Robust Maximum Likelihood Acoustic Energy Based Source Localization in Correlated Noisy Sensing Environments

E. Dranka and R. Coelho, *Member, IEEE*

***Abstract*—Acoustic energy based localization with wireless sensor networks is an interesting solution to locate sources and targets. For simplicity, localization formulation based on the maximum likelihood (ML) approach considers that the source and noise samples are uncorrelated and represented by a Gaussian distribution. However, the acoustic background noise can severely affect the accuracy of the location estimation. This paper proposes an accurate error estimate in which the correlation of the received signals at each wireless sensor is represented by a Hurst exponent and modeled by a fractional Gaussian noise (fGn). The experimental results show that the proposed solution is more appropriate for the source localization estimation under real acoustic noises and even for highly non-stationary sources.**

***Index Terms*—Acoustic source localization, maximum likelihood (ML), energy based localization, Hurst exponent, fractional Gaussian noise.**

I. Introduction

The deployment of efficient and low cost wireless sensor networks [1]–[3] has motivated the proposal of source localization solutions based on acoustic energy sensing [1], [4]. The accurate estimation of an acoustic source position is a very important issue in many research areas and applications such as target tracking, surveillance, video conferences [5], seismic [6] and robotics [7].

The acoustic source localization methods are mainly based on the computation of the time-delay estimation (TDE) or the time-delay of arrival (TDOA) and the acoustic signal energy. The TDE or TDOA algorithms use the time-delay or phase difference measures obtained at the acoustic sensors generally distributed in a microphone array. Source localization estimation methods using the acoustic energy or intensity were proposed for wireless sensors and enable direct source location. The maximum likelihood (ML-Energy) version [4] gives the location estimation of multiple sources even in open-field wireless sensor environment. The authors showed that the ML-Energy outperforms the previous acoustic estimation algorithms.

The major challenge of the source localization estimation area is to achieve accurate measures at each sensor in the presence of real acoustic background noise. For simplicity, in the literature, the location estimation models consider that the noise samples are uncorrelated and represented by a Gaussian distribution. However, this assumption can severely degrade the sensor measurements and thus, the source localization estimation accuracy [8]. Moreover, the acoustic noises and sources can be non-stationary and have different time and frequency statistics [9]–[11].

This paper introduces a novel ML acoustic energy based source localization definition to achieve the estimation accuracy under correlated acoustic noise distortion. In the proposed solution, the correlation degree of the corrupted signals received at each acoustic sensor is represented by the Hurst exponent (H) [12]. These samples are represented by a fractional Gaussian noise (fGn) [13]. Furthermore, since H defines any degree of correlation, the proposed method (H-ML-Energy) provides a more robust localization estimation under a wide range of acoustic scenarios. The H-ML-Energy is investigated considering three real acoustic sources (Car, Helicopter and Speech) and three noises (Babble, Car and F16) collected from different databases. The evaluation experiments are conducted with the signals corrupted by the real acoustic noises and five different values of SNR (signal-to-noise ratio). The experiments also include the computation of the index of non-stationarity (INS) [14] of the acoustic sources and noises. The ML-Energy [4] is adopted as the baseline method for the source localization investigation considering correlated and uncorrelated noisy scenarios. The source localization estimation accuracy is examined in terms of the error probability function (EPF), the Bhattacharryya distance [15], [16] and the root mean square error (RMSE) estimates. The results show that the H-ML-Energy outperforms the baseline ML-Energy for all the acoustic sources in correlated noisy situations. For the highly non-stationary Speech source, the proposed solution achieves lower RMSE results than those obtained with ML-Energy, even when it is corrupted by the non-stationary acoustic noises (e.g., Babble and F16).

This paper is organized as follows. Section II introduces the main concepts of the Hurst exponent, its estimation method and the index of non-stationarity. It also presents the time and spectral characteristics of the acoustic sources and noises

Manuscript received March 14, 2014; revised August 05, 2014, October 23, 2014, December 10, 2014; accepted December 13, 2014. Date of publication December 23, 2014; date of current version February 11, 2015. This work was supported in part by the National Council for Scientific and Technological Development (CNPq) under 304254/2012-6 research grant. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Janusz Konrad.

The authors are with the Laboratory of Acoustic Signal Processing, Military Institute of Engineering (IME), Rio de Janeiro, RJ, 22290-270, Brazil (e-mail: coelho@ime.eb.br).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2014.2385657

examined in this paper. In Section III, the proposed solution, H-ML-Energy, is introduced. Extensive localization experiments and results considering the proposed solution and the baseline method, are provided and discussed in Section IV. Finally, Section V concludes this work.

II. Acoustic Source and Noise Representation: Temporal and Spectral Characteristics

Real acoustic signals and noises are generated by many kinds of sources like animals, vehicles, weapons and people. Consequently, they have different temporal characteristics (amplitude distribution), stationarity, time-scale or degree of correlation and spectral aspects. This Section briefly introduces the Hurst exponent and the estimation method applied in this work. It also shows the spectrogram of the sources and noises and discusses its correspondence with the Hurst values. The index of non-stationarity [14] is also presented and evaluated in this Section.

A. Hurst Exponent

The Hurst exponent $(0 < H < 1)$ expresses the time-scaling degree of a stochastic process. It can also be defined by the decaying rate of the auto-correlation coefficient function $\rho(k)$ $(-1 < \rho(k) < 1)$ as $k \to \infty$. Let a signal be represented by a stochastic process $x(t)$, with finite variance and normalized auto-correlation function (ACF)

$$\rho(k) = \frac{\mathrm{Cov}[x(t), x(t+k)]}{\mathrm{Var}[x(t)]}, \quad k = 0, 1, 2, \dots \tag{1}$$

where Cov[$\cdot$] and Var[$\cdot$] refer to the covariance and variance, respectively, $\rho(k)$ belongs to $[-1, 1]$ and $\lim_{k\to\infty} \rho(k) = 0$. The asymptotic behavior of $\rho(k)$ is given by

$$\rho(k) \sim H(2H-1)k^{2(H-2)} \tag{2}$$

This means that $\rho(k)$ is a slowly decaying function and that when $k \to \infty$, $\rho(k) \sim H(2H-1)k^{2(H-2)}$ and hence, $\rho(k)/H(2H-1)k^{2(H-2)} \sim 1$. According to the value of H, stochastic processes can be classified as:

- Anti-persistent processes or negative correlation degree $(0 < H < \frac{1}{2})$: The ACF rapidly tends to zero and $\sum_{k=-\infty}^{\infty} \rho(k) = 0$.
- Processes with short-range time scale $(H = \frac{1}{2})$: The ACF $\rho(k)$ exhibits an exponential decay to zero, such that $\sum_{k=-\infty}^{\infty} \rho(k) = c$, where $c > 0$ is a finite constant, e.g., uncorrelated Gaussian noise.
- Processes with long-range time scale or strong correlation $(\frac{1}{2} < H < 1)$: The ACF $\rho(k)$ is a slowly-vanishing function, meaning a time dependence degree even between samples that are far apart or $\sum_{k=-\infty}^{\infty} \rho(k) = \infty$.

