

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Robust Audio Localization for Mobile Robots in Industrial Environments

Manuel Manzanares, Yolanda Bolea and Antoni Grau  
*Technical University of Catalonia, UPC, Barcelona  
 Spain*

## 1. Introduction

For autonomous navigation in workspace, a mobile robot has to be able to know its position in this space in a precise way that means that the robot must be able to self-localize to move and perform successfully the different entrusted tasks. At present, one of the most used systems in open spaces is the GPS navigation system; however, in indoor spaces (factories, buildings, hospitals, warehouses...) GPS signals are not operative because their intensity is too weak. The absence of GPS navigation systems in these environments has stimulated the development of new local positioning systems with their particular problems. Such systems have required in many cases the installation of beacons that operate like satellites (similar to GPS), the use of landmarks or even the use of other auxiliary systems to determine the robot's position.

The problem of mobile robot localization is a part of a more global problem because in autonomous navigation when a robot is exploring an unknown environment, it usually needs to obtain some important information: a map of the environment and the robot's location in the map. Since mapping and localization are related to each other, these two problems are usually considered as a single problem called simultaneous localization and mapping (SLAM). The problem of Simultaneous Localization and Map Building is a significant open problem in mobile robotics which is difficult because of the following paradox: to localize itself the robot needs the map of the environment, and, for building a map the robot location must be known precisely.

Mobile robots use different kinds of sensors to determine their position: for instance it is very common the use of odometric or inertial sensors, however it is remarkable to consider that in wheel slippage, sensor drifts a noise causing error accumulation, thus leading to erroneous estimates. Another kind of external sensors used in robotics in order to solve localization are for instance CCD cameras, infrared sensor, ultra sonic sensor, mechanical wave and laser. Other sensors recently applied are the instruments sensible to the magnetic field known as the electronic compass (Navarro & Benet, 2009). Mobile robotics are interested on those able to measure the Earth's magnetic field and express it through an electrical signal. One type of electronic compass is based on magneto-resistive transducers, whose electrical resistance varies with the changes on the applied magnetic field. This type of sensors presents sensitivities below 0.1 milligauss, with response times below 1 sec, allowing its reliable use in vehicles moving at high speeds (Caruso, 2000). In SLAM some applications with electronic compass have been developed working simultaneously with other sensors such as artificial vision (Kim et al., 2006) and ultrasonic sensors (Kim et al., 2007).

In mobile robotics, due to the use of different sensors at the same time to provide localization information the problem of data fusion rises and many algorithms have been implemented. Multisensor fusion algorithms can be broadly classified as follows: estimation methods, classification methods, inference methods, and artificial intelligence methods (Luo et al., 2002); in the latter are remarkable neural networks, fuzzy and genetic algorithms (Begum et al., 2006); (Brunskill & Roy, 2005). Related with the provided sensors information processing in SLAM context, many works can be found, for instance in (Di Marco et al., 2000), where estimation of the position of the robot and the selected landmarks are derived in terms of uncertainty regions, under the hypothesis that the errors affecting all sensor measurements are unknown but bounded, or in (Begum et al., 2006) where an algorithm processes sensor data incrementally and therefore, has the capability to work online.

Therefore a comprehensive collection of researches have been reported on SLAM, most of which stem from the pioneer work of (Smith et al. 1990). This early work provides a Kalman Filter (KF) based statistical framework for solving SLAM. The KF based SLAM algorithms require feature extraction and identification from sensor data, for estimating the pose and the parameters. In the situation that the system noise and measurement obey a Gaussian amplitude distribution, KF uses the state recursive equation that is with the noise estimates the optimal attitude of mobile robots. But there would be generated errors of localization, if the noise does not obey the distribution. KF is also able to the merge low graded multisensor data models. Particle filter is the next probabilistic technique that has earned popularity in SLAM literature. The hybrid SLAM algorithm proposed in (Thrun, 2001) uses particle filter for posterior estimation over a robot's poses and is capable to map large cyclic environments. Another method of fusion broadly used is Extended Kalman Filter (EKF); the EKF can be used where the model is nonlinear, but it can be suitably linearized around a stable operating point.

Several systems have been researched to overcome the localization limitation. For example, the Cricket Indoor Location (Priyantha, 2000) which relies on active beacons placed in the environment. These beacons transmit simultaneously two signals (a RF and an ultrasound wave). Passive listeners mounted, for example, on mobile robots can, by knowing the difference in propagation speed of the RF and ultrasound signals, estimate their own position in the environment. GSM and WLAN technologies can also be used for localization. Using triangulation methods and measuring several signal parameters such as the signal's angle and time of arrival, it becomes possible to estimate the position of a mobile transmitter/receiver in the environment (Sayed et al., 2005). In (Christo et al., 2009), a specific architecture is suggested for the use of multiples iGPS Web Services for mobile robots localization.

Most of the mobile robot's localization systems are based on robot vision, and robot vision is also a hot spot in the research of robotics. Camera which is the most popular visual sensor is widely used for the localization of mobile robots just now. However some difficulties occur because of the limitation of camera's visual field and the dependence on light condition. If the target is not in the visual field of camera or the lighting condition is poor, the visual localization system of the mobile robot cannot work effectively. Nowadays, the role of acoustic perception in autonomous robots, intelligent buildings and industrial environments is increasingly important and in the literature there are different works (Yang et al., 2007); (Mumolo et al., 2003); (Cyszewski, 2003).

Comparing to the study on visual perception, the study on auditory is still in its infancy stage. The human auditory system is a complex and organic information processing system,

it can feel the intensity of sound and space orientation information. Compared with vision, audition has several unique properties. Audition is omni-directional. The sound waves have strong diffraction ability; audition also is less affected by obstacles. Therefore, the audio ability possessed by robot can make up the restrictions of other sensors such as limited view or the non-translucent obstacles. Nevertheless, audio signal processing presents some particular problems such as the effect of reverberations and noise signals, complex boundary conditions and near-field effect, among others, and therefore the use of audio sensors together with other sensors is common to determine the position and also for autonomous navigation of a mobile robot, leading to a problem of data fusion. There are many applications that would be aided by the determination of the physical position and orientation of users. As an example, without the information on the spatial location of users in a given environment, it would not be possible for a service robot to react naturally to the needs of the user. To localize a user, sound source localization techniques are widely used. Such techniques can also help a robot to self-localize in its working area. Therefore, the sound source localization (one or more sources) has been studied by many researchers (Ying & Runze, 2007); (Sasaki et al., 2006); (Kim et al., 2009). Sound localization can be defined as the process of determining the spatial location of a sound source based on multiple observations of the received sound signals. Current sound localization techniques are generally based upon the idea of computing the time difference of arrival (TDOA) information with microphone arrays (Brandstein & Silverman, 1997); (Knapp & Carter, 1976), or interaural time difference (ITD) (Nakashima & Mukai, 2005). The ITD is the difference in the arrival time of a sound source between two ears, a representative application can be found in (Kim & Choi, 2009) with a binaural sound localization system using sparse coding based ITD (SITD) and self-organizing map (SOM). The sparse coding is used for decomposing given sounds into three components: time, frequency and magnitude, and the azimuth angle are estimated through the SOM. Other works in this field use structured sound sources (Yi & Chu-na, 2010) or the processing of different audio features (Rodemann et al., 2009), among other techniques.

The works that authors present in this Chapter are developed with audio signals generated with electric machines that will be used to mobile robots localization in industrial environments. A common problem encountered in industrial environments is that the electric machine sounds are often corrupted by non-stationary and non-Gaussian interferences such as speech signals, environmental noise, background noise, etc. Consequently, pure machine sounds may be difficult to identify using conventional frequency domain analysis techniques as Fourier transform (Mori et al., 1996), and statistical techniques such as Independent Component Analysis (ICA) (Roberts & Everson, 2001).

The wavelet transform has attracted increasing attention in recent years for its ability in signal features extraction (Bolea et al., 2003); (Mallat & Zhang, 1993), and noise elimination (Donoho, 1999). While in many mechanical dynamic signals, such as the acoustical signals of an engine, Donoho's method seems rather ineffective, the reason for their inefficiency is that the feature of the mechanical signals is not considered. Therefore, when the idea of Donoho's method and the sound feature are combined, and a de-noising method based on Morlet wavelet is added, this methodology becomes very effective when applied to an engine sound detection (Lin, 2001). In (Grau et al., 2007), the authors propose a new approach in order to identify different industrial machine sounds, which can be affected by non-stationary noise sources.

It is also important to consider that non-speech audio signals have the property of non-stationary signals in the same way that many real signals encountered in speech processing, image processing, ECG analysis, communications, control and seismology. To represent the behaviour of a stationary process is common the use of models (AR, ARX, ARMA, ARMAX, OE, etc.) obtained from the experimental identification (Ljung, 1987). The coefficient estimation can be done with different criteria: LSE, MLE, among others. But in the case of non-stationary signals the classical identification theory and its results are not suitable.

Many authors have proposed different approaches to modelling this kind of non-stationary signals, that can be classified: i) assuming that a non stationary process is locally stationary in a finite time interval so that various recursive estimation techniques (RLS, PLR, RIV, etc.) can be applied (Ljung, 1987); ii) a state space modelling and a Kalman filtering; iii) expanding each time-varying parameter coefficients onto a set of basis sequences (Charbonnier et al., 1987); and iv) nonparametric approaches for non-stationary spectrum estimation such a local evolving spectrum, STFT and WVD are also developed to characterize non-stationary signals (Kayhan et al., 1994).

To overcome the drawbacks of the identification algorithms, wavelets could be also considered for time varying model identification. The distinct feature of a wavelet is its multiresolution characteristic that is very suitable for non-stationary signal processing (Tsatsanis & Giannakis, 1993).

The work to be presented in this Chapter will investigate different approaches based on the study of audio signals with the purpose of obtaining the robot location (in x-y plane) using as sound sources industrial machines. For their own nature, these typical industrial machines produce a stationary signal in a certain time interval. These resultant stationary waves depend on the resonant frequencies in the plant (depending on the plant geometry and dimensions) and also on the different absorption coefficients of the wall materials and other objects present in the environment.

