COMPARISON OF ACOUSTIC SOURCE LOCALIZATION METHODS IN TIME DOMAIN USING SPARSITY CONSTRAINTS

Thomas Padois,[1] Franck Sgard[1], Olivier Doutres[2] and Alain Berry[3]

IRSST [1]
505 boulevard De Maisonneuve Ouest, Montréal (QC), H3A 3C2, CANADA
ETS [2]
1100 Rue Notre-Dame Ouest, Montréal (QC), H3C 1K3, CANADA
GAUS [3]
2500 Boulevard de l'Université, Sherbrooke (QC), J1K 2R1, CANADA

This paper deals with source localization techniques in time domain for broadband acoustic sources. The goal is to detect accurately and quickly the position and amplitude of noise sources in workplaces in order to prevent employees from hearing loss or safety risk. First, the generalized cross correlation associated with a spherical microphone array is used to get a raw noise source map. Then a linear inverse problem is defined. Commonly, linear inverse problem is solved with an l_2-regularization. In this study, two sparsity constraints are used to solve the inverse problem, the orthogonal matching pursuit and the truncated Newton interior-point method. Synthetic data are used to highlight the abilities of such techniques. High resolution can be achieved for various acoustic sources configurations. Moreover, the amplitudes of the acoustic sources are correctly estimated. Finally, a comparison of computation time shows these techniques are suitable in real scenario.

1 INTRODUCTION

The goal of this study is to develop an acoustic tool to accurately and quickly localize acoustic noise sources in workplaces such as an industrial hall. Many workers are exposed to high sound levels at work that may be harmful and lead to hearing loss or safety risk. Passive solutions have been developed to reduce noise emitted by sources based on acoustic panels, curtains, enclosures or damping materials. However, the first step in an acoustic diagnosis is to accurately localize the position of the noise source in order to act at the right place. Commonly, the dimensions of an industrial hall are large and the workers undergo the direct sound field and multiple reflections. Therefore, the source localization method has to correctly identify all the source positions and reflections in order to adequately design and implement noise control solutions.

[1]email: Thomas.Padois@irsst.qc.ca

Intensimetry is a technique to localize acoustic sources [1]. The goal is to scan the sound field around an object with a two-microphone probe in order to estimate the radiated acoustic power. Then, this information can be used as input to ray tracing software to predict the sound field in a closed environment. However this technique is time-consuming and the source locations have to be known *a priori*.

An alternative technique is to use an array of microphones associated with a source localization algorithm [2]. The goal is to compensate the time or phase delay between two microphones in relation to a virtual scan point. The processing is performed either in time or frequency domain. Frequency methods use the cross spectral matrix between the microphone signals. The most common technique is known as beamforming [3]. The main disadvantage is the poor spatial resolution at low frequencies. Deconvolution techniques have been developed to improve the resolution of the noise source map [4–6]. Recent works based on inverse methods with a l_1-regularization have shown good performances [7]. However, in an industrial hall, the noise sources are generally broadband so that the computational cost is large since the processing has to be done for each frequency.

The most common technique in time domain is known as the Generalized Cross Correlation (GCC) which is based on the time-delay between a pair of microphones [8]. This time-delay generates an hyperbola for the possible source positions over the scan zone. The intersection of all the hyperbolas (for all the microphone pairs) provides the source positions.

Noël *et al.* [9] have used the GCC associated to an inverse problem to localize source positions in an industrial hall. They developed an inverse problem which minimizes the difference between theoretical and measured cross-correlation. They obtained a noise source map with the angular energy flow received from each direction relative to the microphone antenna. The results have shown good performances despite of a small number of scan points and large computational cost. In this study a minimization problem is also proposed but with a different theoretical formulation, associated to a l_1-regularization [10]. Section 2 describes the source localization techniques and its performance is demonstrated in Section 3 with synthetic data.

