

LOCALIZATION AND IDENTIFICATION OF SOUND SOURCES USING “COMPRESSIVE SAMPLING” TECHNIQUES

Antoine Peillot, François Ollivier

Institut d'Alembert, UPMC Université Paris 06, UMR 7190, F-75005, Paris, France, e-mail: antoine.peillot@upmc.fr

Gilles Chardon, Laurent Daudet

Institut Langevin, ESPCI ParisTech, UPMC Univ Paris 06, Univ Paris Diderot 07, UMR 7587, F-75005, Paris, France

“Compressive sampling” (CS) is a new signal acquisition strategy that intends to reduce significantly the amount of recorded data by picking only a limited number of samples. CS theory asserts that one can reconstruct a given signal from a few randomly distributed samples if only the signal is sparse in a proper basis. CS ensures a minimum loss of information but requires, for the reconstruction of the signal, the use of dedicated sparsity-promoting algorithms. In this paper, CS is applied to the source localization problem using an array of randomly distributed microphones. In this case, the signal of interest is sparse in the spatial domain, i.e a few positions in space contain sources. We focus on the near-field beamforming where the array of sensors is sensitive to the sources directivity. The localization method is extended to complex sources and we attempt to identify them in terms of multipoles. Numerical simulations and experimental results prove this sparsity-promoting method to be powerful for source localization. However the identification step, quite successful on ideal data, is not sufficiently robust when applied to experimental data and need further investigations.

1. Introduction

The present work investigates the use of compressive sampling techniques for acoustic sources localization and identification. The localization problem is most commonly solved by the standard beamforming technique [1] (SBF) using an array of microphones. Its main advantages are flexibility and straightforward implementation. It involves spectral processing and uses cross-spectral matrices (CSM) and phase shifting in order to provide maps of the sound intensity in the so called reconstruction space. However this standard method presents drawbacks. It suffers from a resolution limit which is proportional to the wave length to array dimension ratio. Moreover the directivity pattern of the array may present grating lobes if the array step is too high with respect to the wave length. This may reveal non-existent sources. Using a large number of sensors may allow overcoming these limitations but would induce complex issues of synchronization, calibration, A/D conversion, and total data throughput. Other techniques achieving super resolution have been derived such as MUSIC [2], a subspace-based method, or CAPON method [3]. However these methods fail when correlated sources

are to be localized and are not made free from using a large amount of sensors. Compressive sampling [4] (CS) techniques, recently emerging from applied mathematics, intend to reduce significantly the amount of data throughput when recording time signals or spatial fields without loss of information, thus allowing source reconstruction from a limited data set. However, CS requires on one hand that the recorded signal or field be sparse and on the other hand that specific reconstruction algorithms be used for the resolution of the source localization inverse problem. In the first section we introduce the principles of compressive sampling. The next section outlines the implementation of CS for source localization; numerical simulations and experimental results are presented. Afterwards we focus on the case of near-field where it can be expected to get a more precise description of the involved sources. Therefore In the final section, an additional decomposition domain is considered assuming that the sources can be described on a spherical harmonic basis. Thus the CS method is improved so as to allow source identification in terms of multi-poles; numerical results are presented.

2. Sparsity and Compressive Sampling

2.1 Principles

In the standard way, recording a signal consists in a uniform sampling at the Nyquist rate. This involves a large amount of data throughput when the signal to be acquired holds high frequency components. Often following acquisition, the signal undergoes a compression step so as to discard the redundant information. This step is particularly efficient when the considered signal is sparse in a basis to be identified. Then it can be represented with a small number of coefficients without information loss. Compressive Sampling (CS) is a new data acquisition technique that aims to compress the signal directly at the recording stage by picking only a limited amount of samples. Thus, provided that the signal of interest is sparse, CS theoretically allows it to be under sampled far below the Nyquist rate. In the ideal case, this sparsity can be expressed in an orthogonal basis; more generally a dictionary made of redundant vectors can be used. For example, let \mathbf{s} be a signal of dimension N , sparse in a dictionary $\mathcal{D} = [\mathbf{d}_1 \dots \mathbf{d}_n \dots \mathbf{d}_N]$. \mathbf{s} can be written:

