Sensor signal processing

Effective Acoustic Energy Sensing Exploitation for Target Sources Localization in Urban Acoustic Scenes

Marília Alves*, Rosângela Coelho**, and Eloi Dranka

Laboratory of Acoustic Signal Processing, Military Institute of Engineering, Rio de Janeiro 22290-270, Brazil
**Student Member, IEEE*
***Senior Member, IEEE*

Manuscript received August 16, 2019; accepted December 15, 2019. Date of publication December 19, 2019; date of current version February 19, 2020.

Abstract—This letter proposes a new approach to improve the accuracy of the energy-based source localization methods in urban acoustic scenes. The proposed acoustic energy sensing flow estimation (ESFE) uses the sensors signal nonstationarity degree to determine the area with highest energy concentration in the scenes. The ESFE is applied to different acoustic scenes and yields to source localization accuracy improvement with computational complexity reduction. The experiments results show that the proposed scheme leads to significant improvement in source localization accuracy.

Index Terms—Sensor signal processing, acoustic, energy-based source localization, index of nonstationarity (INS), wireless acoustic sensor network (WASN).

I. INTRODUCTION

Wireless acoustic source network (WASN) is a very attractive solution for source localization in urban areas [1], [2]. WASN enables low cost and low power coverage for indoor and outdoor acoustic scene environments. Accurate source location estimation is an ubiquitous issue in a diversity of applications including objects monitoring, seismic event detection, house surveillance, and smart vehicle tracking.

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 [2]–[5]. TDE or TDOA algorithms use time-delay or phase difference measures obtained at the acoustic sensors generally distributed in a microphone array. Energy-based techniques are simple and interesting solutions widely applied for sound source estimation localization in WASN. The main challenge is the background acoustic interference that can severely affect a target location estimation, particularly when considering real acoustic scenes. Generally, each acoustic scene is composed of multiple sources with different temporal and spectral statistics.

The main goal of this letter is twofold. First, it applies energy-based source localization methods in acoustic scenes environment. Second, it introduces an efficient acoustic energy sensing flow exploitation (ESFE) approach for energy-based source localization accuracy improvement. The proposed scheme defines the nonstationary acoustic energy flow formed by individuals sources that composes a scene. The selection of sensors is based on the nonstationarity degree of the collected acoustic sensors amplitude signals. The ESFE enables location estimation accuracy improvement of the energy-based localization methods with reduced number of sensors. The Cramér–Rao lower bound (CRLB) is also derived to examine the robustness of the H-ML-Energy method.

Extensive experiments are conducted to evaluate the effectiveness of the proposed ESFE solution. For this purpose, outdoor Park and indoor Kitchen scenes are simulated using a diversity of real acoustic sources signals. Index of nonstationarity (INS) [6] of the sensor signals are applied for the sensor node selection in each scenario. The maximum likelihood (ML) energy-based source localization methods, i.e., ML-Energy [5] and H-ML-Energy [4], are examined before and after the application of the proposed approach. This solution is also compared to a sensor selection method based on noise reduction [3], [7]. Experiments are conducted with four different values of SNR (signal-to-noise ratio) ranging from 0 to 15 dB. Experimental results demonstrate that the proposed ESFE scheme improves the accuracy of the energy-based source localization methods while reducing the number of sensors in acoustic scenes with real sources.

Corresponding author: Rosangela Coelho (e-mail: coelho@ime.eb.br).
Associate Editor: R. Vida.
Digital Object Identifier 10.1109/LSENS.2019.2960888

II. SOURCE LOCALIZATION IN ACOUSTIC SCENES

Acoustic scenes are composed of multiple sound sources that naturally belongs to this environment, including animals, people, and objects. In order to evaluate the ESFE method, two acoustic scenes[1] were artificially composed of six distinct real omnidirectional acoustic sources randomly placed in the delimited area of each scene. First scene is outdoor Park, that is composed of single sources "speaker," "waterfall," "birds," "dogs barking," "children playing," and "babble." Another scene is indoor Kitchen with sources "speaker," "television," "water," "sizzling," "cutting," and "clanking dishes."