2 SOURCE LOCALIZATION TECHNIQUES

2.1 Acoustic model signal

An omnidirectional acoustic point source at location $\mathbf{s} = (s_x, s_y, s_z)^\dagger$ and a set of M microphones $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_M\}$ at location $\mathbf{m}_i = (m_{ix}, m_{iy}, m_{iz})^\dagger$ are considered. The bold letters denote matrices or vectors. Vectors are arranged column-wise, therefore the $(\cdot)^\dagger$ symbol represents the transposed vector. The acoustic pressure $\mathbf{p}$ recorded by a microphone in free field conditions is given by

$$\mathbf{p}(\mathbf{m}_i, \mathbf{s}, t) = \mathbf{S}(\mathbf{s}, t - \Delta t_i), \tag{1}$$

with $\mathbf{S}(\mathbf{s}, t)$ is the acoustic source signal. The Time of Flight (ToF) Δt_i between the source and the microphone is defined by Euclidean distance

$$\Delta t_i = \frac{1}{c_0} \|\mathbf{m}_i - \mathbf{s}\|_2, \tag{2}$$

where c_0 is the sound velocity and $\| \cdot \|_p$ is the p-norm of a vector or matrix.

2.2 Generalized Cross Correlation (GCC)

Classically, acoustic source localization or acoustic imaging is performed using the energy of the microphone array signal $\mathbf{E}_p$ defined for continuous signal by

$$\mathbf{E}_p(\mathbf{s}) = \int_{-\infty}^{+\infty} \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \mathbf{p}(\mathbf{m}_i, \mathbf{s}, t)\mathbf{p}(\mathbf{m}_j, \mathbf{s}, t)dt. \tag{3}$$

The auto-correlation does not bring information about the time delays therefore to improve the noise source map these terms in Eq. (3) can be subtracted

$$\mathbf{E}'_p(\mathbf{s}) = \int_{-\infty}^{+\infty} \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} \mathbf{p}(\mathbf{m}_i, \mathbf{s}, t)\mathbf{p}(\mathbf{m}_j, \mathbf{s}, t)dt - \frac{1}{M} \sum_{i=1}^{M} \int_{-\infty}^{+\infty} \mathbf{p}^2(\mathbf{m}_i, \mathbf{s}, t - \Delta t_i)dt. \tag{4}$$

An alternative formulation for the energy may be defined

$$\mathbf{E}'_p(\mathbf{s}) = \frac{1}{M^2 - M} \sum_{i=1}^{M} \sum_{j \neq i} (\mathbf{p}_i \star \mathbf{p}_j)(\Delta t_j - \Delta t_i), \tag{5}$$

where the product $(\mathbf{A} \star \mathbf{B})$ corresponds to the Cross-Correlation (CC) at time lag τ defined by

$$(\mathbf{A} \star \mathbf{B})(\tau) = \int_{-\infty}^{+\infty} \mathbf{A}(t)\mathbf{B}(t + \tau)dt. \tag{6}$$

To compute the CC, the inverse Fourier of the weighted cross-spectrum $\mathbf{C}_{AB}(\omega)$, denoted $(\circ)$, is introduced

$$(\mathbf{A} \circ \mathbf{B})(\tau) = \int_{-\infty}^{+\infty} \mathbf{W}(\omega)\mathbf{C}_{AB}(\omega) \exp(j\omega\tau)d\omega, \tag{7}$$

with

$$\mathbf{C}_{AB}(\omega) = \left(\int_{-\infty}^{+\infty} \mathbf{A}(t) \exp(-j\omega t)dt \right) \left(\int_{-\infty}^{+\infty} \mathbf{B}(t) \exp(-j\omega t)dt \right)^*. \tag{8}$$

The symbol $(\cdot)^*$ corresponds to the complex conjugate. The weighted function is used to estimate accurately the time delay between the microphones. The most common technique is the PHAse Transform (PHAT) [8] defined by

$$\mathbf{W}(\omega) = \frac{1}{\mid \mathbf{C}_{AB}(\omega) \mid}, \tag{9}$$

with $\mid \cdot \mid$ the absolute value of the signal. The goal is to whiten the cross-spectrum signal by removing the amplitude and keeping the phase information.