Therefore, the H exponent of a signal is related to its spectral characteristics. Within the whole range $]0, 1[$, the power spectral density $S_x(f)$ can be shown to be proportional to f^{1-2H} when $f \to 0$ [13]. For $H = 1/2$, $S_x(f)$ is constant over the whole frequency spectrum (e.g., white noise), whereas low frequencies are prominent in the case where $H > 1/2$, and in particular when $H \to 1$ ($1/f$ or pink noise).

TABLE I
Hurst Exponent Estimation of the Acoustic Sources Corrupted With Real Noises

		Noise		
Source	SNR	Car (H=0.84)	F16 (H=0.67)	Babble (H=0.58)
Car (H=0.84)	0 dB	0.87	0.81	0.80
	5 dB	0.87	0.84	0.85
	10 dB	0.87	0.84	0.85
	15 dB	0.86	0.83	0.83
	20 dB	0.85	0.82	0.80
Helicopter (H=0.12)	0 dB	0.83	0.63	0.55
	5 dB	0.80	0.60	0.52
	10 dB	0.76	0.56	0.49
	15 dB	0.71	0.51	0.45
	20 dB	0.65	0.46	0.40
Speech (H=0.10)	0 dB	0.82	0.62	0.54
	5 dB	0.78	0.57	0.49
	10 dB	0.72	0.50	0.43
	15 dB	0.63	0.42	0.35
	20 dB	0.53	0.33	0.28

In this work, the Hurst exponent estimation is adopted to examine the correlation degree of the noisy signals. The wavelet-based method [17]–[19] is applied for the estimation of the Hurst exponent. It can be described in three main steps as follows:

1) Wavelet decomposition: the discrete wavelet transform (DWT) is applied to successively decompose the input sequence of samples into approximation $(a(j,n))$ and detail $(d(j,n))$ coefficients, where j is the decomposition scale $(j = 1, 2, \dots, J)$ and n is the coefficient index of each scale.
2) Variance estimation: for each scale j, the variance $\sigma_j^2 = (1/N_j)\sum_n d(j,n)^2$ is evaluated from the detail coefficients, where N_j is the number of available coefficients for each scale j. In [18], it is shown that $E[\sigma_j^2] = \mathcal{C}_H\, j^{2H-1}$, where $\mathcal{C}_H$ is a constant.
3) Hurst computation: a weighted linear regression is used to obtain the slope θ of the plot of $y_j = \log_2(\sigma_j^2)$ versus j. The Hurst exponent is estimated as $H = (1+\theta)/2$.

Table I shows the H values obtained from each acoustic source (Car, Helicopter and Speech), noise (Babble, Car, F16) and also from the corrupted signals considering five SNR values: 0 dB, 5 dB, 10 dB, 15 dB and 20 dB. It can be noted that the sources and noises achieved a wide range of H values $(0.10 \le H \le 0.84)$. It means that all types of correlations are represented, i.e., strong $(H > 1/2)$ and anti-persistent or negative $(H < 1/2)$. It can also be observed that the correlation degree of a source significantly varies depending on the background noise. This is evident with the Speech source corrupted with F16. When the Helicopter source $(H = 0.12)$, which has an anti-persistent correlation, is corrupted with a strong correlated noise as the Car $(H = 0.84)$, the resulting noisy signal shows a long-range correlation $(H = 0.62)$ at SNR $= 20$ dB. Additionally, the corresponding spectral characteristic of the acoustic sources and noises can be examined in Fig. 1.

B. Index of Non-Stationarity

In its definition, a signal is considered as *stationary relatively to an observation scale* if its local short-time spectra at all different time instants, are statistically similar to its global spectrum. The index of non-stationarity (INS) [14]is a time-frequency approach to objectively examine the non-stationarity of a signal. The stationarity test is conducted by comparing the

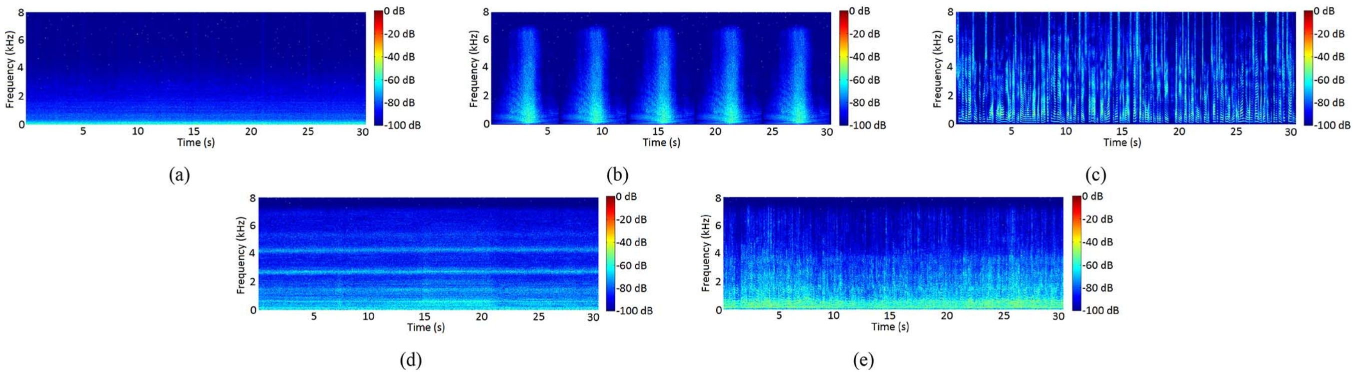


Fig. 1. Spectrogram of signals and noises. (a) Car, (b) Helicopter, (c) Speech, (d) F16, (e) Babble.

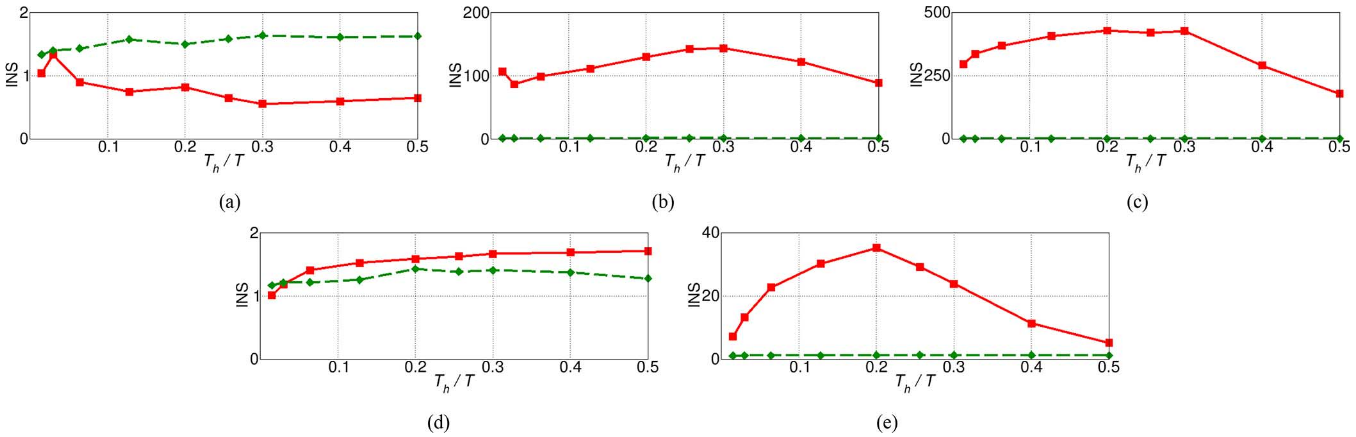


Fig. 2. INS values of source signals and noises. The red line represents the INS value of each time scale T_h/T. The threshold is indicated by the green dashed line. (a) Car, (b) Helicopter, (c) Speech, (d) F16, (e) Babble.