A first approach that authors will investigate is based on the recognition of patterns in the acquired audio signal by the robot in different locations (Bolea et al., 2008). These patterns will be found through a process of feature extraction of the signal in the identification process. To establish the signal models the wavelet transform will be used, specifically the Daubechies wavelet, because it captures very well the characteristics and information of the non-speech audio signals. This set of wavelets has been extensively used because its coefficients capture the maximum amount of the signal energy.

A MAX model (Moving Averaging Exogenous) represents the sampled signals in different points of the space domain because the signals are correlated. We use the closest signal to the audio source as signal input for the model. Only the model coefficients need to be stored to compare and to discriminate the different audio signals. This would not happen if the signals were represented by an AR model because the coefficients depend on the signal itself and, with a different signal in every point in the space domain, these coefficients would not be significant enough to discriminate the audio signals. When the model identification is obtained by wavelets transform, the coefficients that do not give information enough for the model are ignored.

The eigenvalues of the covariance matrix are analyzed and we reject those coefficients that do not have discriminatory power. For the estimation of each signal the approximation signal and its significant details are used following the next process: i) model structure selection; ii) model parameters calibration with an estimation model (the LSE method can be



used for its simplicity and, furthermore a good identified model coefficients convergence is assured); iii) validation of the model.

Another approach that will also be investigated is based on the determination of the transfer function of a room, denoted RTF (Room Transfer Function), this model is an LPV (Linear Parameters Varying) because the parameters of the model vary along the robot's navigation (Manzanares et al., 2009).

In an industrial plant, there are different study models in order to establish the transmission characteristics of a sound between a stationary audio source and a microphone in closed environments: i) the beam theory applied to the propagation of the direct audio waves and reflected audio waves in the room (Kinsler et al., 1995); ii) the development of a lumped parameters model similar to the model used to explain the propagation of the electromagnetic waves in the transmission lines (Kinsler et al., 1995) and the study of the solutions given by the wave equation (Kuttruff, 1979). Other authors propose an RTF function that carries out to industrial plant applied sound model (Haneda et al., 1992); (Haneda et al., 1999); (Gustaffson et al., 2000). In these works the complexity to achieve the RTFs is evident as well as the need of a high number of parameters to model the complete acoustic response for a specific frequency range, moreover to consider a real environment presents an added difficulty.

In this research we study how to obtain a real plant RTF. Due that this RTF will be used by a mobile robot to navigate in an industrial plant, we have simplified the methodology and our goal is to determinate the x-y coordinates of the robot. In such a case, the obtained RTF will not present a complete acoustic response, but will be powerful enough to determine the robot's position.

## 2. Method based on the recognition of patterns of the audio signal

This method is based on the recognition of patterns in the acquired audio signal by the robot in different locations, to establish the signals models the Daubechies wavelets will be used. A MAX model (Moving Averaging Exogenous) represents the sampled signals in different points of the space domain, and for the estimation of each signal the approximation signal and its significant details are used following the process steps mentioned previously: i) model structure selection; ii) model parameters calibration with an estimation model; iii) validation of the model.

Let us consider the following TV-MAX model and be  $S_i = y(n)$ ,

$$y(n) = \sum_{k=0}^q b(n;k)u(n-k) + \sum_{k=0}^r c(n;k)e(n-k) \quad (1)$$

where  $y(n)$  is the system output,  $u(n)$  is the observable input, which is assumed as the closest signal to the audio source, and  $e(n)$  is a noise signal. The second term is necessary whenever the measurement noise is colored and needs further modeling. The coefficients for the different models will be used as the feature vector, which can be defined as  $X_S$ , where

$$X_S = (b_1, b_2, \dots, \overset{q+1}{\dots}, c_1, c_2, \dots, \overset{r+1}{\dots}) \quad (2)$$

where  $q+1$  and  $r+1$  are the amount of  $b$  and  $c$  coefficients respectively. From every input signal a new feature vector is obtained representing a new point in the  $(q+r+2)$ -dimensional

feature space,  $f_s$ . For feature selection, it is not necessary to apply any statistical test to verify that each component of the vector has enough discriminatory power because this step has been already done in the wavelet transform preprocessing.

This feature space will be used to classify the different audio signals entering the system. Some labeled samples with their precise position in the space domain are needed. In this chapter a specific experiment is shown. When an unlabeled sample enters the feature space, the minimum distance to a labeled sample is computed and this measure of distance will be used to estimate the distance to the same sample in the space domain. For this reason a transformation function  $f_T$  is needed which converts the distance in the feature space in the distance in the space domain, note that the distance is a scalar value, independently of the dimension of the space where it has been computed.

The Euclidean distance is used, and the distance between samples  $S_i$  and  $S_j$  in the feature space is defined as

$$d_{fs}(S_i, S_j) = \sqrt{\sum_{k=0}^q (b_{kS_i} - b_{kS_j})^2 + \sum_{k=0}^r (c_{kS_i} - c_{kS_j})^2} \quad (3)$$

where  $b_{kS_i}$  and  $c_{kS_i}$  are the  $b$  and  $c$  coefficients, respectively, of the wavelet transform for the  $S_i$  signal. It is not necessary to normalize the coefficients before the distance calculation because they are already normalized intrinsically by the wavelet transformation.

Because there exist the same relative distances between signals with different models, and with the knowledge that the greater the distortion the farther the signal is from the audio source, we choose those correspondences ( $d_{xy}$ ,  $d_{fs}$ ) between the samples that are closest to the audio source equidistant in the  $d_{xy}$  axis. These points will serve to estimate a curve of  $n$ -order, that is, the transformation function  $f_T$ . An initial approximation for this function is a polynomial of 4th order and there are several solutions for a unique distance in the feature space, that is, it yields different distances in the  $x$ - $y$  space domain.

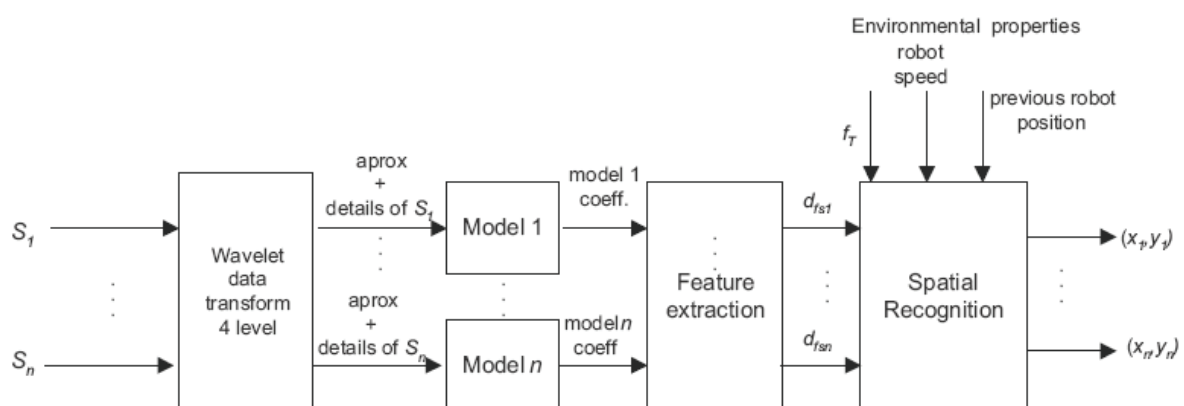


Fig. 1. Localization system in space domain from non-speech audio signals.

We solve this drawback adding a new variable: previous position of the robot. If we have an approximate position of the robot, its speed and the computation time between feature extraction samples, we will have a coarse approximation of the new robot position, coarse enough to discriminate among the solutions of the 4th-order polynomial. In the experiments section a waveform for the  $f_T$  function can be seen, and it follows the model from the sound derivative partial equation proposed in (Kinsler et al., 1995) and (Kuttruff, 1979).

In Figure 1 the localization system can be shown, including the wavelet transformation block, the modeling blocks, the feature space and the spatial recognition block which has as input the environment of the robot and the function  $f_T$ .

## 2.1 Sound source angle detection

As stated in the Introduction section, in order to locate sound sources several works have been developed using a microphone array. Because we work with a unique source of sound, and in order to simplify the number of sensors, we propose a system that detects the direction in which the maximum sound intensity is received and, in this way, emulating the response of a microphone array located in the perimeter of a circular platform. To achieve this effect we propose a turning platform with two opposed microphones. The robot computes the angle respect the platform origin ( $0^\circ$ ) and the magnetic north of its compass. Figure 2 depicts the blocks diagram of the electronic circuit to acquire the sound signals. The signal is decoupled and amplified in a first stage in order to obtain a suitable range of work for the following stages. Then, the maximum of the mean values of the rectified sampled audio signal determines the position of the turning platform.

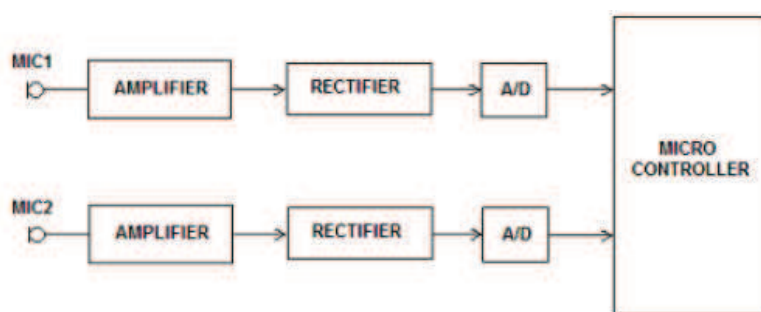


Fig. 2. Angle detection block diagram.

There are two modes of operation: looking for local values or global values. To find the maximum value the platform must turn  $180^\circ$  (because there are two microphones), this mode warranties that the maximum value is determined but the operation time is longer than using the local value detection, in which the determination is done when the system detects the first maximum. In most of the experiments this latter operation mode is enough.