If only the microphone pairs are considered and if the source position is substituted by a set of $\mathcal{Q}$ scan points $\mathcal{Q} = \{\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_k\}$, the energy of the source map, commonly called GCC, can be written as

$$\text{GCC}(\mathbf{q}_k) = \frac{1}{M_p} \sum_{i=1}^{M} \sum_{j>i} \int_{-\infty}^{+\infty} \mathbf{W}_{ij}(\omega)\mathbf{C}_{ij}(\omega) \exp(j\omega\tau)d\omega, \tag{10}$$

with M_p the number of microphone pairs.

2.3 Inverse model with sparsity constraint

In perfect conditions, the energy source map should exhibit a point at source positions. However, the geometry of the microphone array and source-microphone distance induce an imperfect noise source map with spurious lobes due to the superimposition of hyperbolas. One approach to improve the source localization is to define a minimization problem J between the measured source map $\mathbf{y}$ (obtained with GCC) and a modeled source map $\hat{\mathbf{y}}$

$$J(\mathbf{x}) = \min_{\mathbf{x}} \rho(\mathbf{y}, \hat{\mathbf{y}}), \tag{11}$$

where ρ is a cost function and $\mathbf{x}$ is the unknown source amplitude vector. The modeled source map is defined by the following linear system

$$\hat{\mathbf{y}} = \mathbf{A}\mathbf{x}, \tag{12}$$

where $\mathbf{A}$ correspond to a propagation model matrix [10]. To design the propagation model matrix, an acoustic source at location $\mathbf{s}$ is considered. The Tof, denoted $ToF(\mathbf{s}, \mathbf{d}_{ij})$, between a source and a pair of microphones $\mathbf{d}_{ij}$ (at positions $\mathbf{m}_i$ and $\mathbf{m}_j$) is computed. Then the ToF between all the scan points and this pair of microphone $ToF(\mathbf{q}_k, \mathbf{d}_{ij})$ is computed. Finally, the difference between $ToF(\mathbf{s}, \mathbf{d}_{ij})$ and $ToF(\mathbf{q}_k, \mathbf{d}_{ij})$ is calculated. Small difference means that the scan point is potentially close to the source conversely large difference corresponds to scan far away from the source. The propagation matrix can be defined by

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}(s_1, q_1) & \mathbf{a}(s_1, q_2) & \cdots & \mathbf{a}(s_1, q_k) \\ \mathbf{a}(s_2, q_1) & \mathbf{a}(s_2, q_2) & \cdots & \mathbf{a}(s_2, q_k) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}(s_k, q_1) & \mathbf{a}(s_k, q_2) & \cdots & \mathbf{a}(s_k, q_k) \end{pmatrix}, \tag{13}$$

with

$$\mathbf{a}(s_l, q_k) = \frac{1}{M_p} \sum_{i=1}^{M} \sum_{j>i} \begin{cases} 1 & \text{if } \mid ToF(s_l, \mathbf{d}_{ij}) - ToF(q_k, \mathbf{d}_{ij}) \mid \leq \epsilon \quad \epsilon \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

The Tof is given by

$$ToF(\mathbf{q}_k, \mathbf{d}_{ij}) = \Delta t_i - \Delta t_j = \frac{1}{c_0}(\|\mathbf{m}_i - \mathbf{q}_k\|_2 - \|\mathbf{m}_j - \mathbf{q}_k\|_2). \tag{15}$$

Therefore, if ρ represents the Euclidean distance a linear least squares problem can be defined

$$J(\mathbf{x}) = \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2. \tag{16}$$

If the number of scan points is larger than the number of sources, we can use sparse method to solve the linear inverse problem which means minimize the l_0-norm of the $\mathbf{x}$ vector. However the minimization of the l_0-norm is very difficult in practice. Convex relaxation of the l_0-norm using the l_1-norm is preferred. Therefore, the linear inverse problem to solve is

$$J(\mathbf{x}) = \min_{\mathbf{x}} \left(\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right). \tag{17}$$

In the last years, several methods have been proposed to solve this inverse problem. In this study, the solutions given by the Orthogonal Matching Pursuit (OMP) [11] and the truncated Newton interior-point called Large Scale l_1 (LS1) [12] are compared. The inverse problem is solved in different way for each method. Both methods are compared in terms of source localization, level estimation and computational time with the classical GCC.