A. Index of Nonstationarity

A signal sample sequence is defined as stationary if its main statics are time invariant. The INS[1] is a time-frequency approach to objectively examine the nonstationarity of a signal. The stationarity test is conducted by comparing spectral components of the signal to a set of stationary references, called surrogates. For this purpose, spectrograms of the signal and surrogates are obtained by means of the short-time Fourier transform 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 [6], Borgnat *et al.* considered that the distribution of the KL values can be approximated by a Gamma distribution. Therefore, for each window length, a threshold γ is defined for the stationarity test considering a confidence degree of 95%. Thus

$$\text{INS} \begin{cases} \leq \gamma, & \text{signal is stationary} \\ > \gamma, & \text{signal is nonstationary} . \end{cases} \quad (1)$$

[1]Available at lasp.ime.eb.br.

1949-307X © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Authorized licensed use limited to: INSTITUTO MILITAR DE ENGENHARIA. Downloaded on February 21,2020 at 13:37:20 UTC from IEEE Xplore. Restrictions apply.

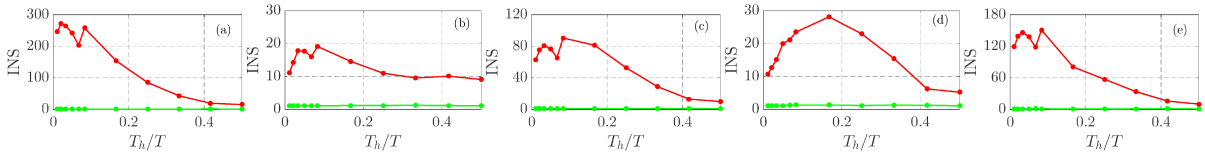

Fig. 1. INS values obtained for 3-s audio signals. The red line represents the INS value of each time scale T_h/T. The threshold is indicated by the green lines. (a) "speaker." (b) Park. (c) "speaker"+Park (15 dB). (d) Kitchen. (e) "speaker"+Kitchen (15 dB).

Fig. 1 illustrates the INS values of five studied signals: the source "speaker," acoustic scenes Park and Kitchen, and the source corrupted by the scenes as background 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 ($T = 3$ s) of the signal. Red lines represent INS values and green lines indicate threshold values. A signal is stationary if its INS value is below the threshold for every time scale. Signals with INS values above the threshold for the majority of the time scales are classified as nonstationary. A signal is considered as highly nonstationary if the maximum INS value $\text{INS}_{\max}$ is greater than 100. The "speaker" source is here classified as highly nonstationary, whereas the two scenes are nonstationary. When corrupted by noise, the $\text{INS}_{\max}$ of a signal decreases since it becomes more stationary. Note that the "speaker" source has $\text{INS}_{\max} \approx 300$. However, its $\text{INS}_{\max}$ is reduced to 80 and 150 when corrupted by the Park and Kitchen scenes, respectively. In other words, the INS is a representative parameter to indicate if a target signal is corrupted by noise. The ESFE method exploits this INS characteristic for the selection of most representative sensors in the WASN.

III. ENERGY-BASED SOURCE LOCALIZATION IN ACOUSTIC SCENES

Energy-based source localization methods are based on the fact that the acoustic energy attenuation is inversely proportional to the distance from the signal to the multiple acoustic sensors distributed in the field [8]. An ML approach was presented in [5] for acoustic source position estimation (ML-Energy) and later in [4] for noisy correlated environments (H-ML-Energy). In this letter, the methods ML-Energy and H-ML-Energy are going to be adjusted for the acoustic scene environment.

Localization methods in acoustic scenes are applied to estimate the target source position, while the summation of the other sources is considered as noise. The signal received at the ith sensor is sampled during the nth time interval with a sampling frequency f_s is defined as $x_i(n) = A_i(n) + W_i(n)$, where $A_i(n)$ is the acoustic signal intensity or energy given by

$$A_i(n) = \sqrt{g_i} \sum_{j=1}^{K} \frac{s_j(n - \tau_{ji})}{|\mathbf{p}_j(n - \tau_{ji}) - \mathbf{r}_i|}, \tag{2}$$

where g_i represents the sensor gain, s_j is the signal intensity of the jth source, $\mathbf{r}_i$ is the sensor position, and $\mathbf{p}_j$ is the jth ($j = 1, \ldots, K$) source spatial coordinates. In acoustic scenes, the total noise intensity $W_i(n)$ is defined as

$$W_i(n) = \sqrt{g_i} \sum_{m=1}^{M} \frac{o_m(n - \tau_{ji})}{|\mathbf{w}_m(n - \tau_{ji}) - \mathbf{r}_i|}, \tag{3}$$