## 2.2 Spatial recognition

This distance computation between the unlabelled audio sample and labeled ones is repeated for the two closest samples to the unlabelled one. Applying then the transformation function  $f_T$  two distances in the x-y domain are obtained. These distances indicate where the unlabelled sample is located. Now, with a simple process of geometry, the position of the unlabelled sample can be estimated but with a certain ambiguity, see Figure 3. In (Bolea et al., 2003) we used the intersection of three circles, which theoretically gives a unique solution, but in practice these three circles never intersect in a point but in an area that induces to an approximation, and thus, to an error (uncertainty) in the localization point.

The intersection of two circles (as shown in Figure 3) leads to a two-point solution. In the correct discrimination of these points the angle between the robot and the sound source is computed.



Since the robot computes the angle between itself and the sound source, the problem is to identify the correct point of the circles intersection. Figure 4 shows the situation.  $I_1$  and  $I_2$  are the intersection points. For each point the angle respect the sound source is computed ( $\alpha_1$  and  $\alpha_2$ ), because the exact source position is known ( $x_s, y_s$ ).

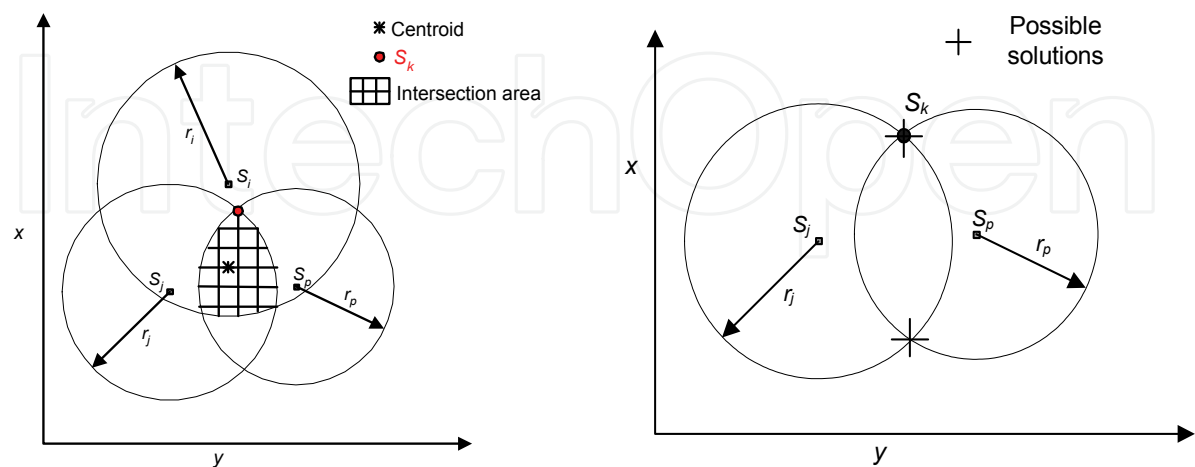


Fig. 3. Geometric process of two (right) or three (left) circles intersection to find the position of unlabeled sample  $S_k$ .

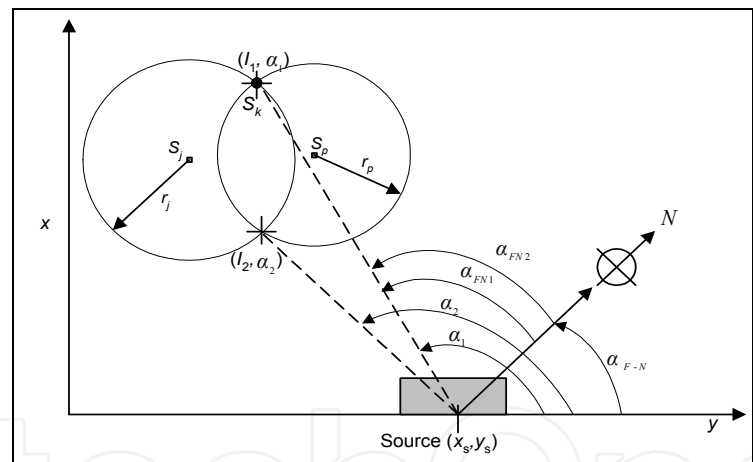


Fig. 4. Angles computation between ambiguous robot localization and sound source. Angles  $\alpha_1$  and  $\alpha_2$  correspond to:

$$\alpha_1 = \arctg \frac{y_{I_1} - y_s}{x_{I_1} - x_s}, \quad \alpha_2 = \arctg \left( \frac{y_{I_2} - y_s}{x_{I_2} - x_s} \right) \tag{4}$$

These angles must be corrected respect the north in order to have the same offset than the angle computed aboard the robot:

$$\alpha_{FN1} = \alpha_1 - \alpha_{F-N}; \quad \alpha_{FN2} = \alpha_2 - \alpha_{F-N} \tag{5}$$

being  $\alpha_{F-N}$  the angle between the room reference and the magnetic north (previously calibrated).

Now, to compute the correct intersection point is only necessary to find the angle which is closer to the angle computed on the robot with the sensor.

### 3. Method based on the LPV model with audio features

In this second approach we study how to obtain a real plant RTF. Due that this RTF will be used by a mobile robot to navigate in an industrial plant, we have simplified the methodology and our goal is to determinate the x-y coordinates of the robot. In such a case, the obtained RTF will not present a complete acoustic response, but will be powerful enough to determine the robot's position. The work investigates the feasibility of using sound features in the space domain for robot localization (in x-y plane) as well as robot's orientation detection.

#### 3.1 Sound model in a closed room

The acoustical response of a closed room (with rectangular shape), where the dependence with the pressure in a point respect to the defined  $(x,y,z)$  position is represented by the following wave equation:

$$L_x \frac{\partial^2 p}{\partial x^2} + L_y \frac{\partial^2 p}{\partial y^2} + L_z \frac{\partial^2 p}{\partial z^2} + k^2 p = 0 \quad (6)$$

$L_x$ ,  $L_y$  and  $L_z$  denote the dimensions of the length, width and height of the room with ideally rigid walls where the waves are reflected without loss, Eq. (6) is rewritten as:

$$p(x, y, z) = p_1(x)p_2(y)p_3(z) \quad (7)$$

when the evolution of the pressure according to the time is not taken into account.

Then Eq. (7) is replaced in Eq. (6), and three differential equations can be derived and it is the same for the boundary condition. For example,  $p_1$  must satisfy the equation:

$$\frac{d^2 p_1}{dx^2} + k_x^2 p_1 = 0 \quad (8)$$

With boundary conditions in  $x = 0$  and  $x = L_x$ :

$$\frac{dp_1}{dx} = 0$$

$k_x$ ,  $k_y$  and  $k_z$  constants are related by the following expression:

$$k_x^2 + k_y^2 + k_z^2 = k^2 \quad (9)$$

Equation (8) has as general solution:

$$p_1(x) = A_1 \cos(k_x x) + B_1 \sin(k_x x) \quad (10)$$

Through Eq. (8) and limiting this solution to the boundary conditions, constants in Eq. (10) take the following values:

$$k_x = \frac{n_x \pi}{L_x}; k_y = \frac{n_y \pi}{L_y} \text{ and } k_z = \frac{n_z \pi}{L_z}$$

being  $n_x$ ,  $n_y$  and  $n_z$  positive integers. Replacing these values in Eq. (10) the wave equation eigenvalues are obtained:

$$k_{n_x n_y n_z} = \pi \left[ \left( \frac{n_x}{L_x} \right)^2 + \left( \frac{n_y}{L_y} \right)^2 + \left( \frac{n_z}{L_z} \right)^2 \right]^{1/2} \quad (11)$$

The eigenfunctions or normal modes associated with these eigenvalues are expressed by:

$$p_{n_x n_y n_z}(x, y, z) = C_1 \cdot \cos\left(\frac{n_x \pi x}{L_x}\right) \cdot \cos\left(\frac{n_y \pi y}{L_y}\right) \cdot \cos\left(\frac{n_z \pi z}{L_z}\right) \cdot e^{j\omega t} \quad (12)$$

$$e^{j\omega t} = \cos(\omega t) - j \sin(\omega t)$$

being  $C_1$  an arbitrary constant and introducing the variation of pressure in function of the time by the factor  $e^{j\omega t}$ . This expression represents a three dimensional stationary wave space in the room. Eigenfrequencies corresponding to Eq. (11) eigenvalues can be expressed by:

$$f_{n_x n_y n_z} = \frac{c}{2\pi} k_{n_x n_y n_z}$$

$$f_{n_x n_y n_z} = \sqrt{f_{nx}^2 + f_{ny}^2 + f_{nz}^2} \quad (13)$$

$$f_{n_x n_y n_z} = \sqrt{\left(\frac{n_x c}{2L_x}\right)^2 + \left(\frac{n_y c}{2L_y}\right)^2 + \left(\frac{n_z c}{2L_z}\right)^2}$$

where  $c$  is the sound speed. Therefore, the acoustic response of any close room presents resonance frequencies (eigenfrequencies) where the response of a sound source emitted in the room at these frequencies is the highest. The eigenfrequencies depend on the geometry of the room and also depend on the materials reflection coefficients, among other factors. Microphones obtain the environmental sound and they are located at a constant height ( $z_1$ ) respect the floor, and thus the factor:

$$\cos\left(\frac{n_z \pi z_1}{L_z}\right) \quad (14)$$

is constant and therefore, if temporal dependency pressure respect the time is not considered, Eq. (12) is:

$$p_{n_x n_y n_z}(x, y) = C_2 \cdot \cos\left(\frac{n_x \pi x}{L_x}\right) \cdot \cos\left(\frac{n_y \pi y}{L_y}\right) \quad (15)$$

In our experiments,  $L_x = 10.54\text{m}$ ,  $L_y = 5.05\text{m}$  and  $L_z = 4\text{m}$ , considering a sound speed propagation of  $345\text{m/s}$ . When Eq. (15) is applied in the experiments rooms, for mode (1, 1,

2), this equation indicates the acoustic pressure in the rooms depending on the x-y robot's position, and this is:

$$p_{n_x n_y n_z}(x, y) = C_2 \cdot \cos\left(\frac{\pi x}{10,54}\right) \cdot \cos\left(\frac{\pi y}{5,05}\right) \quad (16)$$

With these ideal conditions and for an ideal value for constant  $C_2 = 2$ , the theoretic acoustic response in the rooms for this absolute value of pressure, and for this propagation mode, can be seen in Figure 5.