3 NUMERICAL STUDY OF THE PERFORMANCES OF THE SOURCE LOCALIZATION TECHNIQUES

3.1 Numerical set-up

To highlight the abilities of the proposed source localization techniques, synthetic data are used in free-field conditions, reflections can be generated by adding correlated sources (image-source method). The sound pressure recorded by the microphone array is computed using Eq. (1). The microphone array is a sphere composed of three circles. The radius of the larger circle is 0.25 m and it is 0.2 m for the smaller circles. The distance between the smaller circles and the main circle is 0.15 m. Each circle has five microphones, therefore the total number of microphone is $M = 15$ (see Figure 1). In practice, the goal is to set the microphone array at all the worker positions that is why the geometry is compact with a low number of microphones to limit the computational time [9]. Commonly, the noise generated in industrial halls is composed of many sources and reflections over a broad frequency range. Thus the source signal considered here is a Gaussian white noise with a zero mean value and a standard deviation equal to 1. With an acoustic pressure reference equal to $2e^{-5}$ Pa, the level of the source signal is 94 dB. The acoustic time signal is sampled at $44,000$ Hz and 2^{14} fast Fourier transform points are used to compute the cross-spectrum Eq. (8). The sound speed is set to $c_0 = 340$ m/s. The source locations are searched in a plane (including the source positions), called scan zone, at 1 m from the center of the spherical microphone array. The scan zone is a square with side equal to 1 m. The scan zone is sampled with 21 points in each direction which leads to a total number of scan points equal to $L = 441$ points and a space sampling of 5 cm. The GCC Eq. (10) is computed with all the microphone pairs $M_p = 105$ in this case ($M_p = (M \times (M-1))/2 = 105$). The PHAT weighted function is used to whiten the cross-spectrum according to Eq. (9). PHAT removes the magnitude of the cross-spectrum therefore the source levels can not be estimated. To recover the magnitude, the cross-spectrum is multiplied by its root mean square. The GCC Eq. (10) is computed for all the scan points, the result obtained is the noise source map. The origin of the source map corresponds to the center of the spherical microphone array (at 1 m). The noise source map is coded with colors where darker colors correspond to louder noise sources. The dynamic range of the noise source map is 16 dB and 1 dB corresponds to one coded color.

3.2 Case of three uncorrelated sources

First, the case of three uncorrelated point sources is studied. The source spacing is 0.2 m. The noise source maps are computed with the three source localization techniques discussed in section 2 and are shown in Figure 2. In all the noise source maps, GCC, OMP and LS1 are presented in the left, center and right, respectively. The GCC noise source map exhibits three spots at the source locations with large spurious lobes. In this case, an accurate detection of the source positions is difficult. OMP and LS1 provide noise source maps with only three spots at the source locations. Both methods remove the spurious lobes and provide a high resolution source map. GCC and OMP correctly estimate the magnitudes of the three sources whereas LS1 under-estimates by 1 dB the source level. This canonic case validates the three methods and shows the efficiency of OMP and LS1 algorithms at perfectly detecting the positions of three uncorrelated sources.

3.3 Case of three correlated sources

In some situations, the source signals generated by sources may be correlated such as in the case of ground or wall reflections. Now, the input signal is the same for the three sources and the configuration is kept similar. The noise source maps are shown in Figure 3. GCC exhibits a main lobe at the central source position with two smaller spots at the two other source positions.