where o_m represents the mth ($m = 1, \ldots, M$) noise source intensity and $\mathbf{w}_m$ its position. Given the time index t, the acoustic energy in the ith sensor $\mathbb{E}[x_i^2(n)] = u_i(t)$ is given by

$$u_i(t) = g_i \sum_{j=1}^{K} \frac{B_j(t)}{d_{ij}^2(t)} + 2\mathbb{E}[A_i(t)w_i(t)] + \mathbb{E}[W_i^2(t)], \tag{4}$$

where B_j is the jth source acoustic energy, and d_{ij} is the distance between the source j and sensor i. In the ML-Energy, the background noise is modeled as uncorrelated with the source signal, the cross term $\mathbb{E}[A_i(t)W_i(t)]$ is considered equal to zero. This assumption can severely degrade the sensor measurements and the source localization estimation accuracy. In [4], the correlation between the source and the signal is taken in consideration. In the H-ML-Energy method, the cross term $\mathbb{E}|A_i(t)W_i(t)|$ and the term $\mathbb{E}|W_i^2(t)|$ are modeled by a fractional Gaussian noise (fGn) with exponent H, mean μ_H, and variance σ_H, obtained from the sensor readings. Denoting the fGn process by $h_i(t) = 2\mathbb{E}[A_i(t)W_i(t)] + \mathbb{E}[W_i^2(t)]$, the acoustic source localization energy is defined as

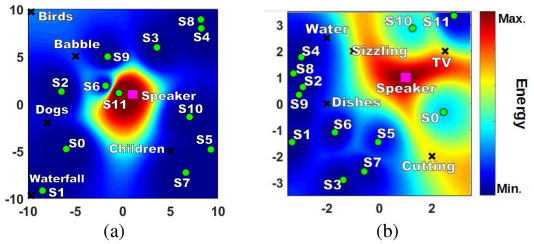

Fig. 2. Energy distribution scene maps. Sensors are represented by circles, target source by square, and other/noise sources by "×." (a) Park (20 × 20 m). (b) Kitchen (7 × 7 m).

$$u_i(t) = g_i \sum_{j=1}^{K} \frac{B_j(t)}{d_{ij}^2(t)} + h_i(t). \tag{5}$$

The fGn process represents the energy measurement error. Since the fGn is able to indicate any degree of correlation by means of its exponent, the H-ML-Energy grants better accuracy to the energy-based localization model. The vector $\mathbf{Z_H}$ represents the normalized acoustic energy in each sensor $i (i = 1, \ldots, L)$. Thus, $\mathbf{Z_H} = [\frac{u_1 - \mu_{H_1}}{\sigma_{H_1}} \cdots \frac{u_L - \mu_{H_L}}{\sigma_{H_L}}]^T$. The joint probability density function $\mathbf{Z_H}$ in matrix form is

$$f(\mathbf{Z_H}|\boldsymbol{\theta}) = (2\pi)^{-L/2} \exp\left\{-\frac{1}{2}(\mathbf{Z_H} - \mathbf{G_H}\mathbf{D}\mathbf{B})^T(\mathbf{Z_H} - \mathbf{G_H}\mathbf{D}\mathbf{B})\right\}, \tag{6}$$

where $\mathbf{G_H}$ represents the gain matrix, $\mathbf{D}$ is the attenuation matrix, $\mathbf{B}$ is the acoustic energy source vector, and $\boldsymbol{\theta} = [\rho_1^T \rho_2^T \cdots \rho_K^T B_1 B_2 \cdots B_K]^T$ is a vector with the source positions ρ_j and their corresponding acoustic energies B_j. These matrices are defined in [4]. In this letter, multiresolution search [5] is applied to obtain the minimum value of the log-likelihood function

$$L(\boldsymbol{\theta}) = ||\mathbf{Z_H} - \mathbf{G_H}\mathbf{D}\mathbf{B}||^2. \tag{7}$$

Fig. 2 illustrates acoustic energy distribution maps of Park and Kitchen scenes where the higher acoustic energy is represented in red and the lower in dark blue. The maps represents the simulation of the acoustic scenes that consists on the random distribution of target sources (magenta square), noise sources (black "×"), and microphones (green circles) in scene area. Energy distribution is estimated based on the sensor readings according to (7). Acoustic target and noise sources are nonstationary, then energy distribution estimation will vary at each signal frame. It also depends on the position and number of sensors distributed.