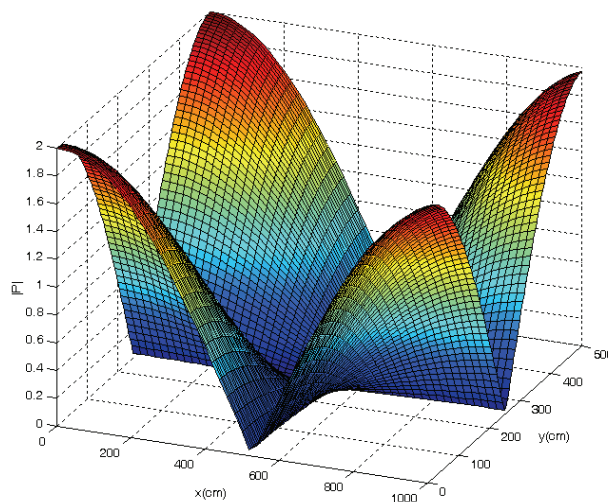


Fig. 5. Room response for propagation mode (1,1,2).

The shape of Figure 5 would be obtained for a sound source that is excited only this propagation mode, really the acoustic response will be more complex as we increase the propagation modes excited by the sound source.

### 3.2 Transfer function in a closed room

In (Gustaffson et al., 2000) a model based in the sum of second order transfer functions is proposed; these functions have been built between a sound source located in a position  $d_s$  emitting an audio signal with a specific acoustic pressure  $P_s$  and a microphone located in  $d_m$  which receives a signal of pressure  $P_m$ ; each function represents the system response in front to a propagation mode.

The first contribution of this work is to introduce an initial variation to this model considering that the sound source has a fixed location, and then this model can be expressed as:

$$\frac{P_m(d_m, s)}{P_s(s)} = \sum_{n=1}^M \frac{K[d_m]s}{s^2 + 2\xi_n \omega_n s + \omega_n^2} \quad (17)$$

Because our objective is not to obtain a complete model of the acoustic response of the industrial plant, it will not be necessary to consider all the propagation modes in the room and we will try to simplify the problem for this specific application without the need to work with models of higher order.

To implement this experiment the first step is to select the frequency of interest by a previous analysis of the audio signal frequency spectrum emitted by the considered sound source (an industrial machine). Those frequency components with a significant acoustic power will be considered with the only requirement that they are close to one of the resonant frequencies of the environment. The way to select those frequencies will be through a band-pass digital filter centered in the frequency of interest. Right now, the term  $M$  in the sum of our model will have the value  $N$ , being this new value the propagation modes resulting from the filtering process.

The spectra of the sound sources used in our experiments show an important component close to the frequency of 100Hz for the climatic chamber, and a component of 50Hz for the PCB insulator, see Figure 10 (right) and Figure 11 (right).

For a concrete propagation mode, the variation that a stationary audio signal receives at different robot's position can be modeled, this signal can be smoothed by the variation of the absorption coefficient of the different materials that make up the objects in the room; those parameters are named  $K[d_m]$  and  $\xi[d_m]$ , and Eq. (17) results:

$$H(s, d_m) = \frac{P_m(d_m, s)}{P_s(s)} = \sum_{n=1}^N \frac{K[d_m]s}{s^2 + 2\xi_n[d_m]\omega_n s + \omega_n^2} \quad (18)$$

where the gain ( $K$ ), smooth coefficient ( $\xi_n$ ) and the natural frequency ( $\omega_n$ ) of the transfer function room system depend on the room characteristics:  $d_m$ ,  $n_x$ ,  $n_y$ ,  $L_x$ , and  $L_y$ , yielding an LPV indoor model.

Using Eq. (17) the module of the closed room in a specific transmission mode  $\omega_{n1}$  is:

$$|H(j\omega_{n1}, d_{m1})| = \frac{K}{2\xi_{n1}\omega_{n1}} \quad (19)$$

The room response in the propagation mode  $\omega_{n1}$  ( $z_1$  is a constant), assuming that the audio source only emits a frequency  $\omega_{n1}$  for a specific coordinate (x,y) of the room, is:

$$|H| = \left| \frac{P_m}{P_s} \right|_{n_x, n_y} = \left| C \cos\left(\frac{n_x \pi x}{L_x}\right) \cos\left(\frac{n_y \pi y}{L_y}\right) \right| \quad (20)$$

with  $f_{n_1} = \sqrt{f_{nx}^2 + f_{ny}^2}$ ,  $\omega_{n_1} = 2\pi f_{n1}$ .

Equating Eq. (19) and (20), it results:

$$\xi_{n1} = \frac{k}{2\omega_{n1} \left| \cos\left(\frac{n_x \pi x}{L_x}\right) \cos\left(\frac{n_y \pi y}{L_y}\right) \right|} \quad (21)$$

If the filter is non-ideal then more than one transmission mode could be considered and therefore the following expression is obtained:

$$\sum_{l=1}^m \frac{K_{nl}}{2\xi_{nl}\omega_{nl}} = \sum_{l=1}^m C \left| \cos\left(\frac{n_{xl}\pi x}{L_x}\right) \cos\left(\frac{n_{yl}\pi y}{L_y}\right) \right| \quad (22)$$



The best results in the identification process in order to determine the robot's position have been obtained, for each considered propagation mode, keeping  $K[d_m]$  coefficient constant and observing the different variations in the acquired audio signal in the smoothing coefficient  $\xi[d_m]$ .

If the zeros of the system are forced to be constant in the identification process for different robot's locations, and we admit that the emitted signal power by the sound sources is also constant and the audio signal power acquired with the microphones varies along the robot's position, then the pole positions in the  $s$  plane, for the considered propagation mode, will vary in the different robot's positions and their values will be:

$$s_{1n}[d_m] = -\xi_n[d_m]\omega_n + \omega_n\sqrt{(\xi_n[d_m])^2 - 1} \quad (23)$$

$$s_{2n}[d_m] = -\xi_n[d_m]\omega_n - \omega_n\sqrt{(\xi_n[d_m])^2 - 1} \quad (24)$$

It is worth noting that this model of reduced order gives good results in order to determine the robot's position and, although it does not provide a complete physical description of the evolution of the different parameters in the acoustic response for the different robot's positions, we can admit that according to the physical model given by the wave equation in Eq. (16), the modules of the proposed transfer functions will vary following a sinusoidal pattern and the pole position in the  $s$  plane will show those variation in the same fashion.

## 4. Experiments and discussions

### 4.1 Method based on the recognition of patterns of the audio signal

In the first proposed method based on the recognition of patterns of the audio signal, in order to prepare a setting as real as possible, we have used a workshop with a CNC milling machine as non-speech audio source. The room has a dimension of 7 meters by 10 meters and we obtain 9 labeled samples (from  $S_1$  to  $S_9$ ), acquired at regular positions, covering the entire representative workshop surface. With the dimensions of the room, these 9 samples are enough because there is not a significant variance when oversampling.

In Figure 6 the arrangement of the labelled samples can be observed. The robot enters the room, describes a predefined trajectory and gets off. In its trajectory the robot picks four unlabeled samples (audio signals) that will be used as data test for our algorithms ( $S_{10}$ ,  $S_{11}$ ,  $S_{12}$  and  $S_{13}$ ). The sample frequency is 8 kHz following the same criteria as (Bielińska, 2002) in order to choose the sampling frequency because its similarity to speech signals.

First, in order to obtain the 9 models coefficients corresponding to the 9 labeled non-stationary audio signals, these signals are decomposed by the wavelet transform in 4 levels, with one approximation signal and 4 detail signals, Figure 7. For the whole samples, the relevance of every signal is analyzed. We observe the more significant decomposition to formulate the prediction model, that is, those details containing the more energy of the signal. With the approximation ( $A4_i$ ) and the detail signal of 4th level ( $D4_i$ ) is enough to represent the original signal, because the mean and deviation for the  $D3_i$ ,  $D2_i$  and  $D1_i$  detail signals are two orders of magnitude below  $A4_i$  and  $D4_i$ . Figure 7 (bottom left) shows the difference between the original signal and the estimated signal with  $A4_i$  and  $D4_i$ . Practically there is no error when overlapped. In this experiment we have chosen the Daubechies 45 wavelets transform because it yields good results in identification (Tsatsanis & Giannakis, 1993), after testing different Daubechies wavelets.

After an initial step for selecting the model structure, it is determined that the order of the model has to be 20 (10 for the  $A4_i$  and 10 for  $D4_i$  coefficients), and an MAX model has been selected, for the reasons explained above. When those 9 models are calibrated, they are validated with the error criteria of FPE (Function Prediction Error) and MSE (Mean Square Error), yielding values about  $10e(-6)$  and 5% respectively using 5000 data for identification and 1000 for validation. Besides, for the whole estimated models the residuals autocorrelation and cross-correlation between the inputs and residuals are uncorrelated, indicating the goodness of the models.