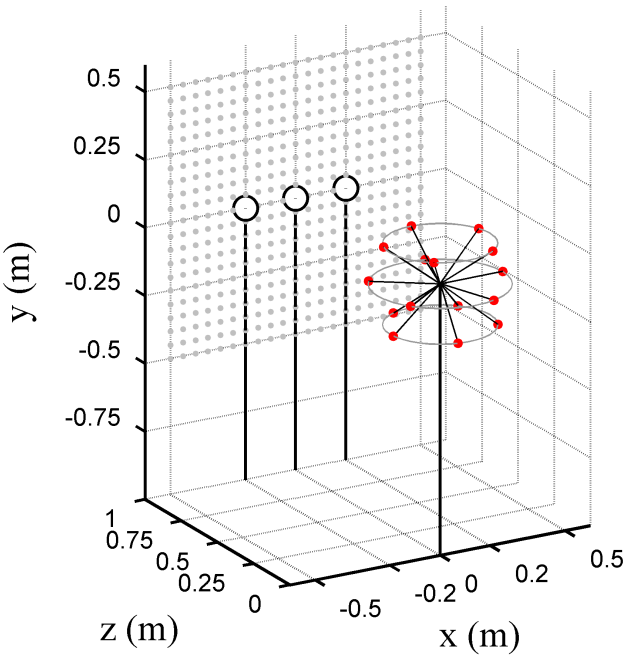

Figure 1: Spherical array composed of 15 microphones (red dots close closed to origin) in the case of three acoustic point sources (large dots at $z = 1$ m). The scan zone is represented by the gray dots (at $z = 1$ m).

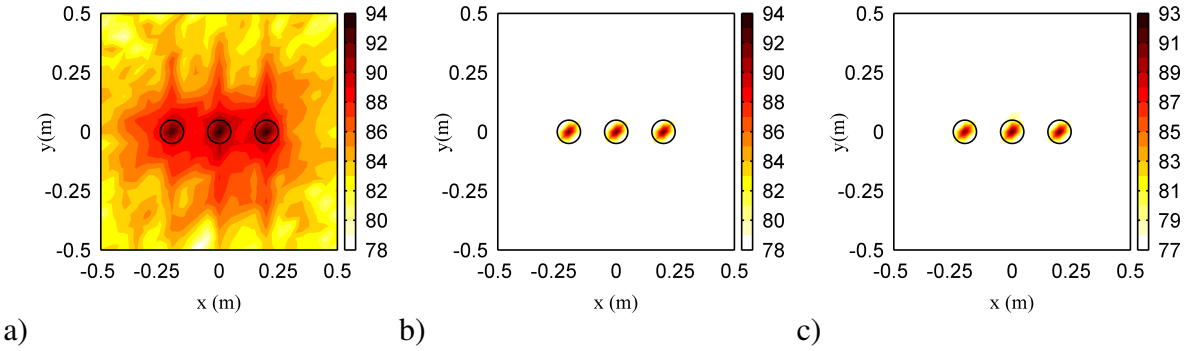

Figure 2: Noise source maps for three broadband uncorrelated sources, a) GCC, b) OMP and c) LS1. The circles are the true source positions. The colorbar is in dB.

In this case the pattern is clearly different from the previous configuration and it is more difficult to detect the three source positions. Due to the correlation between source signals, the side lobes merge to create a louder source at the origin. Both OMP and LS1 algorithms improve the source localization and each source is well detected. However, the source level is under-estimated for the sources on the side with each technique.

3.4 Case of an extended source

In the previous configuration, the noise source was a point source. However, in industrial situations, noise sources are often extended. Therefore an extended source composed of 41 point sources from -0.2 m to 0.2 m is computed (source separation of 1 cm i.e five sources by scan zone sample). The noise source maps are shown in Figure 4. GCC shows an extended source with large side lobes which may impair the localization of sources with a lower source level.

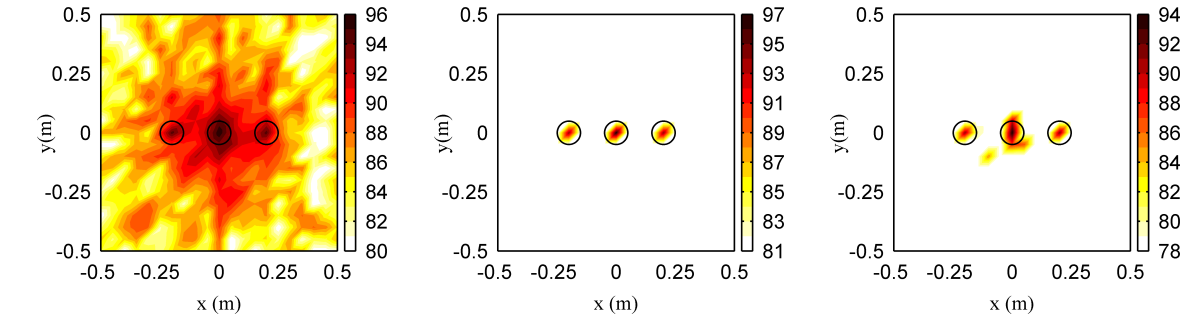

Figure 3: Noise source maps for three broadband correlated sources, a) GCC, b) OMP and c) LS1. The circles are the true source positions. The colorbar is in dB.