spectral components of the signal to a set of stationary references, called *surrogates*. For this purpose, the spectrograms of the signal and *surrogates* are obtained by means of the short-time Fourier transform (STFT) considering a window length T_h. Then, the Kullback-Leibler (KL) divergence is used to measure the distance between the short-time spectra of the analyzed signal and its global spectrum averaged over time. Finally, the INS is given by the ratio between this distance and the corresponding KL values obtained from the stationary *surrogates*. In [14], the authors considered that the distribution of the KL values can be approximated by a Gamma distribution [14]. Therefore, for each window length T_h, a threshold γ can be defined for the stationarity test considering a confidence degree of 95%. Thus,

$$\text{INS}\begin{cases} \leq \gamma & \text{, signal is stationary;} \\ > \gamma & \text{, signal is non-stationary.} \end{cases} \tag{3}$$

In this work, the index of non-stationarity is evaluated for the investigated acoustic sources and noises. If the signal or the noise is stationary, its INS value is expected to be close to unity. On the other hand, the larger the INS the more non-stationary the noise. The time scale T_h/T indicates the relation between the length adopted in the short-time spectral analysis (T_h) and the total length ($T = 30$ s) of the signal.

The stationarity of three acoustic sources (Car, Helicopter and Speech) and three noises (Babble, Car and F16) was examined in this work. The female speech source was chosen from TIMIT [20] database. The F16 and Babble noises and Car source/noise were selected from NOISEX-92 [21]database. The Helicopter source was chosen from FreeSfx[1] database. The INS values are presented in Fig. 2. The minimum time scale T_h/T value chosen for the estimation is 0.016.

In this paper, the signals and noises that have INS values greater than 40 are considered highly non-stationarity. It can be seen that the Speech and Helicopter sources are highly non-stationary. If its INS value is above the threshold for the majority of T_h/T, the noise or source is classified as non-stationary. This is the case of the Babble and F16 noises. The Car noise is considered as stationary since its INS values are below the threshold for every time scale.

III. Proposed H-ML-Energy

The main objective or aim of the proposed approach is to improve the accuracy of the maximum likelihood energy-based acoustic source localization methods, when the sources are corrupted with real acoustic noises. The ML-Energy was first proposed in [1] and it was extended for multiples sources in [4]. Its main principle is based on the fact that the acoustic energy is attenuated as the signal propagates from the source to the sensors [22].

Consider an open-field without obstacles (no reverberation) and with a constant and uniform sound velocity where K sources are positioned in this field, and L energy sensors are deployed with known positions given by $\mathbf{r}_i$ $(i = 1, 2, \ldots, L)$. The signal intensity attenuates in a rate inversely proportional

[1] Available in http://www.freesfx.co.uk/.

to the distance that it is propagated [22]. The signal received at the i-th sensor is sampled during the n-th time interval with a sampling frequency f_s. It is defined as

$$x_i(n) = A_i(n) + w_i(n), \tag{4}$$

where

$$A_i(n) = \sqrt{g_i}\sum_{j=1}^{K} \frac{s_j(n-\tau_{ji})}{|\mathbf{p}_j(n-\tau_{ji}) - \mathbf{r}_i|} \tag{5}$$

is the acoustic signal intensity measured in the i-th sensor. The vector $\mathbf{p}_j$ denotes the j-th $(j = 1, 2, \ldots, K)$ source spatial coordinates and $w_i(n)$ is the background noise. The j-th source signal intensity is represented by $s_j(n - \tau_{ji})$, with τ_{ji} being the propagation delay from the j-th source to the i-th sensor i. $\sqrt{g_i}$ indicates the i-th sensor gain.

Given a time index t, the acoustic energy $u_i(t)$ is defined [1] as $\mathbb{E}[x_i^2(n)] = u_i(t)$, or

$$u_i(t) = g_i \cdot \sum_{j=1}^{K} \frac{\mathbb{E}[s_j^2(n-\tau_{ji})]}{|\mathbf{p}_j(t-\tau_{ji}) - \mathbf{r}_i|^2} + \mathbb{E}[w_i^2(t)] + 2\mathbb{E}[A_i(t)w_i(t)]. \tag{6}$$

Since the energy measurements are averaged over a short-time block of M samples, and considering that the signal energy does not vary significantly during the block duration, the propagation delays are neglected for the model. Therefore, the acoustic energy received in each sensor is

$$u_i(t) = g_i \sum_{j=1}^{K} \frac{B_j(t)}{d_{ij}^2(t)} + 2\mathbb{E}[A_i(t)w_i(t)] + \mathbb{E}[w_i^2(t)] \tag{7}$$

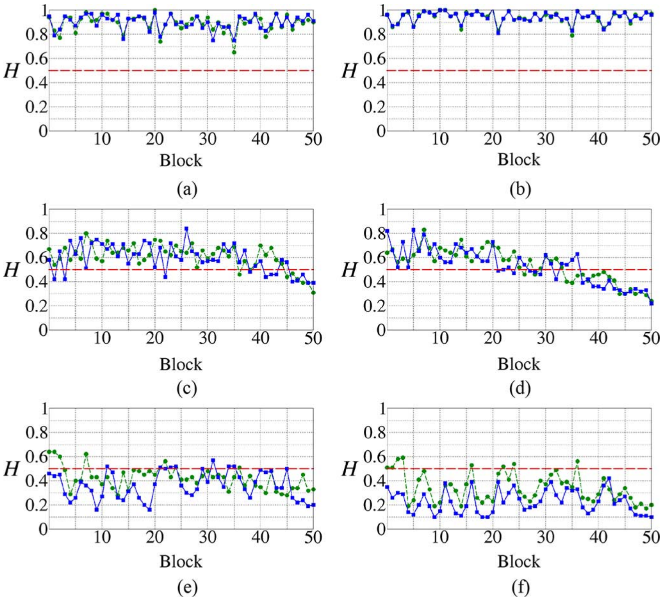
where $B_j(t) = (1/M)\sum_{n=1}^{M}\left[x_j(n) - \mu_{x_j}\right]^2$ denotes the j-th source acoustic energy, and $d_{ij}(t) = |\mathbf{p}_j(t) - \mathbf{r}_i|$ refers to the Euclidean distance between the i-th sensor and the j-th source. The ML-Energy method considers that the cross term $\mathbb{E}[A_i(t)w_i(t)]$ is equal to zero, since the background noise is modeled as uncorrelated with the source signal. In this proposal, the cross term $(\mathbb{E}[A_i(t)w_i(t)])$ and the $\mathbb{E}[w_i^2(t)]$ term are modeled by a fractional Gaussian noise with H exponent, mean μ_H and variance σ_H, obtained from the sensor readings, i.e., from the received noisy signal. Denoting the fGn process by $h(t) = 2\mathbb{E}[A_i(t)w_i(t)] + \mathbb{E}[w_i^2(t)]$, the acoustic source localization model is given by,