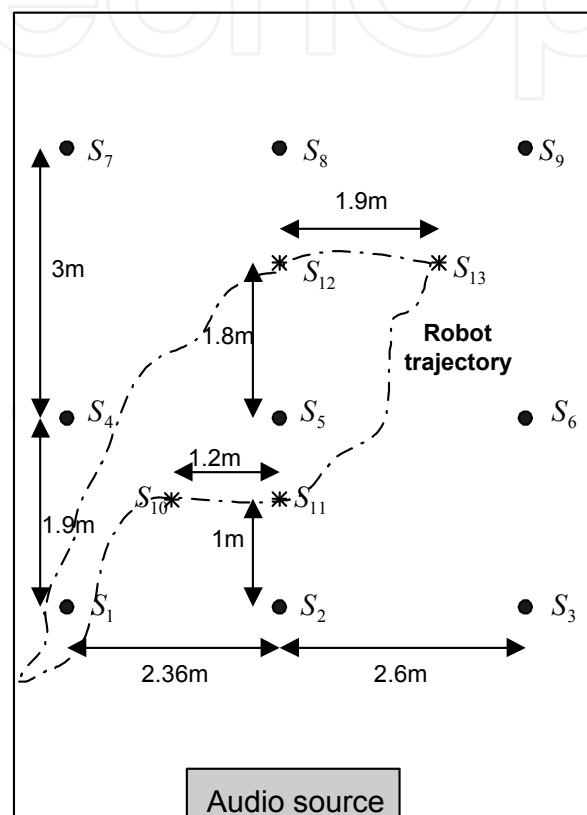


Fig. 6. Robot environment: labeled audio signals and actual robot trajectory with unlabelled signals ( $S_{10}$ ,  $S_{11}$ ,  $S_{12}$ ,  $S_{13}$ ).

These coefficients form the feature space, where the relative distances among all the samples are calculated and related in the way explained in section 2 in order to obtain the transform function  $f_T$ . With these relations, the curve appearing in Figure 8 is obtained, under the minimum square error criteria, approximated by a 4th-order polynomial with the following expression:

$$f_T = d_{fs} = 9.65e(10)d_{xy}^4 + 1.61e(5)d_{xy}^3 - 8.49e(2)d_{xy}^2 + 144.9d_{xy} + 107.84$$

which is related with the solution of the sound equation in (Kinsler et al., 1995); (Kuttruff, 1979) with a physical meaning.

With the transform function  $f_T$  we proceed to find the two minimum distances in the feature space to each unlabelled sample respect the labeled ones, that is, for audio signals  $S_{10}$ ,  $S_{11}$ ,  $S_{12}$  and  $S_{13}$ , respect to  $S_1$ , ...,  $S_9$ .

We obtain four solutions for each signal because each distance in the feature space crosses four times the  $f_T$  curve. In order to discard the false solutions we use the previous position information of the robot, that is the  $(x_i, y_i)_{prev}$  point. We also know the robot speed ( $v = 15\text{cm/sec}$ ) and the computation time between each new position given by the system, which is close to 3 sec. If we consider the movement of the robot at constant speed, the new position will be  $(x_i, y_i)_{prev} \pm (450, 450)\text{mm}$ .

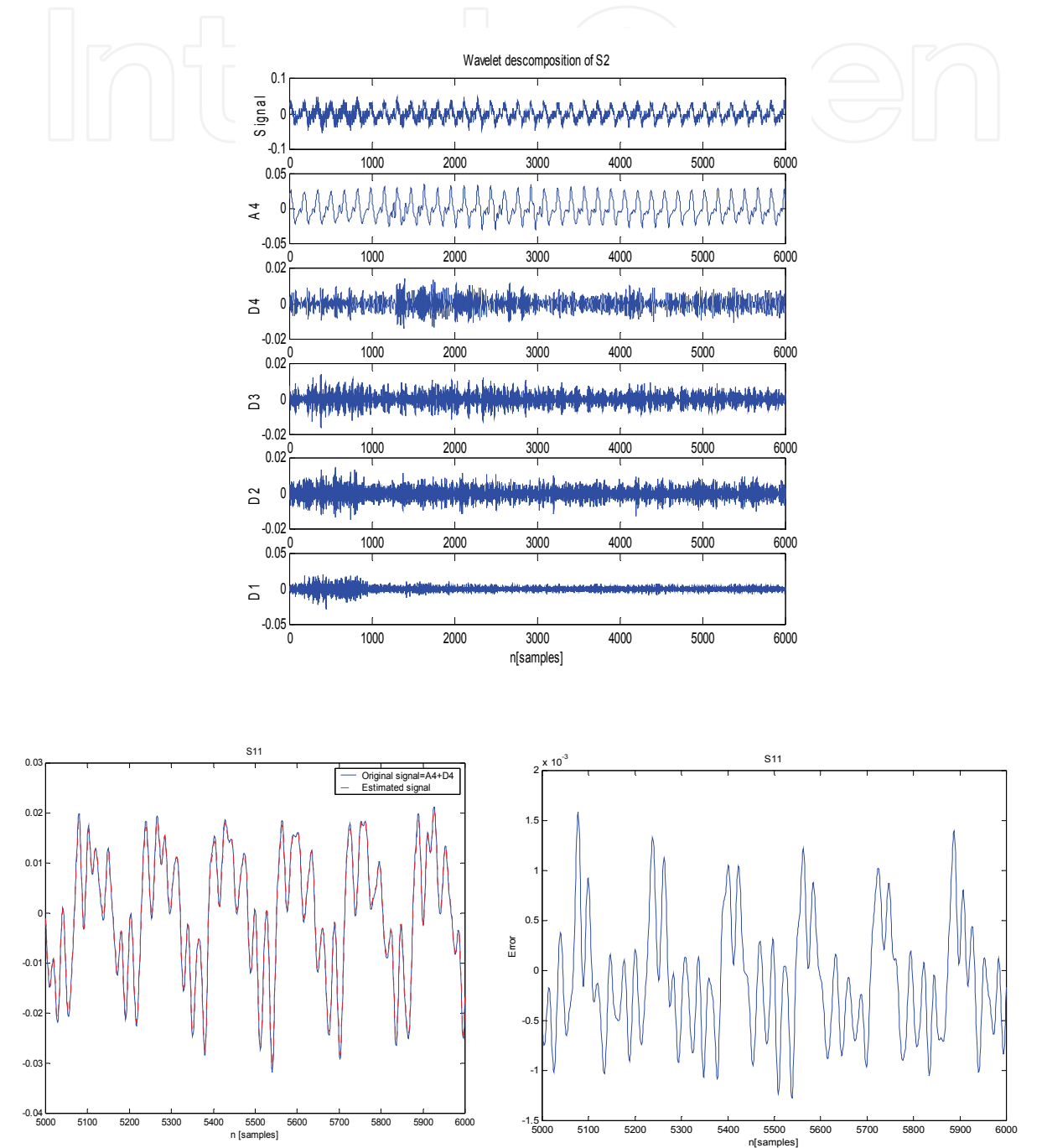


Fig. 7. (Up) Multilevel wavelet decomposition of a non-speech signal ( $S_2$ ) by an approximation signal and four signal details; (down) comparison between (left) original signal ( $A_4+D_4$ ) and the estimated signal and (right) its error for  $S_{11}$ .

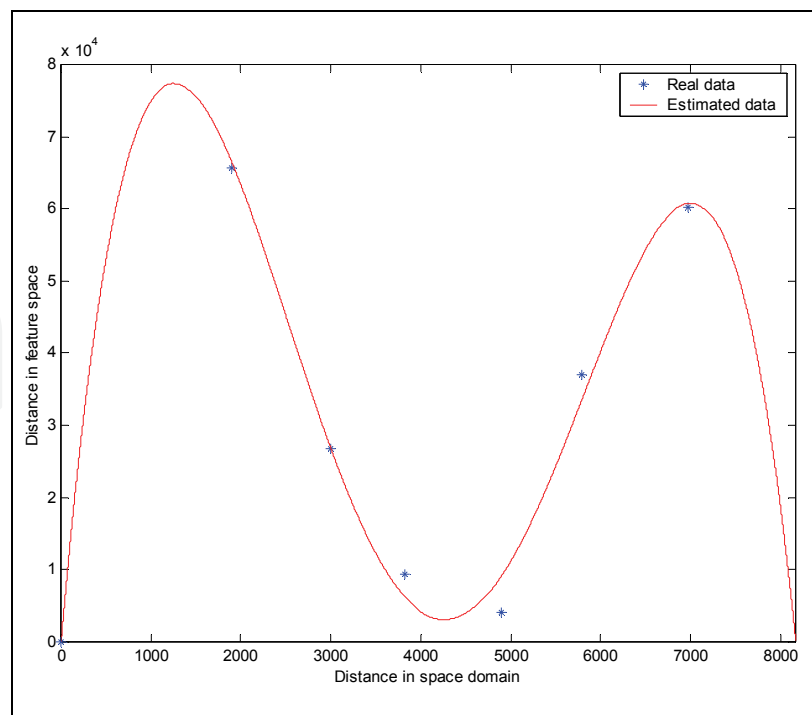


Fig. 8. Transform function  $f_T$ .

With this information we choose the solution that best fits with the crossing circles solution and the possible robot movement. In order to solve the ambiguity of the two intersection points, the angle that the robot has computed ( $\gamma$ ) is compared with the angles ( $\alpha_{FN1}$ ,  $\alpha_{FN2}$ ) analytically computed between the two intersection points and the sound source (corrected respect the magnetic north). The solution is the angle  $\alpha_{FNi}$  closer to  $\gamma$ . The uncertainty of this location is bounded by  $d \cdot \sin(\epsilon)$ , being  $\epsilon$  the difference between the actual angle of the robot respect the sound source and the  $\max\{\alpha_{FNi}, \gamma\}$ , and  $d$  is the actual distance of the robot to the sound source. In our experiments, we have verified that  $\epsilon$  is limited to  $1.9^\circ$  for  $d=1\text{m}$  and  $1^\circ$  for  $d=2.5\text{m}$  (between 3.3 and 4.3 cm of absolute error in localization).

#### 4.2 Method based on the LPV model

In the second proposed method based on the LPV model, the methodology applied to determine the robot's position is the following:

1. The robot acquires an audio signal in its current position and performs an identification process taking as input signal the filtered sound source signal and as output signal the acquired and filtered signal. The parameters corresponding to the obtained poles in this identification process will be the features components for further steps.
2. The Euclidean distances in the feature space are calculated between the current position and the different labeled samples.
3. The two first samples are chosen and the distance between them and the robot's position are then calculated. Through a transformation function  $f_T$ , in the same way that the previous approach, the distance in the feature domain is converted to a distance in the space domain. These two distances in the space domain give two possible positions by the crossing circles of distances.
4. To discriminate between both possible solutions, the angle between each one and the platform containing the microphone array (which contains a compass) are calculated,

and the closest one to the platform angle will be chosen as discriminatory variable to select the current robot's position.