OMP improves the noise source map, however several spots, with low levels which do not correspond to source locations, are present. The best noise source map is obtained using LS1 where the source position is well detected without any spurious lobes. The source level estimated by all the methods is close to 100 dB. The number of sources by scan point is equal to 5, thus if the contribution of each source is summed, we can define an overall source level by scan point ($10\log_{10}((5\times1)/4e^{-10}) = 101$ dB). Therefore, LS1 correctly detects the source positions and moreover is able to estimate the source level with a small error.

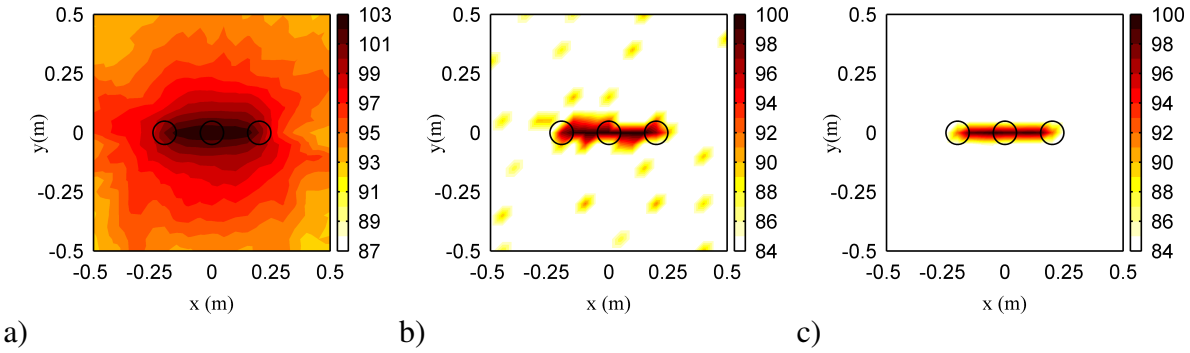

Figure 4: Noise source maps for an extended source, a) GCC, b) OMP and c) LS1. The circles are the left, center and right limits of the extended source. The colorbar is in dB.

3.5 Case of three sources with unequal magnitudes

In this section, the ability to detect sources with unequal magnitudes is investigated. The configuration is similar that in the section 3.2 where three uncorrelated sources are 1 m from the array. However, the magnitude is decreased by 3 dB and 6 dB for the left and right sources, respectively. The noise source maps are shown in Figure 5.a-c. All the source localization techniques correctly detect the source positions but the best results are obtained with OMP and LS1. To gain insight into the noise source maps, slices at $y = 0$ m are plotted in Figure 5.d-f. These figures clearly show the high resolution ability of OMP and LS1. The best sound level estimation is given by OMP method whereas GCC and LS1 under-estimate the sound level by 1 dB.