$$u_i(t) = g_i \sum_{j=1}^{K} \frac{B_j(t)}{d_{ij}^2(t)} + h_i(t). \tag{8}$$

Thus, the fGn process represents the energy measurement error and, since the fGn is able to represent any degree of correlation by the means of its H exponent, the proposed solution grants better accuracy to the energy based localization model.

The fractional Gaussian noise (fGn) [13] is a series of N identical Gaussian random variables $X_1, X_2, \ldots, X_N$, with correlation degree represented by H, and with the property

$$\mathcal{A}_N = \frac{X_1 + X_2 + X_3 + \ldots + X_N}{N^H} \overset{d}{\approx} X, \tag{9}$$



(a) (b) (c) (d) (e) (f)

Fig. 3. Hurst exponent estimated from the sources corrupted with real noise (green) and from fGn process proposed for the representation of the correlation degree (blue). The dashed red lines indicate the value $H = 1/2$. (a) Car + F16 (0 dB), (b) Car + F16 (10 dB), (c) Helicopter + F16 (0 dB), (d) Helicopter + F16 (10 dB), (e) Speech + Babble (0 dB), (f) Speech + Babble (10 dB).

where $\overset{d}{\approx}$ denotes similarity in the probability distribution. Alternatively, it can be given in terms of its sample variance, i.e., $\mathrm{Var}(X_N) = N^{H-1}\mathrm{Var}(X_1)$ where

$$\mathrm{Var}(X_N) = \frac{\mathrm{Var}(X_1 + X_2 + X_3 + \ldots + X_N)}{N\mathbb{E}[X_1]}. \tag{10}$$

The fGn autocorrelation function can be defined by

$$\mathbb{E}[X_0 X_k] = R_H(k) = \frac{1}{2}[(k-1)^{2H} - 2k^{2H} + (k+1)^{2H}], \tag{11}$$

where k is the lag and H is the Hurst exponent. According to [13], the fGn spectral density can be approximated, for all H values, by $S_H(f) \sim C\sigma^2|f|^{1-2H}$, when $|f| \to 0$. It follows that $H = 1/2$ corresponds to the particular case of the ML-Energy, i.e.,

$$\mathbb{E}[X_0 X_k] = \delta(k) = \begin{cases} 0, & k \neq 0 \\ 1, & k = 0 \end{cases}. \tag{12}$$

Fig. 3 depicts the Hurst exponent results estimated from blocks of the source signals corrupted by the real acoustic noises at 0 dB and 10 dB. The H values of the signals are plotted using a dashed green line whereas the solid blue line represents the H values of the sources corrupted by fGn with H estimated from the real acoustic noises. The uncorrelated case is illustrated by the dashed red line over $H = 1/2$. It can be seen that the three sources are highly different from the uncorrelated Gaussian assumption in almost all the blocks, i.e., their H value is not equal to 1/2. This demonstrates that the fGn is a good candidate to represent the correlation degree of the received signals, since their H values obtained from each block is followed the H values of the fGn.

A. H-ML-Energy Localization Estimation

In the proposed acoustic source localization with the correlation degree represented by H, (8) must be redefined to compute the maximum likelihood estimation function to achieve the source localization considering the L sensor readings. Thus,

$$\mathbf{Z_H} = \begin{bmatrix} \frac{u_1 - \mu_{H_1}}{\sigma_{H_1}} & \cdots & \frac{u_L - \mu_{H_L}}{\sigma_{H_L}} \end{bmatrix}^T \tag{13}$$

where $\frac{u_i - \mu_{H_i}}{\sigma_{H_i}}$ is the normalized acoustic energy evaluated for each sensor i $(i = 1, \ldots, L)$, and (8) can be rewritten as $\mathbf{Z_H} = \mathbf{H_H B} + \boldsymbol{\xi}_\mathbf{H}$, with

$$\begin{aligned} \mathbf{G_H} &= \operatorname{diag}\begin{bmatrix} \frac{g_1}{\sigma_{H_1}} & \frac{g_2}{\sigma_{H_2}} & \cdots & \frac{g_L}{\sigma_{H_L}} \end{bmatrix} \\ \mathbf{D} &= \begin{bmatrix} \frac{1}{d_{11}^2} & \frac{1}{d_{12}^2} & \cdots & \frac{1}{d_{1K}^2} \\ \frac{1}{d_{21}^2} & \frac{1}{d_{22}^2} & \cdots & \frac{1}{d_{2K}^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{d_{L1}^2} & \frac{1}{d_{L2}^2} & \cdots & \frac{1}{d_{LK}^2} \end{bmatrix} \\ \mathbf{B} &= \begin{bmatrix} B_1 & B_2 & \ldots & B_K \end{bmatrix}^T \\ \mathbf{H_H} &= \mathbf{G_H D} \\ \boldsymbol{\xi}_\mathbf{H} &= \begin{bmatrix} \epsilon_{H_1} & \epsilon_{H_2} & \ldots & \epsilon_{H_L} \end{bmatrix}^T, \end{aligned} \tag{14}$$

where $\mathbf{G_H}$ is the gain matrix, $\mathbf{D}$ represents the attenuation matrix, $\mathbf{B}$ is the acoustic energy source vector and $\boldsymbol{\xi}_\mathbf{H}$ denotes the error vector $(\epsilon_{H_i} = \mathbb{E}[h_i^2(t)])$.

The joint probability density function of $\mathbf{Z_H}$ in matrix form is

$$f(\mathbf{Z_H} \mid \theta) = (2\pi)^{-N/2} \exp\{-\frac{1}{2}(\mathbf{Z_H} - \mathbf{H_H B})^T(\mathbf{Z} - \mathbf{H_H B})\} \tag{15}$$

where,

$$\theta = \begin{bmatrix} \mathbf{p}_1^T & \mathbf{p}_2^T & \ldots & \mathbf{p}_K^T & B_1 & B_2 & \ldots & B_K \end{bmatrix}^T \tag{16}$$

is a vector with the source positions $\mathbf{p}_j$ and their corresponding acoustic energies B_j. Applying the logarithm in (15), it is obtained the log-likelihood function [6],

$$L(\theta) = \|\mathbf{Z_H} - \mathbf{H_H B}\|^2, \tag{17}$$

and its minimum can be found by a computational low-cost multiresolution search (MR) [4] or an exhaustive search (ES). In this work, the MR search is applied in the experiments.