5. Steps 3 and 4 are repeated with the remaining labeled samples, and the solution is chosen among the closest angle to the robot's platform.

The acoustic response of the environment is very directional, and this fact leads to consider some uncertainty in the determination of the transformation function which relates the distance in the feature space and the domain space.

The robot, in order to determine its location, will perform the identification process between the emitted sound signal by the sound sources and the acquired signal by the microphone. As it can be seen in Figure 9, the robot follows the trajectory indicated by the arrows. In the map sound sources are indicated (climate chamber and PCB insulator). Two experiments are carried out using both sound sources separately. There are two kind of audio samples:  $R_1, R_2, R_3, R_4, R_5, R_6$  and  $R_7$  which are used in the recognition step whereas  $M_1, M_2, M_3, M_4$  and  $M_5$  are labeled samples used in the learning step.

The acquired signal in the climatic chamber will be used in the identification process. This signal is time-continuous and, initially, non-stationary; but because the signal is generated by revolving electrical machines it has some degree of stationarity when a high number of samples is used, in this case, 50,000 samples (1.13 seconds).

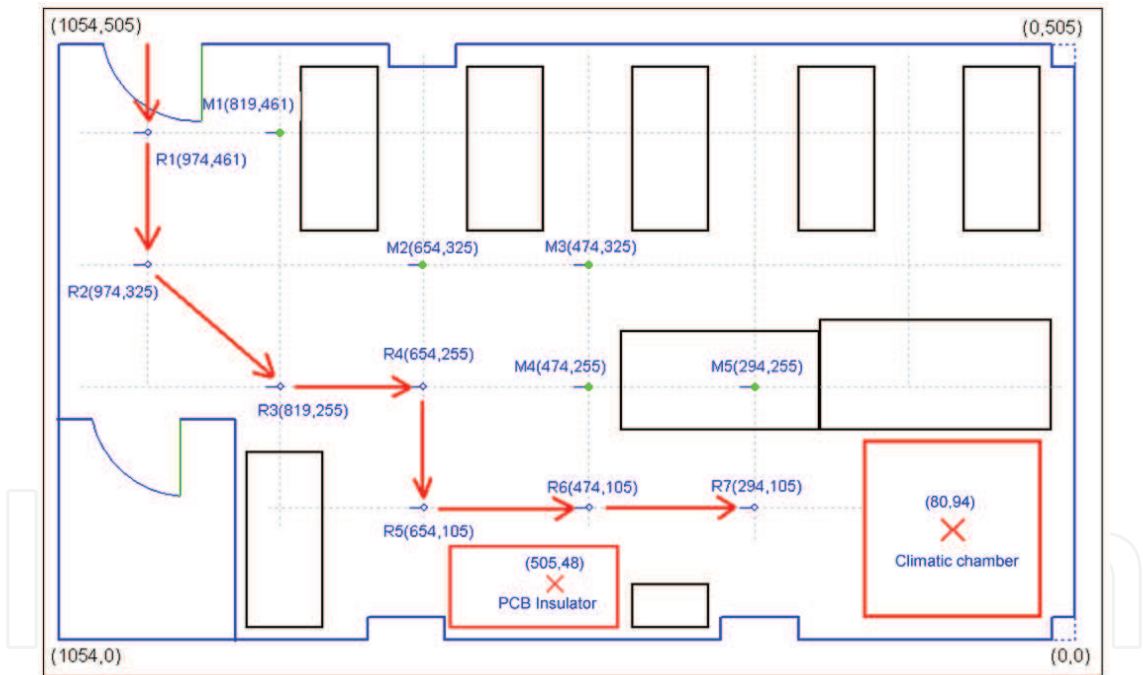


Fig. 9. Robot environment: labeled audio signals and actual robot trajectory with unlabeled signals ( $R_1, R_2, R_3, R_4, R_5, R_6$  and  $R_7$ ).

The fundamental frequency is located at 100Hz, see Figure 10, and there are also some significant harmonics above and below it. In order to simplify the identification process only the fundamental frequency at 100Hz will be taken into account.

In this approach the sampling frequency is 44,100Hz. Other lower frequencies could be used instead, avoiding working with a high number of samples, but this frequency has been chosen because in a near future a voice recognition system will be implemented aboard the robot and it will be shared with this audio localization system.



The emitted signal for the PCB insulator machine and its spectrum can be seen in Figure 11. To facilitate the plant identification process centering its response in the 100Hz component, the input and output signals will be filtered and, consequently, the input-output relationship in linear systems is an ARX model. To do that, a band-pass filter is applied to the acquired sound signals by the robot, specifically a 6th-order digital Cauer filter. Figure 12 shows the results of the filter for the input signal in, for instance, robot position  $R_4$  in the climatic chamber (experiment 1). After an initial step for selecting the model structure, an ARX has been selected, for the reasons explained above of stationery (Charbonnier et al., 1987), with  $n_a = 10$ ,  $n_b = 4$  and a delay of 2 for the case of the climatic chamber (experiment 1), and  $n_a = 10$ ,  $n_b = 2$  and a delay of 4 in the case of PCB insulator (experiment 2). When those 5 models are calibrated, they are validated with the error criteria of FPE (Function Prediction Error) and MSE (Mean Square Error), yielding values about  $10e(-10)$  and 3% respectively using 5000 data for identification and 3000 for validation. Besides, for the whole estimated models the residuals

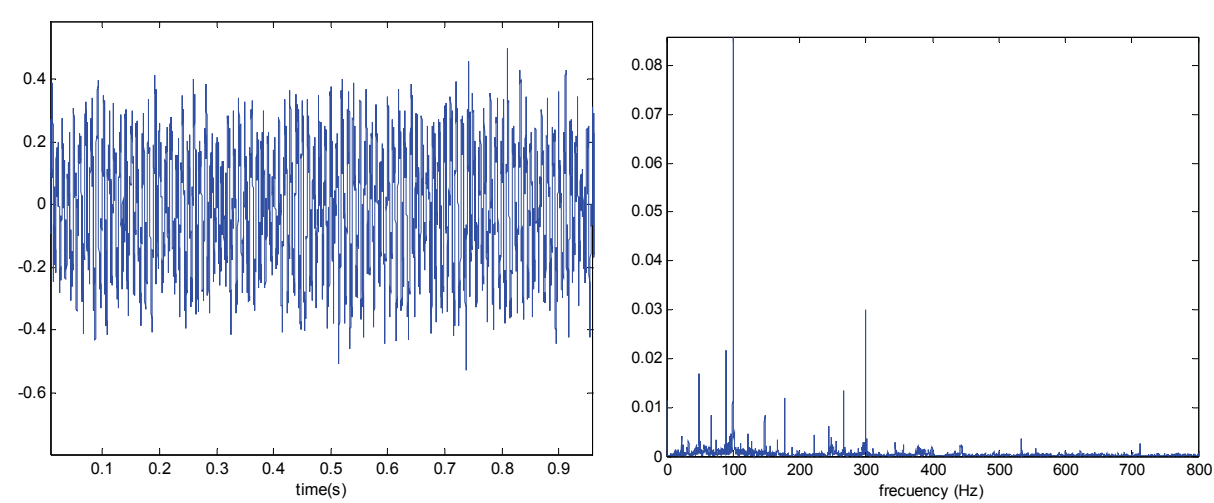


Fig. 10. Source signal (climate chamber) and its frequency spectrum.

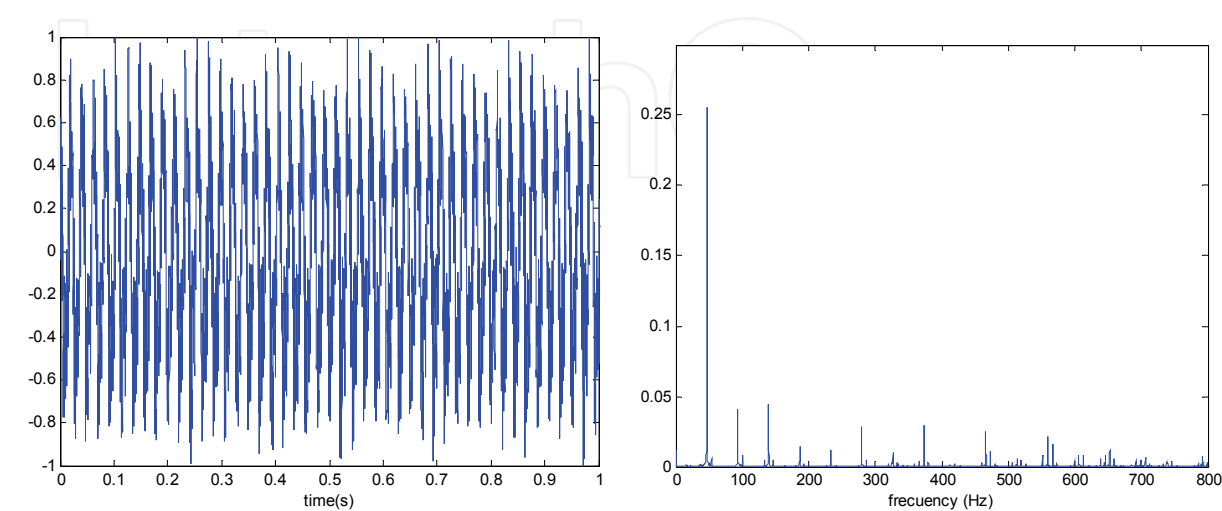


Fig. 11. Source signal (PCB insulator) and its frequency spectrum.