In this letter, the derivation of the CRLB is introduced in order to evaluate the performance of the H-ML-Energy estimator. The

Authorized licensed use limited to: INSTITUTO MILITAR DE ENGENHARIA. Downloaded on February 21,2020 at 13:37:20 UTC from IEEE Xplore. Restrictions apply.

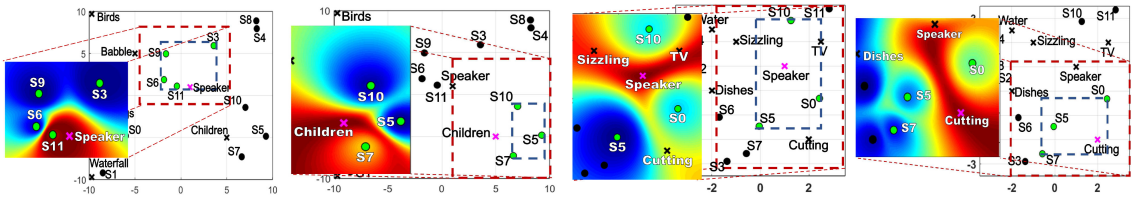

Fig. 3. ESFE method applied in different scenarios. The blue dashed rectangle is the smallest space delimited by the selected sensors, and the red dashed rectangle is the new target source search area. (a) Scene:Park/Source:"speaker." (b) Scene:Park/Source:"children." (c) Scene:Kitchen/Source:"speaker." (d) Scene:Kitchen/Source:"cutting."

CRLB is a theoretical lower bound of the variance of an unbiased parameter estimate [5], [9]. First, the Fisher matrix $\mathbf{F}$ is calculated as $\mathbf{F} = -\mathbb{E}[\frac{\partial}{\partial \boldsymbol{\theta}}(\frac{\partial}{\partial \boldsymbol{\theta}} \ln f(\mathbf{Z_H}|\boldsymbol{\theta})^T]$. From (6), the Fisher matrix can be rewritten as $\mathbf{F} = \frac{\partial(\mathbf{DB})}{\partial \boldsymbol{\theta}} \mathbf{G_H^T G_H} \frac{\partial(\mathbf{DB})}{\partial \boldsymbol{\theta}^T}$. The term $\frac{\partial(\mathbf{DB})}{\partial \boldsymbol{\theta}}$ can be derived as

$$\mathbf{C^T}_j = \frac{\partial(\mathbf{DB})^T}{\partial \rho_j} = -2B_j \left[\frac{c_{1j}}{d_{1j}^3} \cdots \frac{c_{Nj}}{d_{Nj}^3} \right], \tag{8}$$

where $c_{ij} = \frac{\partial d_{ij}}{\partial \rho_j} = \frac{(\rho_j - r_i)}{d_{ij}}$ is a unit vector from jth source to the ith sensor. Considering $\frac{\partial(\mathbf{DB})^T}{\partial \mathbf{B}} = \mathbf{D}$, then $\mathbf{F}$ can be expressed as

$$\mathbf{F} = \begin{bmatrix} \mathbf{C}^T \\ \mathbf{D}^T \end{bmatrix} \mathbf{G^T G} \Big[\mathbf{CD} \Big]. \tag{9}$$

Finally, the CRLB is computed as $\text{CRLB} = \sqrt{\sum_{i=1}^{Q} \frac{[\mathbf{F}_{11}^{-1}] + [\mathbf{F}_{22}^{-1}]}{Q}}$, where Q is the number of blocks.

IV. PROPOSED ESFE METHOD

Acoustic sensing flow is defined as the area with the highest energy in the scene space. This flow is created by the energy emitted from the target sources. Hence, the WASN sensors placed near the source are mainly affected by the energy flow and thus forming an acoustic concentration region. The sensors signals in this region are less corrupted by the scene noise and consequently are more nonstationary. The ESFE consists on the selection of the sensors according to their $\text{INS}_{\max}$ values that lead to new sensing flow area and thus search field reduction. The location estimation focuses on the most nonstationary sensor signals. For this purpose, let $x_i, i = 1, \ldots, L$ be the signals from the L sensors. Selected signals are those with the highest values of $\text{INS}_{\max}$. The selection criterion is defined as

$$\frac{|\text{INS}_{\max i} - \max(\text{INS}_{\max})|}{\max(\text{INS}_{\max})} \geq \alpha, \tag{10}$$