For the H-ML-Energy implementation, the fGn is generated using the midpoint displacement technique [23] with the μ_H, σ_H and H exponent parameters estimated from the signals corrupted with the real acoustic noises, received at each sensor. The Hurst exponent is estimated using the wavelet-based estimator with the 12 coefficients of a Daubechies digital filter [17]. For the ML-Energy, the Gaussian noise is generated using the Box-Muller method [24].

B. Computational Complexity

The computational complexity of the proposed localization solution can be divided into two main parts: the estimation of the H exponent, μ_H and σ_H parameters of the H-ML-Energy has the computational cost of $O(M)$, where M is the number of samples of each block. This is the same cost of the estimation of mean and variance of the baseline ML-Energy solution. And, the computational cost of the search for the minimum value of $L(\theta)$ in (17) depends on the chosen method. Considering N grid points and K sources, the extensive search needs to examine N^{2K} points. However, for the MR search, only Q points are evaluated in each of m iterations, where $Q^m = N$. Thus, the number of search points in the multiresolution search is reduced from N^{2K} to $m \times Q^{2K}$.

IV. Results and Discussion

The proposed H-ML-Energy is evaluated by a series of simulation experiments. All the experiments were performed using real corrupted or noisy signals. In the H-ML-Energy, the acoustic energy of the corrupted signals is computed for each frame using (7). These values and also the μ_H and σ_H parameters, are then used to compose the H-ML matrices (Section III-A) which are finally applied to estimate the source location. The performance of the proposed method is compared to the baseline ML-Energy under two different situations:

- In the first situation, the acoustic energy is obtained with (7), i.e., considering the cross-correlation of the real noisy signals. In this work, this is defined as the practical baseline reference since it enables the validation of the ML-Energy method in real environments.
- In the second situation, the acoustic energy is also calculated using (7), but the cross-correlation term is ignored $(\mathbb{E}[A_i(t)w_i(t)] = 0)$. Since in this second situation it is considered that the source and the noise are uncorrelated (which is the ML-Energy assumption) this is defined as the ideal or ground reference [4].

The acoustic energy values obtained from these two situations are further applied to obtain the ML matrices [4] used to estimate the source location.

Three sources (Car, Helicopter and Speech) and three acoustic noises (Babble, Car, F16) are employed in the Monte Carlo experiments. The real acoustic sources and noises are re-sampled to 16 kHz and have 30 seconds time duration. All these sources and noises are non-stationary, except the Car (see Section II). The three acoustic noises are used to corrupt the sources with five different SNR values: 0 dB to 20 dB, with 5 dB increments, i.e., severe noisy conditions. The SNR is obtained in a point position that is 1 meter away from the source position.

Three different measures are applied for the examination of the source localization accuracy: error probability distribution, Bhattacharyya distance and RMSE. For the localization tests, a two dimensional square field of 100 m × 100 m is used with its origin located at the center of the square. Two sensors configurations are considered: four and ten sensors that are randomly positioned in the field. For each localization experiment, a single acoustic source is randomly positioned in the field and its variable acoustic energy is measured at each sensor. Blocks of M = 1024 samples, i.e., a total of 473 blocks for each source, are used in the experiments. Therefore, for each sensor configuration, 21285 tests are conducted in the evaluation experiments. All the sensors gain are set to $g_i = 1$. The minimum of the log-likelihood function of the H-ML-Energy and the ML-Energy are found using the multiresolution search with 1 meter

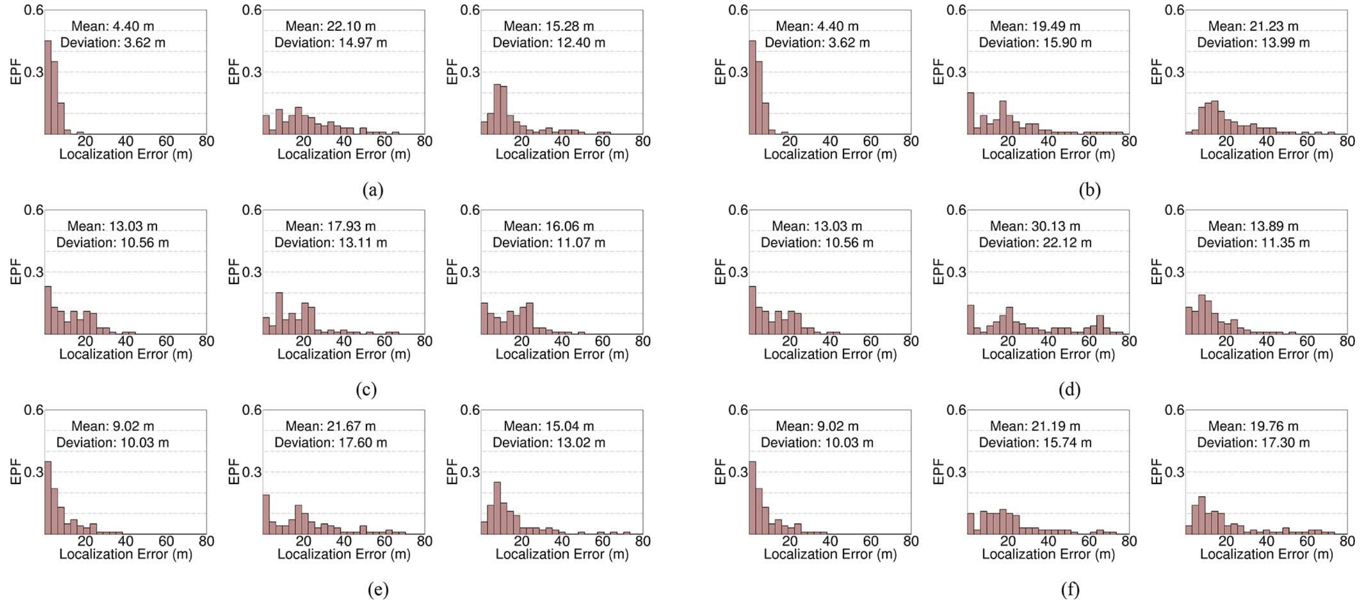


Fig. 4. Distribution of the localization error magnitude of the three sources corrupted by the real acoustic noises with SNR $= 5$ dB, using four sensors. Left: ideal ML-Energy, center: baseline ML-Energy, right: H-ML-Energy. (a) Car + Babble, (b) Car + F16, (c) Helicopter + Car, (d) Helicopter + F16, (e) Speech + F16, (f) Speech + Babble.

increment. For this purpose, it is adopted $m = 2$ iterations with $Q = 10$ search points, such that $Q^m = 100$.