$$\mathbf{s} = \sum_{n=1}^N x_n \mathbf{d}_n \quad (1)$$

where $\mathbf{x} = [x_1 \dots x_n \dots x_N]^T$ is the representation of \mathbf{s} in the dictionary \mathcal{D} . The signal \mathbf{s} is said to be K -sparse when just K coefficients in \mathbf{x} are non-zero ($K \ll N$), i.e \mathbf{s} is a linear combination of K components of \mathcal{D} . The aim of CS is to reconstruct the sparse vector \mathbf{x} from a few measurements $\mathbf{y} = [y_1 \dots y_m \dots y_M]$, ($M \ll N$). Let Φ be the transfer matrix from \mathbf{s} to \mathbf{y} ($\mathbf{y} = \Phi \mathbf{s}$), of dimension $M \times N$, and Θ be the transfer matrix between the measurement vector \mathbf{y} and the source vector \mathbf{x} . In order to identify \mathbf{x} , the problem to be inverted writes:

$$\mathbf{y} = \Theta \mathbf{x} \quad \text{where} \quad \Theta = \Phi \mathcal{D} \quad (2)$$

For a specific application, a crucial point is the choice of a relevant reconstruction dictionary \mathcal{D} . Besides, it has been shown that if Θ is incoherent, the vector \mathbf{x} can be reconstructed using only $M = O(K \log(N/K))$ measurements [4]. It is also known that this incoherence property of Θ can be guaranteed simply by choosing a random distribution of the measurement set. Finally, since the system in Eq.2 is highly under-determined, dedicated algorithms have to be used to recover the sparsest solution. The design of such sparsity-promoting algorithms is a huge emerging research field. The next section describes such algorithms that were used for the present work.

2.2 Sparsity-promoting algorithms

A first class of algorithms formulates the sparse decomposition problem as the minimization of the ℓ_0 -norm of the vector \mathbf{x} which yields :

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad s.t. \quad \mathbf{y} = \Theta \mathbf{x} \quad (3)$$

where $\|\mathbf{x}\|_0$ is the number of non-zero components of \mathbf{x} . However, the non-convexity of the ℓ_0 -norm, and the approximate nature of the representation in most practical cases, make it a difficult optimization problem. In the general case indeed, exhaustive search is excluded. Yet, recent advances in signal processing showed that this problem may advantageously be replaced by the following statement involving the ℓ_1 -norm known as the Basis Pursuit Denoising framework [5]:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad s.t. \quad \|\mathbf{y} - \Theta \mathbf{x}\|_2^2 \leq \varepsilon \quad (4)$$

where $\|\mathbf{x}\|_1 = \sum_{k=1}^K |x_k|$ and ε is a specific error rate.

In a second class of algorithms called the “greedy” algorithms, one finds Orthogonal Matching Pursuit (OMP) [6]. It addresses the sparse optimization problem by a recursive process. At each iteration, a component of the signal \mathbf{x} is selected by correlation of the measurement \mathbf{y} with the columns of Θ . Its contribution is extracted from \mathbf{y} by means of an orthogonal projection providing a residual. The process is iterated until the desired number of selected components is reached. This number being the sparsity of the signal to reconstruct. The precise implementation of OMP to our case is presented section 3.2.

3. Methods for sources localization

3.1 Problem formulation and standard beamforming

Let \mathbf{y} be the vector of stationary pressure signals received by an array of M sensors ($m = 1 \dots M$) from an unknown number of harmonic sources. Let the vector \mathbf{x} describe reconstruction domain supposed to contain the sources. \mathbf{x} is composed of N grid points x_n ($n = 1 \dots N$). \mathbf{y} writes:

$$\mathbf{y} = \mathbf{A} \mathbf{x} + \varepsilon \quad (5)$$

$\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_n \dots \mathbf{a}_N]$ is the steering matrix. ε depicts noise. Standard near-field beamforming considers sources as monopoles *i.e* omnidirectional sources. The steering vector from the n^{th} target grid point to every sensor position thus writes:

$$\mathbf{a}_n = \left(\frac{e^{-jk|r_{1n}|}}{|r_{1n}|} \dots \frac{e^{-jk|r_{mn}|}}{|r_{mn}|} \dots \frac{e^{-jk|r_{Mn}|}}{|r_{Mn}|} \right)^T, \quad n = 1 \dots N \quad (6)$$

The use of the cross spectral matrix (CSM), denoted \mathbf{R} in the following, has proved to make more robust the beamforming technique [1]. At a given frequency f , \mathbf{R} writes :

$$\mathbf{R}(f) = E[\mathbf{y}(f)\mathbf{y}^H(f)] \quad (7)$$

$E[\]$ is the expectation operator and the upper script H denotes the conjugate transpose.