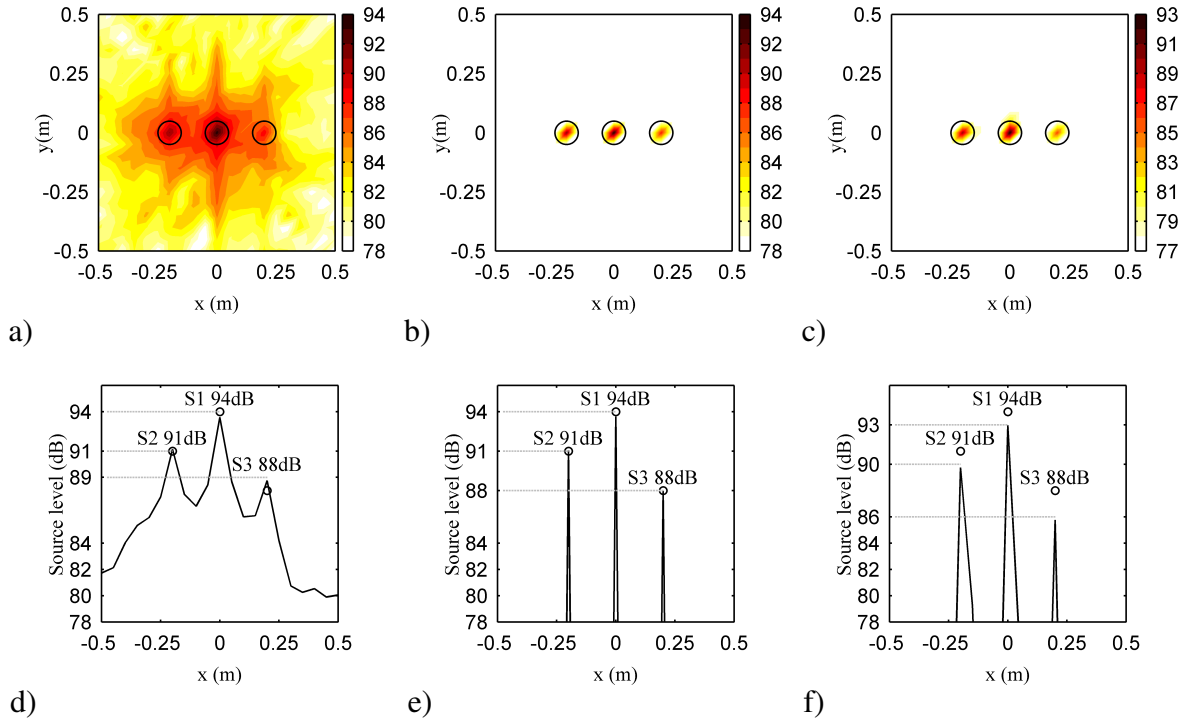

Figure 5: Noise source maps for three broadband uncorrelated sources with unequal magnitudes, a) GCC, b) OMP and c) LS1 and slices at $y = 0$ m d) GCC, e) OMP and f) LS1. The circles are the true source positions and sound levels. The colorbar is in dB.

3.6 Computation time

The previous sections compared the efficiency of the source localization techniques to detect source levels and positions. In an industrial context, all the worker positions have to be tested therefore the computational time of the methods should remain reasonable. The technique proposed by Noël *et al.* requires several hours with 648 scan points [9].

The computational time of the techniques is compared for several numbers of scan points. The time is given by the tic-toc function of Matlab R2014a. A dual core processor at 3.33 GHz is used with 4 Go of Ram. The time for building the propagation matrix $\mathbf{A}$ (Eq. 13), for solving the problem using OMP and LS1 is provided for comparison. The construction of matrix $\mathbf{A}$ and the implementation of OMP are custom-made codes whereas LS1 is based on the toolbox Large-Scale l_1-Regularized Least Squares Problems [13]. The computation time of GCC is very low and mainly dependent on the number of microphone pairs and is therefore not compared with the other methods. The total number of scan points ranges from $L = 361$ points (19×19 grid size) up to $L = 3025$ points (55×55 grid size). The result is shown in Figure 6. From the trend of the curves, it is possible to define a power law depending on the number of scan points. OMP and LS1 times increase with the square of the number of scan points. OMP is the fastest method and the time required to build matrix $\mathbf{A}$ is over a minute for a number of scan points larger than 3000. Finally, even with a large number of scan points the computation time is still reasonable (less than 2 minutes) and can be applied at different workstations.

4 CONCLUSION

In this study, various source localization techniques in time domain for broadband acoustic sources have been presented. The Generalized Cross Correlation (GCC) provides a coarse noise source map. Then a linear inverse problem is defined and solved with a sparsity constraint.