A. Error Probability Function

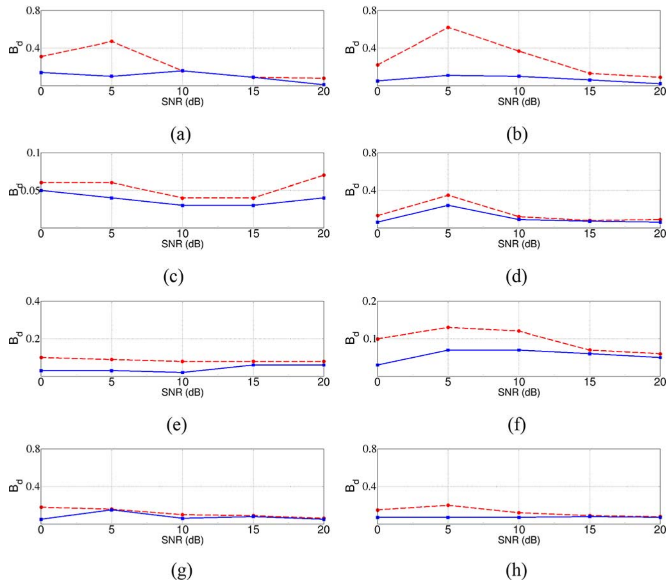
The location error is here defined as the distance between the target source position and the estimated position. The location error values are presented in terms of the error probability density function (EPF) for the three kinds of experiments. The EPF curves from all sources corrupted by the real acoustic noises with SNR $= 5$ dB and the four sensors configuration are summarized in the Fig. 4. The histograms are plotted with a 5 m bin. They also include the mean and the standard deviation of the localization errors. Each group of three histograms represents a source corrupted by a certain noise. From the left to the right, the histogram corresponds to the results obtained with the ideal reference ML-Energy, the practical baseline ML-Energy and the H-ML-Energy, respectively.

It can be noted that the localization errors obtained with the methods are significantly different in distribution. Fig. 4 also shows that the H-ML-Energy outperforms the practical baseline ML-Energy. For example, considering the Car source corrupted with the Babble noise, the mean localization error is 22.10 m with ML-Energy. On the other hand, the H-ML-Energy method achieves a location error mean of 15.28 m, i.e., much lower than the practical reference. As expected, in the situation where the cross-correlation is ignored, the ideal ML-Energy obtains a mean location error of 4.40 m.

In the case where the Speech source is corrupted with the Babble noise, the mean and the standard deviation of the location error obtained by the proposed method are 19.76 m and 17.30 m, respectively. These are lower than the localization error found with the ML-Energy (mean $=$ 21.19 m, standard deviation $=$ 15.74 m), i.e., H-ML-Energy outperforms the practical baseline method. In the same scenario, the ideal reference achieves 9.02 m and 10.03 m, as the mean and the standard deviation localization errors, respectively. This can be considered as an unrealistic localization result particularly for the severe noisy condition (SNR $= 5$ dB).

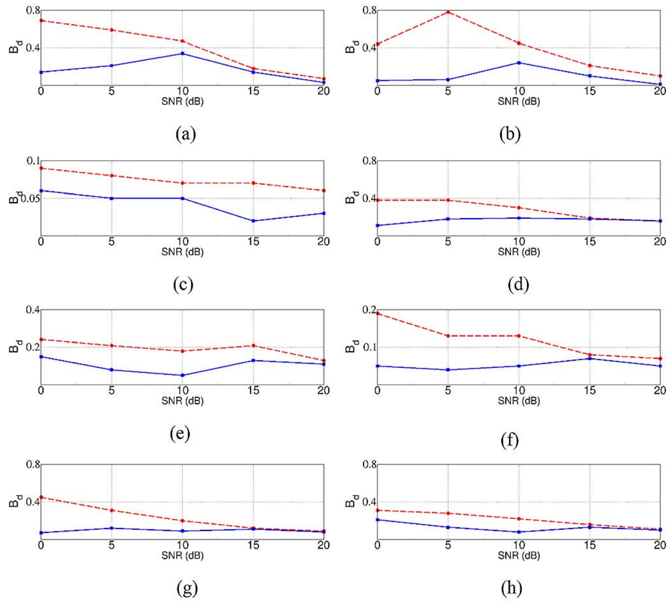
B. Bhattacharyya Distance

The second evaluation measure adopted in this work is the Bhattacharyya distance (B_d) [15], [16]. The Bhattacharyya distance is a real number $(0 \leq B_d < \infty)$ and it is equal to zero when two testing sample sequences have similar distributions. The B_d distance is here applied to measure the distance between the error location distribution of each localization method (H-ML-Energy and the ideal ML-Energy) and the error location distribution of the practical baseline ML-Energy. Figs. 5 and 6 show the Bhattacharyya distance results obtained considering the four and ten sensors configurations, respectively. The blue line indicates the distance between the H-ML-Energy and the localization with the practical baseline. The red line shows the distance measured from the ideal ML-Energy results. It can be observed that for all sources and noises, the H-ML-Energy method achieves the lowest Bhattacharyya distance values, i.e., the localization errors are closer to the ones obtained with the practical ML-Energy method. Considering the four sensors configuration, it can be seen that the Car source corrupted with the Babble noise presents B_d values of 0.62 with the ideal ML-Energy, and SNR $= 5$ dB, while the proposed solution obtains $B_d = 0.11$. For SNR $= 10$ dB, the B_d difference between both methods is reduced, but it is still significant, i.e., 0.11 and 0.37 for the H-ML-Energy and the ML-Energy, respectively. For the highly non-stationary Speech source corrupted with the Babble noise with SNR $= 0$ dB, the B_d values are 0.03 and 0.10 for the H-ML-Energy and the ML-Energy, respectively. These results demonstrate that the H-ML-Energy outperforms the baseline ML-Energy even for a highly non-stationary source and a low value of SNR.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 5. Bhattacharyya distance from location error results obtained from the ML-Energy (red lines) and the H-ML-Energy (blue), using 4 sensors configuration and block size of 1024 samples. (a) Car + F16, (b) Car + Babble, (c) Helicopter+Car, (d) Helicopter+F16, (e) Helicopter+Babble, (f) Speech+Car, (g) Speech + F16, (h) Speech + Babble.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 6. Bhattacharyya distance from location error results obtained from the ML-Energy (red lines) and the H-ML-Energy (blue), using 10 sensors configuration and block size of 1024 samples. (a) Car + F16, (b) Car + Babble, (c) Helicopter + Car, (d) Helicopter + F16, (e) Helicopter + Babble, (f) Speech + Car, (g) Speech + F16, (h) Speech + Babble.