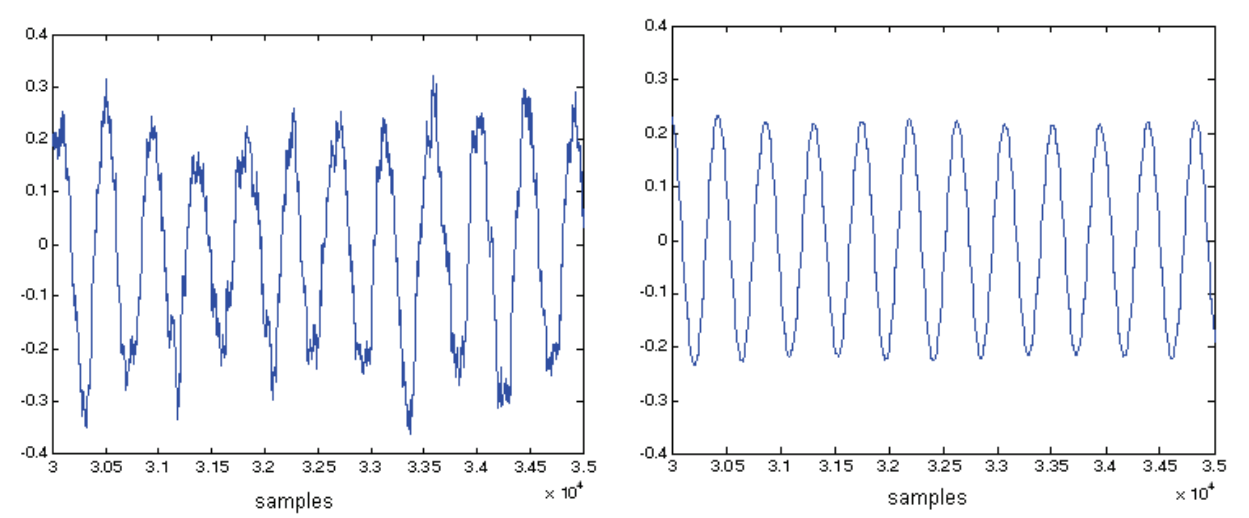


Fig. 12. R4 sound signal (left) and its filtered signal (right).

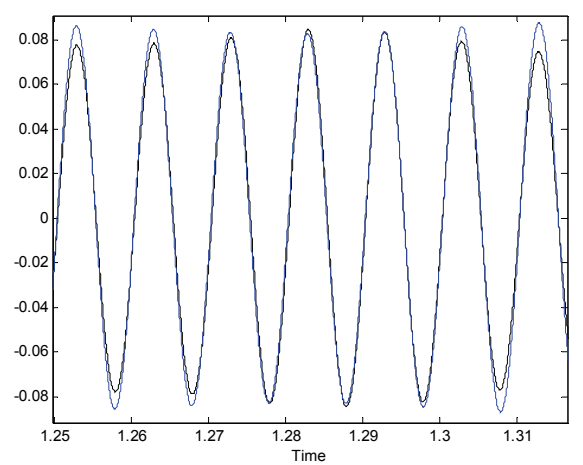


Fig. 13. Original  $M_5$  signal and its estimation.

autocorrelation and cross-correlation between the inputs and residuals are uncorrelated, indicating the goodness of the models.

For instance, for labeled  $M_5$  sample the signal and its estimation can be seen in Figure 13 in the first experiment, validating the model.

When observing the diagram of poles and zeros for the different transfer function models in the identification process for the labeled signals, there exists no difference between the zero positions, and, in the other hand, there is a significant variation in pole positions, due mainly to obstacles presence, reverberations among other effects, see Figure 14. Therefore, we will focus in poles to determinate the points in the feature space.

In experiment 1, in order to determine the transformation function, for every point in the feature space, the distances between them and the source signal are calculated, and these distances are plotted together with their corresponding distances in the space domain.

With these values, after an interpolation process, the transform function  $f_T$  is computed. In order to estimate the robot localization, we use other information such as the robot speed (in this case 15cm/sec), the computation time between each new position (3 sec). This fact is a source of uncertainty that adds in average  $\pm 45$  cm in the robot's position.

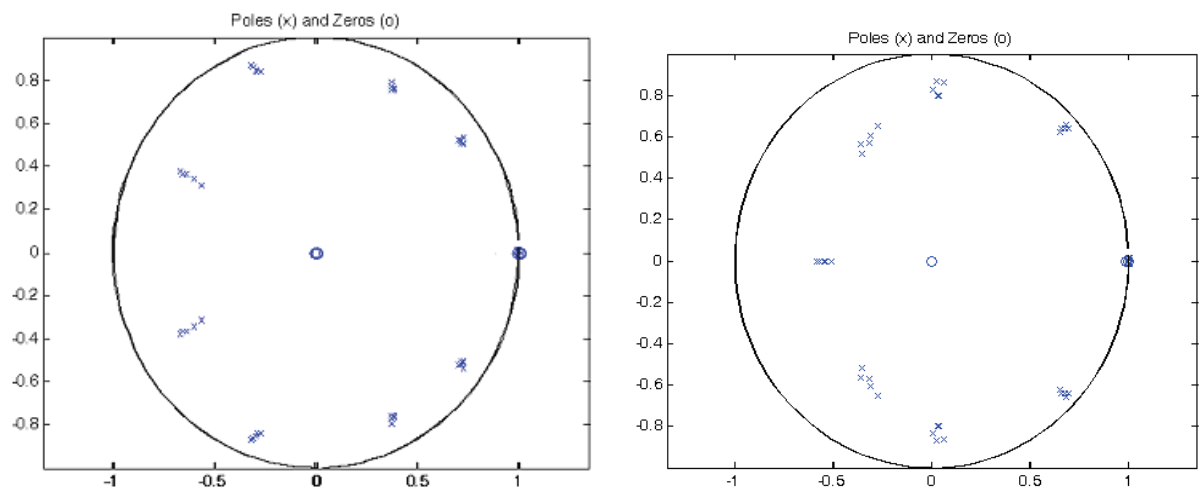


Fig. 14. Poles and zeros positions in experiment 1 (left) and 2 (right).

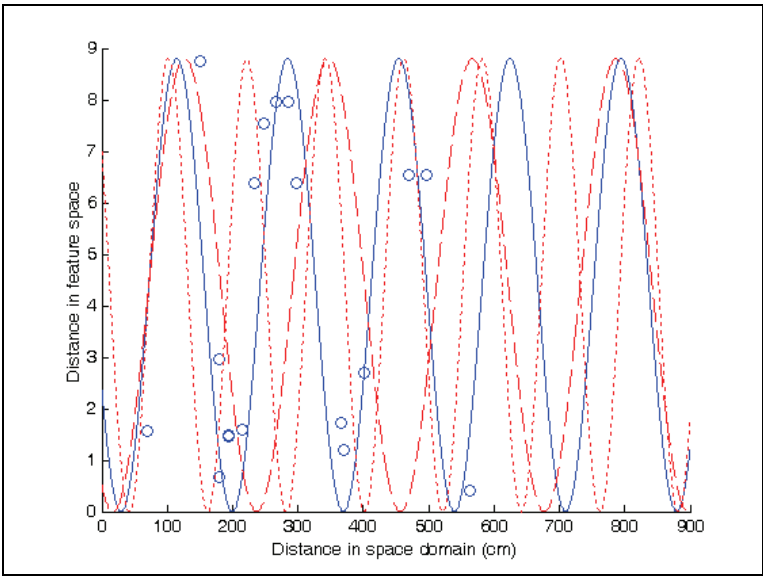


Fig. 15. Nominal transformation function and the limits of the interval for the uncertainty in experiment 1.

In experiment 1, when the climatic chamber is used as sound source the obtained transformation function is:

$$y = 4,4 + 4,4.\sin\left(\frac{2\pi x}{170} - \frac{80\pi}{170}\right)$$

Now, if an uncertainty interval is supposed ( $\pm 50$  cm) the transformation function that covers this variability in the robot’s position can be expressed (for both experiments) as:

$$y = A + A.\sin\left(\frac{2\pi x}{170 \pm 50} - \frac{\phi}{170 \pm 50}\right)$$

In Figure 15, the nominal transformation function and the limits for the uncertainty interval transformation functions can be seen.

There exists another uncertainty of about  $\pm 7.5$  degrees in the angle determination due to the rotary platform in the robot that contains the microphones. Finally, to determine the current robot's position the solution that provides the closest angle to the robot's platform will be chosen. The results of our experiments yield an average error in the X axis of -1.242% and in the Y axis of 0.454% in experiment 1 and 0.335% in the X axis and -0.18% in the Y axis, providing estimated x-y positions good enough and robust.

## 5. Conclusion

With the approaches presented in this Chapter we have achieved some interesting results that encourage the authors to keep on walking in this research field. The room feature extraction is carried out by identification of the sound signals. Besides to reinforce the localization, avoiding ambiguity and reducing uncertainty and incorporating robustness, a sensorial system is used aboard the robot to compute the angle between itself and the sound source. The obtained feature space is related with the space domain through a general approach with acoustical meaning. The validation of this novel approach is tested in different environments obtaining good results. The results keep on being very good when the uncertainty is incorporated in the transformation function.

## 6. References

- Navarro, D. & Benet, G. (2009). Magnetic Map Building for Mobile Robot Localization Purpose, *14th International Conference on Emerging Technologies and Factory Automation*, Palma de Mallorca, September, 2009.
- Caruso, M. (2000). Applications of magnetic sensors for low cost compass systems, *IEEE Position Location and Navigation Symposium*, pp. 177-184, San Diego, CA, USA, March 2000.
- Kim, H.-D.; Kim, D.-W. & Sim, K.B. (2006). Simultaneous Localization and Map Building using Vision Camera and Electrical Compass, *SICE-ICASE International Joint Conference*, Korea, October, 2006.
- Kim, H.-D.; Seo, S.-W.; Jang, I.-H. & Sim, K.B. (2007). SLAM of Mobile Robot in the indoor Environment with Digital Magnetic Compass and Ultrasonic Sensors, *International Conference on Control, Automation and Systems*, Oct. 17-20, 2007, Seoul, Korea.
- Luo, R.C.; Yih, C.-C. & Su, K.L. (2002). Multisensor Fusion and Integration: Approaches, Applications and Future Research Directions, *IEEE Sensors Journal*, Vol. 2, no. 2, April 2002.
- Begum, M.; Mann, G.K.I. & Gosine, R. (2006). A Fuzzy-Evolutionary Algorithm for Simultaneous Localization and Mapping of Mobile Robots, *IEEE Congress on Evolutionary Computation*, Canada, July, 2006.
- Brunskill, M. & Roy, N. (2005). SLAM using Incremental Probabilistic PCA and Dimensionality Reduction, *Proc. of the IEEE International Conference on Robotics and Automation*, Spain, April, 2005.
- Di Marco, M.; Garulli, A.; Lacroix, S. & Vicino, A. (2000). A Set Theoretic Approach to the Simultaneous Localization and Map Building Problem, *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia, December, 2000.