SBF produces a map of the acoustic power over the reconstruction domain. This power map \mathbf{P} is derived as:

$$\mathbf{P} = \text{diag}[\mathbf{A}^H \mathbf{R} \mathbf{A}] \quad (8)$$

The sources are located at the maxima of \mathbf{P} . This simple technique has two essential limitations. The resolution of the technique is dependant on the array dimensions with respect to the wavelength (larger the array better the resolution). In parallel spatial aliasing occurs when the array step is lower

than half the wave length, thus imposing a large number of microphones to be used in order to avoid the location of ghost sources. Note that the sensors distribution must be carefully designed since it sets the level of secondary lobes in the radiation pattern which drive the acoustic maps dynamics, that is to say the ability of the system to resolve sources with different levels.

An attempt to improve SBF proceeds by solving an ℓ_2 -norm minimization problem. The solution is derived via the classical Tikhonov regularization method. But this finally does not improve SBF, exhibiting on the contrary a poorer dynamics. Going on with optimization and regularization, a strategy based on the ℓ_1 -norm minimization has been chosen instead in what is called the “Generalized inverse beamforming” [7]. It assumes that the number of sources is small that is to say that the source distribution \mathbf{x} is sparse. This formulation provides satisfying results but involves relatively computational resources.

With another approach, the well known MUSIC and CAPON methods provide high resolution maps of the acoustic power. These methods show to be successful when the signal to noise ratio is high, the involved sources are uncorrelated and the number of sensors is sufficient.

3.2 The sparse way

In this section we propose the use of sparsity-promoting methods, that can be taken advantage of in order to derive compressive sampling and reduce the number of required sensors and yet be able to achieve high resolution source localization.

A preliminary process consists in computing the eigenvalue decomposition of the matrix \mathbf{R} in order to extract the signal subspace. This decomposition writes :

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \quad (9)$$

$\mathbf{\Lambda}$ and \mathbf{U} are the diagonal matrix of eigenvalues of \mathbf{R} and the associated matrix of eigenvectors respectively. In the case of (coherent ?) harmonic sources, it should be noted that the signal subspace is made of a sole eigenvector \mathbf{u}_s associated with the maximum eigenvalue λ_s . All the sources share the same eigenvector and the relevant signal \mathbf{y} is given by the following projection:

$$\mathbf{y} = \sqrt{\lambda_s}\mathbf{u}_s \quad (10)$$

We now seek to isolate the different sources in the spatial domain described by the target vector \mathbf{x} using a compressive sampling method. The relation between the target domain and the signal \mathbf{y} expresses :

$$\mathbf{y} = \mathbf{\Theta}\mathbf{x} \quad (11)$$

According to the principles of CS listed in section 2.1, the matrix $\mathbf{\Theta}$ decomposes into a transfer matrix $\mathbf{\Phi}$ and a dictionary \mathcal{D} defining the sparsity domain of the sources. In our case this domain is the spatial domain : a few positions in space contain the sources. There is no need of a particular dictionary and \mathcal{D} will be chosen as the identity operator. As for the transfer matrix $\mathbf{\Phi}$ it simply equals the propagation operator or matrix of steering vectors \mathbf{A} defined in Eq.6 : $\mathbf{\Theta} = \mathbf{A}$.

In solving this problem of source localization, the OMP algorithm described in section 2.2 proves to be the more efficient and presents lower computational costs when compared to the ℓ_1 -norm minimization. The algorithm is implemented as follows:

1. Initialize the residual $\mathbf{r}_0 = \mathbf{y}$ and set the iteration counter l to 1
2. Select the most correlated target point $l : l = \arg \max_i |\mathbf{a}_i^H \mathbf{r}_l|$
3. Remove the orthogonal projection of \mathbf{r}_l in the space spanned by $\{\mathbf{a}_j, j = 1 \dots l\}$ to get \mathbf{r}_{l+1}

4. Iterate $l = l + 1$ until the number L of expected sources is reached

Finally, the estimated sources position are the selected target points x_l , $l = 1 \dots L$ associated with the steering vectors \mathbf{a}_l .