where $\max(\text{INS}_{\max}) = \max_{1 \leq i \leq L} \text{INS}_{\max i}$, and the selection threshold $\alpha \in [0, 1]$ is a function of the number of sensors (L) and the highest scene dimension (v) defined by the ESFE algorithm, i.e.,

$$\alpha = \frac{1}{\kappa(v + L)} + \xi, \tag{11}$$

where κ is the measurements adjustment factor and ξ is the estimation error. The values $\kappa = 0.087$ and $\xi = \pm 5\%$ were defined according to extensive experiments. Finally, source location is estimated using only the selected sensors. The vector Z'_H of the normalized acoustic energy in the selected sensors is given by $\mathbf{Z'_H} = [\frac{u_1 - \mu_{H_1}}{\sigma_{H_1}} \cdots \frac{u_N - \mu_{H_N}}{\sigma_{H_N}}]^T$. The log-likelihood function is $L(\boldsymbol{\theta}) = ||\mathbf{Z'_H} - \mathbf{G'_H D' B}||^2$, where N is the number of selected sensors, and $\mathbf{G_H}$ and $\mathbf{D_H}$ represent the gain and attenuation matrices of the selected sensors, respectively. The proposed scheme also leads to a reduction on the location search field based on the selected sensors positions. Fig. 3 shows the ESFE approach applied in four different scenarios. The green circles correspond to the selected sensors, and the black circles other sensors. The blue-dashed rectangle represents the sensing flow area delimited by the selected sensors and red one the new target source search area. The new search area is the one delimited by the selected sensors with the addition of a security area that is defined according to the scene dimension. In these experiments, 20% of v were added to each side of the rectangle. The energy distribution in the energy flow cluster is also presented in the left of each scene. Note that the flow area has different sizes and location depends on the scene and the target source.

Table 1. $\text{INS}_{\max}$ and B_d Results of Each Sensor

Sensor	Park				Kitchen			
	"speaker" $\alpha = 66\%$		"children" $\alpha = 63\%$		"speaker" $\alpha = 28\%$		"cutting" $\alpha = 36\%$	
	$\text{INS}_{\max}$	B_d	$\text{INS}_{\max}$	B_d	$\text{INS}_{\max}$	B_d	$\text{INS}_{\max}$	B_d
S0	8.048	0.105	15.452	0.218	**43.696**	**0.016**	**35.062**	**0.010**
S1	3.072	0.091	7.694	0.157	21.442	0.065	12.503	0.067
S2	9.446	0.095	13.431	0.223	25.801	0.044	14.975	0.053
S3	**18.955**	**0.081**	16.241	0.228	26.846	0.064	16.497	0.045
S4	8.157	0.116	14.478	0.243	20.708	0.049	12.058	0.061
S5	9.000	0.113	**30.379**	**0.132**	**36.499**	**0.030**	**27.527**	**0.017**
S6	**32.921**	**0.039**	16.572	0.204	32.535	0.039	17.636	0.041
S7	11.077	0.110	**44.143**	**0.079**	27.331	0.053	**22.717**	**0.030**
S8	5.962	0.118	13.048	0.246	21.722	0.053	12.370	0.065
S9	**20.707**	**0.070**	14.485	0.220	24.387	0.049	14.952	0.054
S10	16.104	0.091	**30.661**	**0.127**	**43.525**	**0.015**	12.060	0.056
S11	**55.571**	**0.005**	17.910	0.162	28.922	0.038	11.171	0.066

For the evaluation of the proposed ESFE scheme, a method based on SNR is adopted for sensor selection. This approach was adopted in [10] for large WASN ($L > 80$) in order to save energy and extend network lifetime. This technique is here adapted to the energy-based source localization methods and compared with the proposed ESFE. The sensors are chosen according to the highest values of SNR *a posteriori*, which is defined as $\text{SNR}_{\text{post}} = \frac{E[x(t)^2]}{E[n(t)^2]} = \frac{\sigma_x^2}{\sigma_n^2}$, where $x(t)$ is the noisy signal and $n(t)$ is the estimated noise given the time index t. According to the authors, the procedure stops when half of the sensors are selected by the algorithm.

V. SIMULATIONS AND RESULTS

Extensive experiments are conducted to evaluate the accuracy improvement in energy-based source localization methods. Networks of $L = 12$ and $L = 20$ omnidirectional sensors are randomly positioned in Park and Kitchen scenes. Each sequence has time duration of 3 s and is sampled at 16 kHz. One target source and five noise sources were chosen for each experiment. The "speaker" is considered as the target source in both scenes while the other sources are assumed as noise. Furthermore, the source "children" in the Park scene and the source "cutting" in the Kitchen scene are adopted as target sources.

Table 1 shows the $\text{INS}_{\max}$ values for x_i on four different scenarios considering 12 sensors network and the selection threshold α. The Bhattacharyya distance (B_d) [11] is here used to confirm the efficiency of $\text{INS}_{\max}$ as a sensor selection parameter. B_d compares two probability distributions $p_1(x)$ and $p_2(x)$ of two acoustic signals $s_1(t)$ and $s_2(t)$ as $B_d = -\ln \int \sqrt{p_1(x) p_2(x)} dx$. In this letter, $s_1(t)$ corresponds to the original target source signal and $s_2(t)$ to the sensor signals at each scenario. The sensors signals less corrupted by the scene are going to have the lowest values of B_d. Note that, as expected, the sensor with lower B_d corresponds to the ones with higher $\text{INS}_{\max}$ proving that the INS can be used as selection parameter. The selected sensors are highlighted in bold numbers.

The source localization estimation is conducted using energy-based methods, H-ML-Energy, and ML-Energy, before and after the application of the ESFE scheme. Four different SNR conditions are

Authorized licensed use limited to: INSTITUTO MILITAR DE ENGENHARIA. Downloaded on February 21,2020 at 13:37:20 UTC from IEEE Xplore. Restrictions apply.

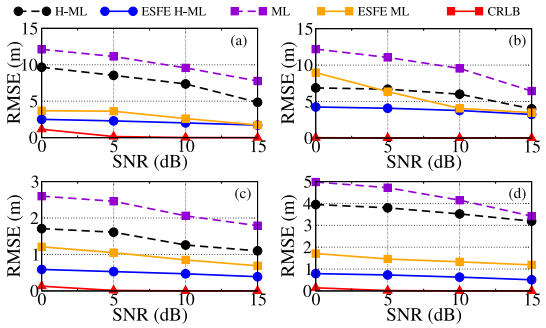

Fig. 4. RMSE analysis L = 12 sensors. (a) Park/"speaker." (b) Park/"children." (c) Kitchen/"speaker." (d) Kitchen/"cutting."

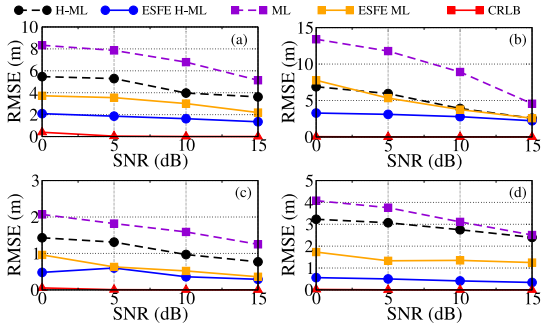

Fig. 5. RMSE analysis L = 20 sensors. (a) Park/"speaker." (b) Park/"children." (c) Kitchen/"speaker." (d) Kitchen/"cutting."

evaluated: from 0 to 15 dB, with 5 dB increments. The SNR is calculated in a position 1 m away from the source position. Blocks of $M = 1024$ samples, i.e., a total of 46 blocks, are used in the evaluation experiments. Therefore, for each scene and each WASN with 12 and 20 sensors deployment, 552 tests are conducted considering three different target sources ("speaker," "children," and "cutting") and four SNR values. The gain of the sensors 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 [5] with 0.1 m in the Park scene (20 × 20 m), and 0.035 m in the Kitchen scene (7 × 7 m). The root mean squared error (RMSE) is applied as evaluation measure in the experiments. It is defined as $\text{RMSE} = \sqrt{\frac{1}{Q}\sum_{i=1}^{Q} |\hat{r}_i - r_i|^2}$, where r_i denotes the target source location of ith $(i = 1, 2, \ldots, Q)$ block and $\hat{r}_i$ represents its estimated position. The RMSE is used to verify how close the estimated localization are from the target source positions.

Figs. 4 and 5 depict the RMSE values obtained with 12 and 20 sensors, respectively. They include H-ML-Energy and ML-Energy, the proposed ESFE for the H-ML-Energy and ML-Energy, and the CRLB estimation. Note that the application of the proposed scheme in the energy-based location methods reduced the error estimation for studied scenarios and made it approaches the CRLB. The new scheme was more effective for the source "children" in the Park scene ($L = 12$), RMSE values in the ML reduced from 12.19 to 2.66 m in 0 dB. The lowest reduction was for the source "speaker" in the Kitchen scene ($L = 20$) where the RMSE varied from 0.77 to 0.29 m in 15 dB. It can also be observed that the H-ML-Energy outperforms the ML-Energy before and after the ESFE application in both scenarios, mainly for low SNR. This is explained by the fact that the H-ML-Energy takes in consideration the correlation between the signal from the source and the interference of the scene.

Table 2 indicates the computational complexity which refers to the processing time required for each algorithm evaluated for 1024 samples per frame. In other compare to the processing time of the different methods, the values were normalized according to the ESFE method with $L = 12$ sensors as reference (execution time = 1). The

Table 2. Normalized Mean Processing Time.

H-ML-12	ESFE-12	H-ML-20	ESFE-20
3.33	1.00	4.30	1.19

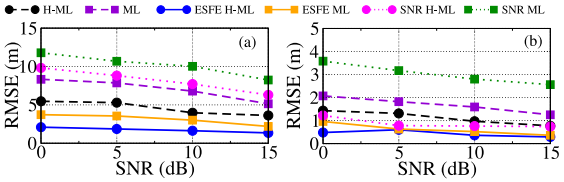

Fig. 6. RMSE comparison analyses. (a) Park. (b) Kitchen.

new approach not only improves the localization estimation accuracy but also decreases the processing period more than three times in both network sizes.

The selection of sensors based on SNR was evaluated and compared with the proposed method for the Park and Kitchen scenes with $L = 20$. The target source was the "speaker," since SNR-based selection methods depend on voice activity detectors [10]. Fig. 6 presents the RMSE values for the different approaches. The best results of the SNR-based selection were achieved in the Kitchen scene for H-ML-Energy method.

VI. CONCLUSION

This letter introduced an effective acoustic energy sensing approach ESFE to improve the accuracy of energy-based source localization methods in acoustic scenes. The new scheme detects the scene energy flow based on the nonstationarity index of the sensor signals readings. Several experiments were conducted with different acoustic scenes and target sources. The results demonstrated that the proposed approach consistently improves the localization estimation while reducing the number of sensors.

ACKNOWLEDGMENT

This work was supported in part by the National Council for Scientifc and Technological Development (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) under Grant 307866/2015 and Grant 203075/2016.

REFERENCES

[1] W. Meng and W. Xiao, "Energy-based acoustic source localization methods: A survey," *Sensors*, vol. 17, no. 2, 2017, Art. no. 376.
[2] L. Lu, H. Zhang, and H.-C. Wu, "Novel energy-based localization technique for multiple sources," *IEEE Syst. J.*, vol. 8, no. 1, pp. 142–150, Mar. 2014.
[3] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Commun. Mobile Comput.*, vol. 2017, 2017, Art. no. 3956282.
[4] E. Dranka and R. Coelho, "Robust maximum likelihood acoustic energy based source localization in correlated noisy sensing environments," *J. Sel. Topics Signal Process.*, vol. 9, no. 2, pp. 259–267, 2015.
[5] X. Sheng and Y.-H. 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.
[6] 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.
[7] F. Deng *et al.*, "Energy-based sound source localization with low power consumption in wireless sensor networks," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 4894–4902, Jun. 2017.
[8] L. Kinsler, *Fundamentals of Acoustics*. New York, NY, USA: Wiley, 1982.
[9] I. Djurovi, "Achieving Cramer Rao lower bounds in sensor network estimation," *IEEE Sensors Lett.*, vol. 2, no. 1, pp. 1–4, Mar. 2018.
[10] J. Szurley, A. Bertrand, M. Moonen, P. Ruckebusch, and I. Moerman, "Energy aware greedy subset selection for speech enhancement in wireless acoustic sensor networks," in *Proc. 20th Eur. Signal Process. Conf.*, 2012, pp. 789–793.
[11] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. 15, no. 1, pp. 52–60, Feb. 1967.

Authorized licensed use limited to: INSTITUTO MILITAR DE ENGENHARIA. Downloaded on February 21,2020 at 13:37:20 UTC from IEEE Xplore. Restrictions apply.