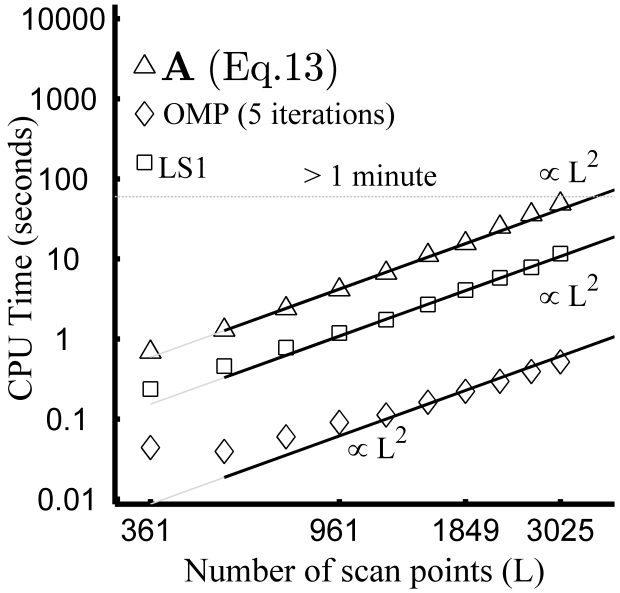

Figure 6: Computation time of OMP and LS1 versus the total number of scan points (using Matlab R2014a, running on a dual core processor at 3.33 GHz and 4 Go of Ram).

Synthetic data generated for different source configurations are used to highlight the abilities of such techniques. As compared to GCC, sparsity constraint methods provide a high resolution imaging with a correct estimation of the source levels. Moreover, the computation time is reasonable for industrial applications. However, these techniques need user-defined parameters. In a subsequent work, a parametric study will be carried out to define the optimal choice of these parameters. Moreover the spherical microphone array allows for recording the acoustic pressure coming from all the directions, thus the next step is to develop a scan zone in spherical coordinates. Finally an experiment will be carried out in hemi-anechoic and reverberant rooms to assess the proposed techniques.

REFERENCES

[1] J. Chatillon. Influence of source directivity on noise levels in industrial halls: Simulation and experiments. *Applied Acoustics*, 68:682–698, 2007.

[2] M. Brandstein and D. Ward. *Microphone Arrays*. Signal Processing Techniques and Applications. Springer-Verlag Berlin Heidelberg, 2001.

[3] U. Michel. History of acoustic beamforming. In *Berlin Beamforming Conference (BEBEC)*, 2006.

[4] P. Sijtsma. Clean based on spatial source coherence. *International Journal of Aeroacoustic*, 6:357–374, 2007.

[5] T. F. Brooks and W. M. Humphreys. A deconvolution approach for the mapping of acoustic sources (damas) determined from phased microphone arrays. *Journal of Sound and Vibration*, 294:856–879, 2006.

[6] T. Padois P-A. Gauthier and A. Berry. Inverse problem with beamforming regularization matrix applied to sound source localization in closed wind-tunnel using microphone array. *Journal of Sound and Vibration*, 333:6858–6868, 2014.

[7] T. Yardibi J. Li P. Stoica and L. N. Cattafesta III. Sparsity constrained deconvolution approaches for acoustic source mapping. *Journal of the Acoustical Society of America*, 123:2631–2642, 2008.

[8] C. H. Knapp and G. C. Carter. The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust. Speech, Signal Process.*, 24:320–327, 1976.

[9] C. Noel V. Planeau and D. Habault. A new temporal method for the identification of source directions in a reverberant hall. *Journal of Sound and Vibration*, 296:518–538, 2006.

[10] J. Velasco D. Pizarro and J. Macias-Guarasa. Source localization with acoustic sensor arrays using generative model based fitting with sparse constraints. *Sensors*, 12:13781–13812, 2012.

[11] A. Peillot F. Ollivier G. Chardon and L. Daudet. Localization and identification of sound using compressive sampling techniques. *18^{th} International Congress on Sound and Vibration (ICSV), Rio de Janeiro, Brazil, 10-14 July*, 2011.

[12] S. Kim K. Koh M. Lustig S. Boyd and D. Gorinevsky. An interior-point method for large-scale l1-regularized least squares. *IEEE J. Sel. Top. Signal Process.*, 1:606–617, 2007.

[13] K. Koh S. Kim and S. Boyd. A matlab solver for large-scale l1-regularized least squares problems. 2012.