For the experiments using the ten sensors configuration (Fig. 6), it can be observed that the Car source corrupted with the F16 noise (SNR $=$ 0 dB), presents $B_d = 0.14$ for the H-ML-Energy and $B_d = 0.69$ for the ideal ML-Energy. When the Babble noise corrupts the Helicopter source, with SNR $=$ 5 dB, the proposed solution obtains $B_d = 0.08$ while the Bhattacharyya distance with the ML-Energy is 0.21. Note that, for this case, the acoustic signal samples received at the sensors are uncorrelated $(H = 0.52)$, as shown in Table I. However, since both the Helicopter source and the Babble noise are non-stationary (refer to Fig. 2), all the parameters of the corrupted signal are expected to vary in different short-time blocks. The H exponent estimated from each block enables the detection of the acoustic energy variability.

In summary, the proposed solution obtains lower Bhattacharyya distance results mainly for the low SNR values, when the background noise effects are more evident.

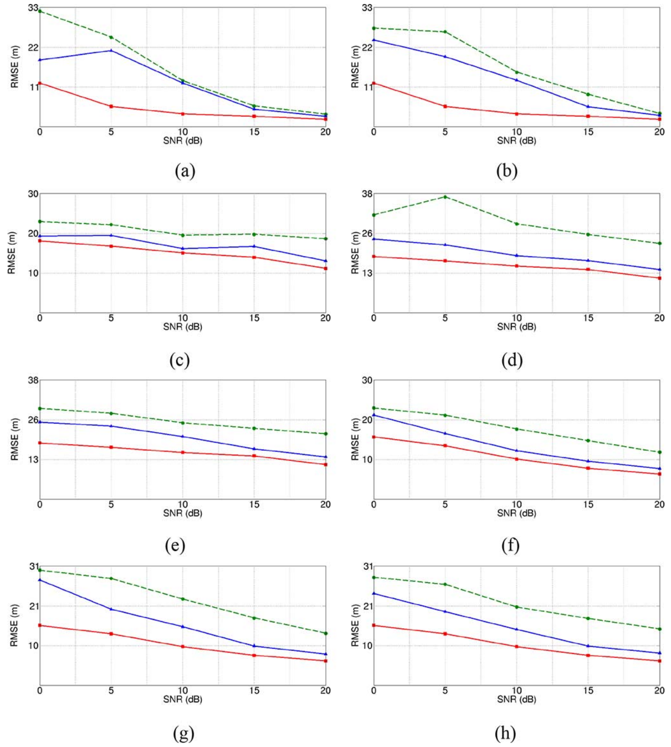
C. Root Mean Squared Error

The root mean squared error (RMSE) is also applied as a third evaluation measure in the experiments. It is defined as,

$$\mathrm{RMSE} = \sqrt{\frac{1}{Q}\sum_{i=1}^{Q}(\hat{\mathbf{r}}_i - \mathbf{r}_i)^2}, \quad (18)$$

where $\mathbf{r}_i$ denotes the target source location during i-th $(i = 1, 2, \ldots, Q)$ block and $\hat{\mathbf{r}}_i$ represents its estimated position during the same block. Here, the purpose of RMSE is to verify how close the estimated localization obtained with the methods are from the target source positions. Figs. 7 and 8 illustrate the RMSE computed for all the SNR values obtained with the four and ten sensors configurations, respectively. The H-ML-Energy RMSE values are represented by the blue line. The green dashed line corresponds to the RMSE from the practical baseline method. The red lines illustrate the RMSE obtained with the ideal ML-Energy. It can be noted from Fig. 7 that for the Car source corrupted with the Babble noise (SNR $=$ 0 dB), the H-ML-Energy RMSE (24.31 m) is lower than the practical baseline value (27.66 m). Once again, this demonstrates that the proposed method outperforms the baseline ML-Energy under noisy condition. On the other hand, the RMSE obtained with the ideal ML-Energy differs in more than 15 m from the practical baseline. For this case, increasing the SNR value to 10 dB, the H-ML-Energy obtains a RMSE value of 13.02 m, while the practical ML-Energy finds a RMSE value of 15.32 m. Interesting results can be observed when the Car source is corrupted with the F16 noise, for SNR $>$ 10 dB. For such cases, the proposed method still outperforms the practical baseline ML-Energy, but the difference between their RMSE values is reduced to less than 1 meter. For the Speech source corrupted with the Babble noise the practical baseline reference presents RMSE equal to 28.19 m, i.e., about 5 meters above the RMSE value obtained with the H-ML-Energy, which is 23.96 m. On the other hand, the ideal ML-Energy method shows a RMSE value of 15.89 m, which differs in almost 13 meters from the practical baseline.

From Fig. 8, it can be seen that for the Helicopter source corrupted with the Babble noise with SNR $=$ 0 dB, the proposed solution obtains a RMSE of 34.72 m, while the practical baseline ML-Energy achieves the RMSE of 40.85 m. In the ideal ML-Energy situation, the obtained RMSE is 17.68 m, which is very distant from the H-ML-Energy and from the practical baseline ML-Energy results. It is important to observe that for some cases, for example, the Helicopter source corrupted with the Car noise, the RMSE obtained with the H-ML-Energy is very close to the ones achieved with the practical baseline ML-Energy (less than 1 meter for the 0 dB case), contrasting with the ideal ML-Energy.



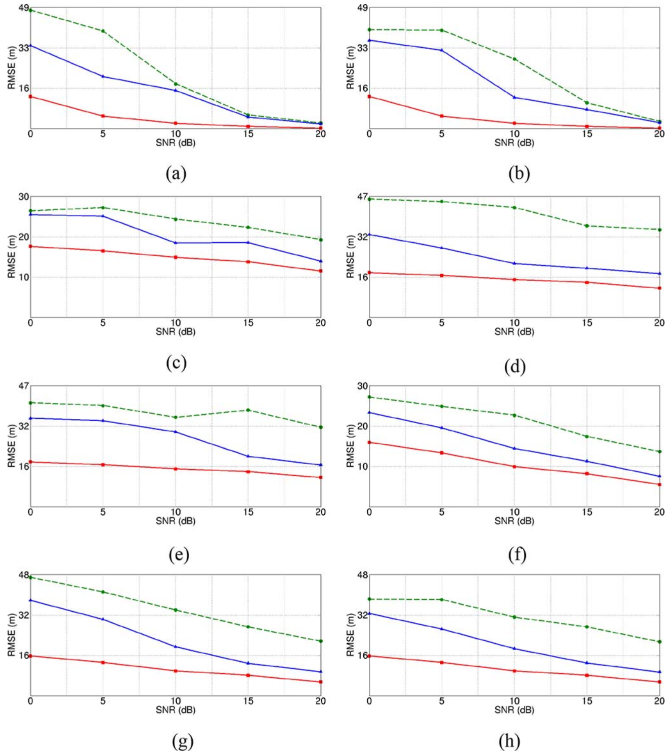
(a) (b)

(c) (d)

(e) (f)

(g) (h)

Fig. 7. RMSE obtained with the practical baseline ML-Energy (green lines) and H-ML-Energy (blue), using 4 sensors configuration. Ideal baseline ML-Energy values are plotted using red lines. (a) Car + F16, (b) Car + Babble, (c) Helicopter + Car, (d) Helicopter + F16, (e) Helicopter + Babble, (f) Speech + Car, (g) Speech + F16, (h) Speech + Babble.



(a) (b)

(c) (d)

(e) (f)

(g) (h)

Fig. 8. RMSE obtained with the practical baseline ML-Energy (green lines) and H-ML-Energy (blue), using 10 sensors configuration. Ideal baseline ML-Energy values are plotted using red lines. (a) Car + F16, (b) Car + Babble, (c) Helicopter + Car, (d) Helicopter + F16, (e) Helicopter + Babble, (f) Speech + Car, (g) Speech + F16, (h) Speech + Babble.

Although the ideal ML-Energy shows smaller RMSE values when compared to the practical baseline and the proposed method, this can be considered as an unrealistic or inaccurate indication. Moreover, it can be seen that despite the different source and noise statistics (refer to Figs. 1 and 2), the ideal localization method achieves almost similar RMSE results for the severe SNR values (<15 dB). This is expected since the ideal ML-Energy ignores the cross-correlation term in its formulation. However, for SNR ≥ 15 dB, i.e., the noise effect is smoothed, the results become much closer to the practical reference and the proposed method.

V. Conclusion

This paper has introduced a novel representation of the acoustic samples cross-correlation for the source localization estimation in real acoustic noise environments. In the H-ML-Energy proposal, the error of the energy readings due to the noise correlation is represented by the Hurst exponent of a fractional Gaussian noise. Several experiments with different real acoustic sources and noises, SNR values and non-stationarity characteristics were conducted to examine the proposed solution. The accuracy of the proposed method was compared to the practical baseline ML-Energy. The results demonstrated that the proposed approach consistently outperforms the baseline ML-Energy when considering real noisy environments.

The investigation of signal enhancement techniques [25], [26] to improve the source localization estimation is worthy for future research.

References

[1] D. Li and Y. Hu, "Energy based collaborative source localization using acoustic micro-sensor array," *EURASIP Appl. Signal Process.*, vol. 4, pp. 321–337, 2003.

[2] D. Blatt and A. Hero, "Energy-based sensor network source localization via projection onto convex sets," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3614–3619, Sep. 2006.

[3] S. Mini, S. Udgata, and L. Sabat, "Sensor deployment and scheduling for target coverage problem in wireless sensor networks," *IEEE Sens. J.*, vol. 14, no. 3, pp. 636–643, 2014.

[4] X. Sheng and Y. Hu, "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 44–53, Jan. 2005.

[5] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'97)*, Munich, Germany, 1997, vol. 1, no. 1, pp. 187–190.

[6] P. Chung and J. Boehme, "The methodology of the maximum likelihood approach—Estimation, detection, and exploration of seismic events," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 40–46, May 2012.

[7] J. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS'03)*, 2003, vol. 2, pp. 1228–1233.

[8] V. Trifa, A. Koene, J. Moren, and G. Cheng, "Real-time acoustic source localization in noisy environments for human-robot multimodal interaction," in *Proc. 16th IEEE Int. Symp. Robot Human Interact. Commun.*, 2007, pp. 393–398.

[9] R. Webster, "Ambient noise statistics," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2249–2253, Jun. 1993.

[10] J. Ming, T. Hazen, J. Glass, and D. Reynolds, "Robust speaker recognition in noisy conditions," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1711–1723, Jul. 2007.

[11] L. Zão and R. Coelho, "Generation of coloured acoustic noise samples with non-gaussian distribution," *IET Signal Process.*, vol. 6, no. 7, pp. 684–688, 2012.

[12] H. Hurst, "Long term storage capacity of reservoirs," *Trans. Amer. Soc. Civil Eng.*, vol. 116, pp. 770–799, 1951.

[13] B. Mandelbrot and J. Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, vol. 10, no. 4, 1968.

[14] P. Borgnat, P. Flandrin, P. Honeine, C. Richard, and J. Xiao, "Testing stationarity with surrogates: A time-frequency approach," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3459–3470, Jul. 2010.

[15] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, 1943.

[16] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. 15, no. 1, pp. 52–60, Feb. 1967.

[17] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.

[18] D. Veitch and P. Abry, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 2–15, Jan. 1998.

[19] R. Sant'Ana, R. Coelho, and A. Alcaim, "Text-independent speaker recognition based on the Hurst parameter and the multidimensional fractional Brownian motion model," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 3, pp. 931–940, May 2006.

[20] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," *Linguist. Data Consortium*, 1993.

[21] A. Varga and H. Steeneken, "Assessment for automatic speech recognition II: NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, pp. 247–251, 1993.

[22] L. Kinsler, *Fundamentals of Acoustics*. New York, NY, USA: Wiley, 1982.

[23] M. Barnsley, R. Devaney, B. Mandelbrot, H. Peitgen, D. Saupe, and R. Voss, *The Science of Fractal Images*. New York, NY, USA: Springer-Verlag, 1988.

[24] G. Box and M. Muller, "A note on the generation of random normal deviates," *Ann. Math. Statist.*, vol. 29, no. 2, pp. 610–611, 1958.

[25] L. Zão, R. Coelho, and P. Flandrin, "Speech enhancement with emd and hurst-based mode selection," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 5, pp. 897–909, May 2014.

[26] T. Gerkmann and R. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, May 2012.

Eloi Dranka Junior obtained the M.Sc. degree from the Military Institute of Engineering (IME) of Rio de Janeiro in 2014. From the same Institute, he received the B.Sc. degree in electrical engineering in 2009. His current research mainly focuses on acoustic signal processing, detection and estimation of non-stationary signals and signal processing for source localization in wireless sensor networks.

Rosângela Fernandes Coelho received the Ph.D. degree from the Ecole Nationale Supérieure des Télécommunications (ENST-Télécom ParisTech) in 1995 and the M.Sc. degree from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) in 1991, both in electrical engineering.

She joined the Military Institute of Engineering (IME) of Rio de Janeiro, in 2002, where she is Associate Professor at the Electrical Engineering Department. Prof. Coelho founded and heads the Laboratory of Acoustic Signal Processing (LASP). In 2003, she received the University Research Program grant award from CISCO/USA. She also served as editorial board member of the IEEE Communications Surveys and Tutorials from 1999–2007. Since 2008, she is responsible for the International Scientific Collaboration IME-ParisTech that includes 10 french engineering schools. Prof. Coelho was President-Adjoint of the Brazilian Telecommunications Society from 2008–2010 and she is member of the IEEE Signal Processing Society. In 2011, Prof. Coelho received the USPTO patent of an automatic speaker recognition method based on a new speech feature and speaker classifier. Her main research interests include acoustic signal processing, speech enhancement and intelligibility, speech and speaker recognition, nonlinear and non-stationary signal analysis, acoustic emotion detection and classification, acoustic speech features, and statistical signal processing.