- Smith, R.; Self, M. & Cheeseman, P. (1990). Estimating uncertain spatial relationships in robotics, *Autonomous Robot Vehicles*, vol. 8, pp. 167-193, 1990.
- Thrun, S. (2001). A probabilistic online mapping algorithm for teams of mobile robots, *Journal of Robotics Research*, vol. 20, no. 5, pp. 335-363, 2001.
- Begum, M.; Mann, G.K.I. & Gosine, R.G. (2006). An Evolutionary SLAM Algorithm for Mobile Robots, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 9 - 15, 2006, Beijing, China.
- Priyantha, N.B. (2000). The Cricket location-support system, *Proc. of the 6th Annual International Conference on Mobile Computing and Networking*, pp. 32-43, 2000.
- Sayed, A.H.; Tarighat, A. & Khajehnouri, N. (2005). Network-Based Wireless Location: Challenges faced in developing techniques for accurate wireless location information, *IEEE Signal Processing Magazine*, vol. 22, no. 4, July 2005.
- Christo, C.; Carvalho, E.; Silva, M.P. & Cardeira, C. (2009). Autonomous Mobile Robots Localization with Multiples iGPS Web Services, *14th International Conference on Emerging Technologies and Factory Automation*, Palma de Mallorca, September, 2009.
- Yang, P.; Sun, H. & Zu, L. (2007). An Acoustic Localization System Using Microphone Array for Mobile Robot, *International Journal of Intelligent Engineering & Systems*, 2007.
- Mumolo, E.; Nolich, M. & Vercelli, G. (2003). Algorithms for acoustic localization based on microphone array in service robotics, *Robotics and Autonomous Systems*, vol. 42, pp. 69-88, 2003.
- Csyzewski, A. (2003). Automatic identification of sound source position employing neural networks and rough sets, *Pattern Recognition Letters*, vol. 24, pp. 921-933, 2003.
- Ying, J. & Runze, Y. (2007). Research Status and Prospect of the Acoustic Localization Techniques, *Audio Engineering*, vol. 31, no. 2, pp. 4-8, 2007.
- Sasaki, Y.; Kagami S. & Mizoguchi, H. (2006). Multiple sound source mapping for a mobile robot by selfmotion triangulation, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, 2006.
- Kim, Y.-E.; Su, D.-H.; Chung, G.-J.; Huang, X. & Lee, C.-D. (2009). Efficient Sound Source Localization Method Using Region Selection, *IEEE International Symposium on Industrial Electronics*, ISIE 2009, Seoul Olympic Parktel, Seoul, Korea July 5-8, 2009.
- Brandstein, M.S. & Silverman, H. (1997). A practical methodology for speech source localization with microphone arrays, *Computer Speech and Language*, vol. 11, no. 2, pp. 91-126, 1997.
- Knapp, C.H. & Carter, G.C. (1976). The generalized correlation method for estimation of time delay, *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. Assp-24, no. 4, 1976.
- Nakashima, H. & Mukai, T. (2005). 3D Sound source localization system based on learning of binaural hearing, *IEEE International Conference on Systems, Man and Cybernetics*, vol. 4, pp. 3534-3539, 2005.
- Kim, H.S. & Choi, J. (2009). Binaural Sound Localization based on Sparse Coding and SOM, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 11-15, 2009, St. Louis, USA.
- Yi, H. & Chu-na, W. (2010). A New Moving Sound Source Localization Method Based on the Time Difference of Arrival, *Proc. of the International Conference on Image Analysis and Signal Processing*, pp. 118-122, 9-11 April, 2010, Zhejiang, China.



- Rodemann, T.; Joublin, F. & Goerick, C. (2009). Audio Proto Objects for Improved Sound Localization, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 11-15, 2009, St. Louis, USA
- Mori, K.; Kasashima, N.; Yoshioha, T. & Ueno, Y. (1996). Prediction of Spalling on a Ball Bearing by Applying the Discrete Wavelet Transform to Vibration Signals, *Wear*, vol. 195, no. 1-2, pp. 162-168, 1996.
- Roberts, S. & Everson, R. (2001). Independent Component Analysis: Principles and Practice, *Cambridge Univ. Press*, Cambridge, UK, 2001.
- Bolea, Y.; Grau, A. & Sanfeliu, A. (2003). Non-speech Sound Feature Extraction based on Model Identification for Robot Navigation, *8th Iberoamerican Congress on Pattern Recognition*, CIARP 2003, Lectures Notes in Computer Science, LNCS 2905, pp. 221-228, Havana, Cuba, November 2003.
- Mallat, S. & Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries, *IEEE Trans. on Signal Processing*, vol.45, no.12, pp. 3397-3415, 1993.
- Donoho, D.-L. (1999). De-noising by soft-thresholding, *IEEE Trans. on Information Theory*, vol. 33, no. 7, pp. 2183-2191, 1999.
- Lin, J. (2001). Feature Extraction of Machine Sound using Wavelet and its Application in Fault Diagnosis, *NTD&E International*, vol. 34, pp. 25-30, 2001.
- Grau, A.; Bolea, Y. & Manzanares, M. (2007). Robust Industrial Machine Sounds Identification based on Frequency Spectrum Analysis, *12th Iberoamerican Congress on Pattern Recognition*, CIARP 2007, Lecture notes in Computer Science, LNCS 4756, pp. 71-77, Valparaiso, Chile, November 2007.
- Ljung, L. (1987). System identification: Theory for the user. *Prentice-Hall*, 1987.
- Charbonnier, R.; Barlaud, M.; Alengrin, G. & Menez, J. (1987). Results on AR-modeling of nonstationary signals, *IEEE Trans. Signal Processing*, vol. 12, no. 2, pp. 143-151.
- Kayhan, A.S.; Ei-Jaroudi, A. & Chaparro, L.F. (1994). Evolutionary periodogram for nonstationary signals, *IEEE Trans. Signal Process*, vol. 42, no. 6, pp. 1527-1536.
- Tsatsanis, M.K. & Giannakis, G.B. (1993). Time-varying system identification and model validation using wavelets, *IEEE Trans. Signal Process*, vol. 41, no. 12, pp. 3512-3523.
- Kinsler, L.; Frey, A.; Coppens, A. & Sanders, J. (1995). Fundamentals of Acoustics, *Limusa Ed.*, Barcelona, 1995.
- Kuttruff, H. (1979). Room Acoustics, *Applied Science Publishers Ltd.*, 1979.
- Haneda, Y.; Makino, S. & Kaneda, Y. (1992). Modeling of a Room Transfer Function Using Common Acoustical Poles, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, ICASSP-92, vol.2, pp. 213-216, 1992.
- Haneda, Y.; Kaneda, Y. & Kitawaki, N. (1999). Common-Acoustical-Pole and Residue Model and Its Application to Spatial Interpolation and Extrapolation of a Room Transfer Function, *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, Nov. 1999.
- Gustaffson, T.; Pota, H.R.; Vance, J.; Rao, B.D. & Trivedi, M.M. (2000). Estimation of Acoustical Room Transfer Functions, *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia, December 2000.
- Bolea, Y.; Manzanares, M. & Grau, A. (2008). Robust robot localization using non-speech sound in industrial environments, *IEEE International Symposium on Industrial Electronics*, ISIE 2008, Cambridge, United Kingdom, 30 June- 2 July 2008.

- Manzanares, M.; Guerra, E.; Bolea, Y. & Grau, A. (2009). Robot Localization Method by Acoustical Signal Identification, 14th *IEEE Emerging Tech and Factory Automation*, ETFA'09, Palma de Mallorca, Spain, 2009.
- Bielińska, E. (2002). Speaker identification, *Artificial Intelligence Methods*, AI-METH, 2002.

IntechOpen

IntechOpen



## **Advances in Sound Localization**

Edited by Dr. Pawel Strumillo

ISBN 978-953-307-224-1

Hard cover, 590 pages

**Publisher** InTech

**Published online** 11, April, 2011

**Published in print edition** April, 2011

Sound source localization is an important research field that has attracted researchers' efforts from many technical and biomedical sciences. Sound source localization (SSL) is defined as the determination of the direction from a receiver, but also includes the distance from it. Because of the wave nature of sound propagation, phenomena such as refraction, diffraction, diffusion, reflection, reverberation and interference occur. The wide spectrum of sound frequencies that range from infrasounds through acoustic sounds to ultrasounds, also introduces difficulties, as different spectrum components have different penetration properties through the medium. Consequently, SSL is a complex computation problem and development of robust sound localization techniques calls for different approaches, including multisensor schemes, null-steering beamforming and time-difference arrival techniques. The book offers a rich source of valuable material on advances on SSL techniques and their applications that should appeal to researches representing diverse engineering and scientific disciplines.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Manuel Manzanares, Yolanda Bolea and Antoni Grau (2011). Robust Audio Localization for Mobile Robots in Industrial Environments, *Advances in Sound Localization*, Dr. Pawel Strumillo (Ed.), ISBN: 978-953-307-224-1, InTech, Available from: <http://www.intechopen.com/books/advances-in-sound-localization/robust-audio-localization-for-mobile-robots-in-industrial-environments>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen