points out in an opposite direction to the closest angle at the map. For this reason, it starts a 180 degrees rotation in the first part of the trajectory. This effect is due to the reference angle  $\theta_r$ , as it is calculated using the *arctan* function (see Eq. 16). As a consequence, the angle error  $\theta_e$  would change direction to the right until it reaches a mirror trajectory towards the desired position. This effect could be avoided by selecting the robot orientation to be facing toward the security zone. Nonetheless, this case is useful at confirming that even in a very bad initialization, the control successfully converges for the 900 Hz square signal case, and reasonably approximates the robot toward the desire position for the case of its natural sound.

In test P8, the control achieves its objective even when the robot has an initial position different from the one considered by the control. However, when such situations arise, it is not always possible for the robot to position itself near the target point, even when the robot moves in the direction of the target, as in test P3. This is because an incorrect estimation of the TDOA method directs the robot erroneously and the control fails to recompose that action.

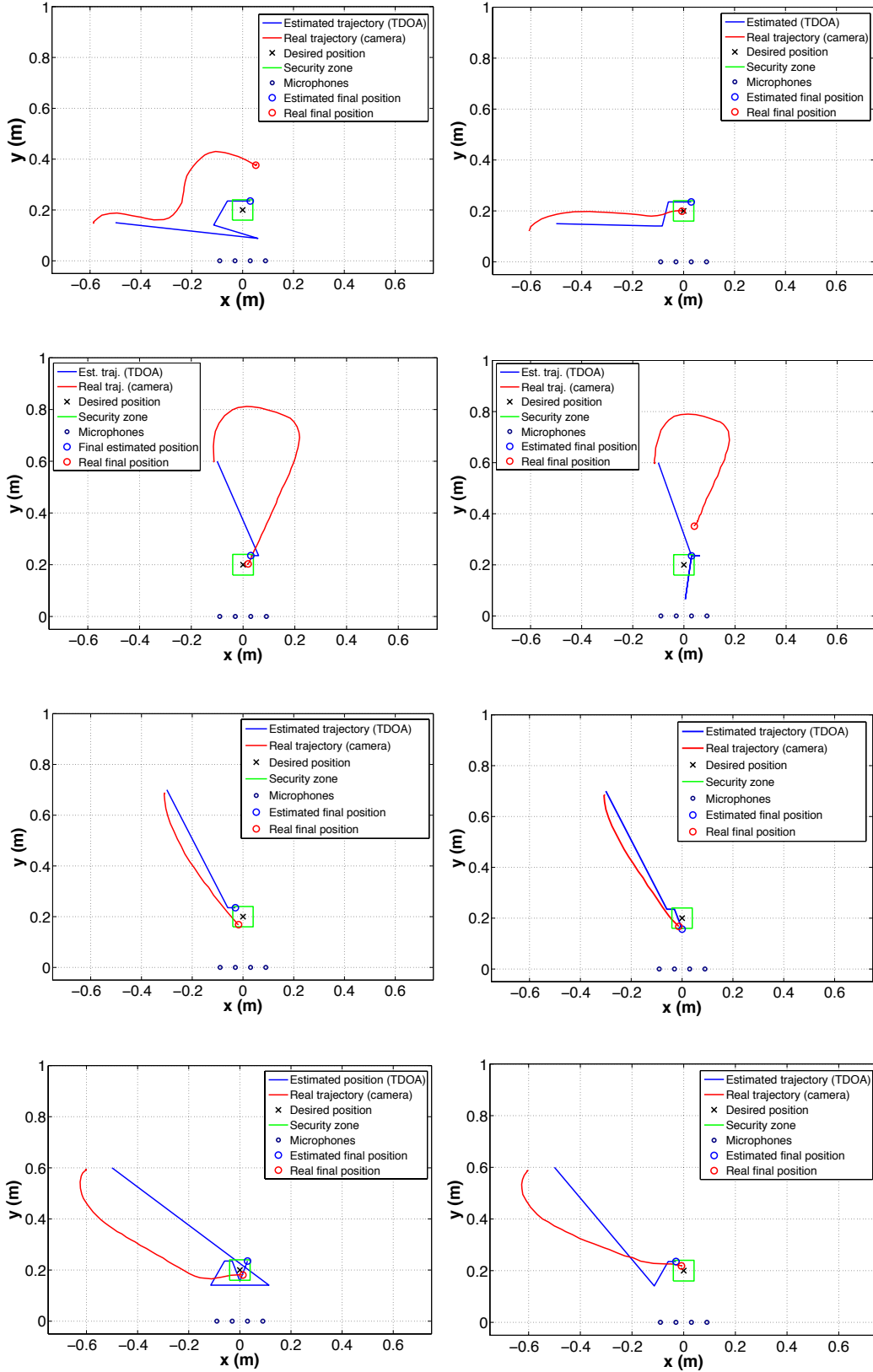


Figure 6.: **Experimental results.** The left column shows the case of using a 900 Hz square signal, while the right column shows the case of using the natural sound of the robot. Tests P3, P5, P7 and P8 are shown from top to bottom.

### 4.3. Acoustic signal: natural robot sound

It is also possible to use the sound produced by the motors and wheels of the moving robot to estimate its location during the control tests. To check the above, in these tests the speaker will not emit any sound. In Figure 6 (right column) we show results related with this case.

Note how, by using the sound of the robot during the tests, successful results are also obtained, as in tests P3, P7 and P8. This is because in these tests the robot is directed by the control to a position close to the desired point in such a way that, when it is estimated that the robot is within the security region, the robot stops and the control task is considered complete.

In particular, in tests P3 and P8, the control achieves its objective even when the robot has an initial position different from that considered by the control. This is corrected once the TDOA method estimates a new position of the robot, thus allowing the control to drive the robot to the desired position.

In test P5, which represents a challenge for control because most of the time the robot stays in an area where its position cannot be accurately estimated, the robot performs a trajectory that drives the robot towards the desired point. However, because the robot is placed in a position that causes an incorrect TDOA estimation, it does not reach the safety in reality. This type of incorrect estimation is due to the low frequency of sampling of the natural sound of the robot.

Table 2 shows the errors calculated by the vision system at the moment when the robot arrives at the desired point or at a location close to it. These errors indicate the distance between the desired position and the final position of the robot for each of the tests performed, as well as the mean square error.

Test No	900 Hz square signal			Robot sound		
	$e_x$ (m)	$e_y$ (m)	$RMSE$ (m)	$e_x$ (m)	$e_y$ (m)	$RMSE$ (m)
P1	-0.02486	0.20756	0.14782	-0.04324	0.22378	0.16116
P2	0.07135	0.21297	0.15882	-0.02162	0.24000	0.17039
P3	0.05189	0.17838	0.13136	-0.00648	0.00108	0.00464
P4	-0.07567	0.17513	0.13490	-0.03567	0.13513	0.09883
P5	0.01945	0.00540	0.01428	0.04108	0.15351	0.11237
P6	-0.04756	0.00864	0.03418	0.05513	0.03999	0.04816
P7	-0.01729	-0.02918	0.02399	-0.01297	-0.03026	0.02328
P8	0.01189	-0.01729	0.01484	-0.00864	0.02053	0.01575
P9	0.00972	-0.02270	0.01746	0.01081	0.04972	0.03598
P10	0.03783	0.06702	0.05442	0.00000	0.05513	0.03898
P11	0.01945	-0.06918	0.05082	0.00216	0.05297	0.03748

Table 2.: Experimental errors.

### 4.4. Discussion

The results presented in this section suggest that, when using the acoustic navigation map as a tool in the point regulation control, the trajectory performed by the robot continues the behavior of the gradient in the acoustic navigation map.

An important feature in the results obtained is that even when the correct starting position of the robot is not known, it is possible to drive it to the desired final position

using the control shown in Equation 17.

In the presented results, it is observed that once the robot has reached the desired position, it is stopped due to the proposed stopping condition, and it is considered that the robot has fulfilled the target when its estimated position is not changed for a minimum of one minute. In addition, this position must be located within the safety zone of the desired point. [An application example for this approach can be found within a multiple sensor fusion scheme, where the robot must reach visual cues, but there may be some regions where camera-based localization alone would not be sufficient, for instance, in texture-less environments. Therefore the design of the safety zone would depend on the placement of easily recognizable landmarks such as QR codes, or color signals, where microphone arrays could be assigned so as to diminish drifting away from a desired localization task.](#)

Although the results indicate that the estimated position of the robot always reaches such safety zone, in some cases the real position of the robot is not the same as the estimated. This is caused by the nature of TDOA, because in some positions of the plane it is not possible to locate the sound source. To improve the accuracy of the estimation it would be necessary to considerably increase the sampled frequency of the information obtained by the microphones.

Some other factors that cause a poor estimation of the position of the robot are stated below:

- In the case of detection of the sound of the robot, there is a possibility that the amplitude of the sound generated by one of the motors is greater than the rest of them, which causes a poor estimation in the position of the robot, that is due to an acoustic phenomenon known as masking.
- When the square signal is used, reverberation problems occur because the signal generated by the speaker impacts directly against the wooden base, which causes the original signal not to be detected, instead, the detected signal is the one that contains the rebounds.
- Another factor that causes this type of problem is the noise generated by sources external to the experimental system. In order to avoid such problems, a preprocessing stage should be implemented, in order to obtain only information from the signal belonging to the robot.

The results obtained by using the two different types of acoustic signals to determine the position of the robot are very similar. However, there are two tests in which for one of the acoustic signals the robot manages to position itself near the target, while when using the other signal the robot ends in a position away from the target. This behavior occurs in tests P3 and P5 and is a consequence of an erroneous estimation of the position of the robot, causing the robot to move in an incorrect direction or to consider that it has reached the desired position.

It should be noted that when using the square signal an acoustic problem arises due to the reverberation generated when the acoustic signal collides with the solids found in the laboratory. This causes in some cases an erroneous estimation of the position of the robot. In order to avoid this situation, it would be necessary to have a working environment that has walls specially designed to avoid the rebound of the acoustic signals, as well having the working area clear of obstacles.

Based on these facts, it can be concluded that the point regulation control with acoustic navigation map presented in this paper manages to lead a mobile robot to a desired position or, failing that, to an area close to it. In addition, with the experiments presented, it is verified that it is possible to apply a new working tool, the acoustic



navigation map, in conjunction with the TDOA acoustic localization method in the area of wheeled robot control.

## 5. Conclusions and future work

We have presented a novel approach to point regulation control based on information from an acoustic navigation map that can be used to drive a mobile robot to a position within the vision area of camera. In addition, surface smoothing tools have been successfully incorporated to generate a new working tool that is capable of providing navigation trails to a mobile robot: the acoustic navigation map. Based on the results obtained in the simulations described in Section 3, we conclude that with a novel approach of the point regulation control in conjunction with the TDOA method of acoustic localization and a map of acoustic navigation as source of information, it is possible to carry out a regulation task at a point without having a predefined path or with a control that depends directly on the errors in the robot states.

The interest of exploring sound source localization, specially the natural noise of the robot, is not for replacing cameras, but for complementing spaces that are out of view. Camera positioning and optimization in closed spaces is an open problem, and it is of great difficulty to cover the whole working space without having “invisible” areas. It is in this sense that our approach would improve the sensing and autonomous capabilities of robots for indoor applications.

In order to increase the area of sound detection, an audio acquisition card could be used to sample, at higher frequencies, the acoustic signal detected by the microphones. For example, if a 192 kHz sampling frequency is considered, i.e. 12 times greater than the one explored in our experiments, the working area in which a sound source could be detected would be of  $4 \times 5$  meters (20 square meters), meaning an increase of around 10 times our current working area.

As reported, the rate of root mean square error greater than 10 cm was of 3 out of 11 for the natural noise of the robot *vs* 4 out of 11 for the 900Hz square signal. Although this does not represent a 100% success rate, it does suggest the possibility of localizing the robot using its natural sound with a reasonable degree of accuracy, even in scenarios which are not isolated from external noises. The authors consider that, for a much better localization rate more sensors should be considered, for example, cameras mounted on the robot, range finders and bird-eye cameras, among other sensors which could be combined with microphones into a filtering strategy such as Kalman or Particle Filtering.

Finally, it is important to recall one limitation of the method when the robot noise is used. In fact, the presence of other noises, like those from air conditioner and computer fans, which are acoustically similar to the sound produced by the robot, may cause localization to fail. To address this problem, it would be necessary to rely on a robust interference cancellation technique.

## References

- [1] A. Martinelli, N. Tomatis, A. Tapus and R. Siegwart. *Simultaneous localization and odometry calibration for mobile robot*. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, pp. 1499-1504, vol. 2, 2003.
- [2] J. Elseberg, D. Borrmann and A. Nüchter. *Automatic and Full Calibration of Mobile*

- Laser Scanning Systems*. In J.P. Desai, G. Dudek, O. Khatib and V. Kumar, editors, *Experimental Robotics: The 13th International Symposium on Experimental Robotics*, Springer Tracts in Advanced Robotics, pp 907-917, vol. 88, 2013.
- [3] D. Scaramuzza and F. Fraundorfer. *Visual Odometry: Part I - The First 30 Years and Fundamentals*. IEEE Robotics and Automation Magazine, pp. 80-92, vol. 18, no. 4, 2011.
- [4] F. Fraundorfer and D. Scaramuzza. *Visual Odometry: Part II - Matching, Robustness, and Applications*. IEEE Robotics and Automation Magazine, pp. 78-90, vol. 19, no. 2, 2012.
- [5] J. Hyup Ko, S. Do Kim and M. Jin Chung. *A method of indoor mobile robot navigation using acoustic landmarks*. Proceedings IEEE International Conference on Robotics and Automation, pp. 1726-1731, vol. 2, 1996.
- [6] J.M. Valin, F. Michaud, J. Rouat and D. Letourneau. *Robust sound source localization using a microphone array on a mobile robot*. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, pp. 1228-1233, vol. 2, 2003.
- [7] J- ShengHua, C. YuChana, C. KangWanga, M. TangLeea and C. Yi Kuo *Simultaneous Localization of a Mobile Robot and Multiple Sound Sources Using a Microphone Array*, Advanced Robotics, 25, pp.135-152, 2011.
- [8] Yuanqing Lin, P. Vernaza, Jihun Ham and D.D. Lee. *Cooperative relative robot localization with audible acoustic sensing*. IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, pp. 3764-3769, 2005.
- [9] DongJu Kim and JongSuk Choi. *Multi-robot team outdoor localization using active marker and high frequency signal sources*. 11th International Conference on Control, Automation and Systems, ICCAS, pp. 192-196, 2011.
- [10] Jing Guo, Qing-Hao Meng, Yu-Xiu Wu, Wei-Xing Yang, Ming Zeng and Wei Li. *Hearing based relative localization for mobile robots in outdoor environments*. IEEE International Conference on Robotics and Biomimetics, ROBIO, pp. 2126-2131. 2012.
- [11] K. Nakamura, K. Nakadai and Hiroshi G. Okuno *A real-time super-resolution robot audition system that improves the robustness of simultaneous speech recognition*, Advanced Robotics, 27:12, 933-945, 2013.
- [12] D. Macho, J. Padrell, A. Abad, C. Nadeu, J. Hernando, J. McDonough, M. Wölfel, U. Klee, M. Omologo, A. Brutti, P. Svaizer, G. Potamianos, and S. M. Chu. *Automatic speech activity detection, source localisation, and speech recognition on the CHIL seminar corpus*. IEEE International Conference on Multimedia and Expo, pp. 876-879, 2005.
- [13] R. Chakraborty and C. Nadeu. *Sound-model based acoustic source localization using distributed microphone arrays*. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, pp. 619-623, 2014.
- [14] M. Omologo and P. Svaizer. *Use of the cross-power spectrum phase in acoustic event location*. IEEE Transactions on Speech and Audio Processing, pp. 288-292, vol. 5, no. 3, 1997.
- [15] M. S. Brandstein and H. F. Silverman. *A practical methodology for speech source localization with microphone arrays*. Computer Speech & Language Journal, pp. 91-126, vol. 11, no. 2, 1997.
- [16] J. P. Dmochowski and J. Benesty. *Steered beamforming approaches for acoustic source localization*. In *Speech Processing in Modern Communication*, Springer Topics in Signal Processing pp 307-337, vol. 3, 2010.
- [17] H. Krim and M. Viberg. *Two decades of array signal processing research: The parametric approach*. IEEE Signal Processing Magazine, vol. 13 (4), pp. 67-94, 1996.
- [18] C. H. Knapp and G. C. Carter. *The Generalized Correlation Method for Estimation of Time Delay*. IEEE Transactions on Acoustics, Speech, and Signal Processing, TASSP, vol. ASSP-24, no. 4, 1976.
- [19] Yuanqing Lin, P. Vernaza, Jihun Ham and D.D. Lee. *Cooperative relative robot localization with audible acoustic sensing*. IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, pp. 3764-3769, 2005.
- [20] R. T. Frankot and R. Chellappa. *A method for enforcing integrability in shape from*

- shading algorithms*. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 439-451, vol. 10, no. 4, 1988.
- [21] R. Klette and K. Schlüns. *Height data from gradient fields*. Proceedings of SPIE (the international Society for Optical Engineering) on Machine Vision Applications, Architectures, and Systems Integration, pp. 204-215, 1996.
- [22] G. Arechavaleta, A. Morales-Díaz, H. Pérez and M. Castelán. *Hierarchical Task-Based Control of Multirobot Systems With Terminal Attractors*. IEEE Transactions on Control Systems Technology, 25 (1), pp. 334 - 341, 2017,
- [23] M. Kumon, K. Fukushima, S. Kunimatsu and M. Ishitobi. *Motion planning based on simultaneous perturbation stochastic approximation for mobile auditory robots*, Intelligent Robots and Systems (IROS) IEEE/RSJ International Conference on, pp. 431-436, 2010.
- [24] A. Magassouba, N. Bertin and F. Chaumette. *Sound-based control with two microphones*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg pp. 5568-5573, 2015.
- [25] K. Sekiguchi, Y. Bando, K. Itoyama, and K.i Yoshii. *Optimizing the Layout of Multiple Mobile Robots for Cooperative Sound Source Separation*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, pp. 5548-5554, 2015.
- [26] K. Sekiguchi, Y. Bando, K. Nakamura, K. Nakadai, K. Itoyama and K.i Yoshii. *Online Simultaneous Localization and Mapping of Multiple Sound Sources and Asynchronous Microphone Arrays*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Korea, pp. 1973-1979, 2016.