Let us stress out that the algorithm implies the choice of the parameter L which is the desired sparsity of the signal to reconstruct i.e in our case, the actual number of sources. Yet this parameter is unknown *a priori*. However, monitoring the ℓ_2 -norm of the residue $\|\mathbf{r}_l\|_2$ at each iteration allows stopping the algorithm when a significant drop is observed. When all the sources have been selected indeed, $\|\mathbf{r}_l\|_2$ vanishes to zero.

3.3 Simulation and experimental results

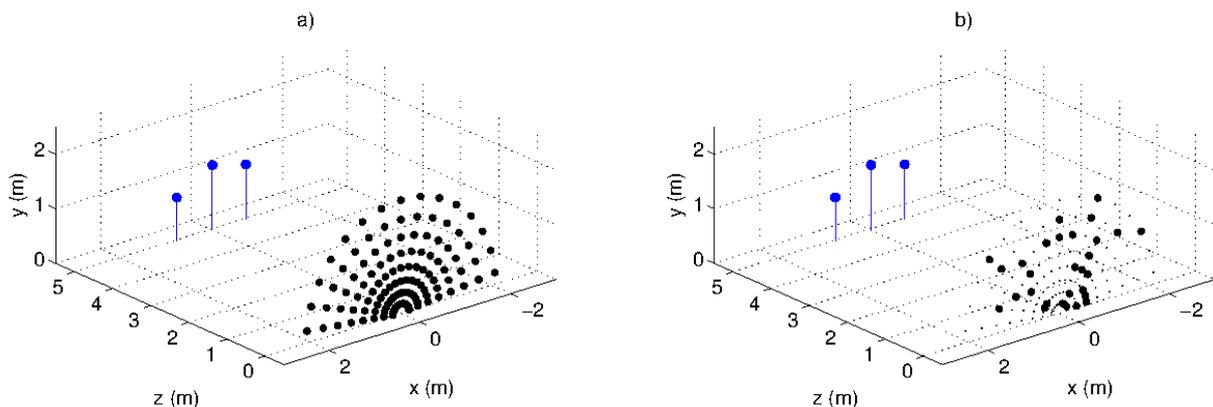


Figure 1. Microphone array (black dots) at 5m from the source plane (blue dots). a) Full set of 120 microphones. b) CS antenna of 30 microphones - random draw among the complete set

This section presents results of source localisation using compressive sampling compared to SBF using the arrays drawn on Fig. 1. The set-up describes as follows. Three sources are located in a vertical plane 5m away from the vertical plane array of 120 microphones (1-a). The central loudspeaker is a B&K type 4295 omnidirectional source. The two lateral sources baffled loudspeakers with a 10cm radius membrane. The whole set-up was located in an anechoic chamber providing free field conditions. The signal driving the sources is a 5s duration white noise. A cross power spectral density is processed averaging 190 FFT blocks on each sensor signal. This provides spectral data from which the CSM matrix \mathbf{R} can be computed at any frequency of interest. On Fig.2, we compare the results of source localization using SBF (background grey scale map with a 15dB dynamics) and CS using the OMP algorithm (red crosses). The true sources location is indicated by the blue circles. The first row exhibits numerical simulations, the second row, the results obtained from experimental measurements and processed using the complete 120 microphones array. For the third row a random set of 30 microphones drawn among the complete set (see Fig.1-b) was used. Three frequencies are studied according to columns : 1kHz, 2.5kHz and 6kHz.

Concerning the SBF maps obtained using the complete array, at 1kHz the intrinsic resolution is about 40cm and the sources being 75cm apart are difficult to separate. At the 2.5kHz medium frequency, the resolution allows to distinguish and locate the three sources quite accurately. At 6 kHz, side lobes arise rendering a noisy sound map with poor dynamics and low level ghost sources, even if the actual sources are well identified. For the three frequencies, when using the strongly under-sampled 30 microphones CS array (third row), the resulting high rate of grating lobes makes source localization impossible.

Whatever the frequency and the number of microphones used, the sparsity-promoting method provides a location of the three sources studied with less than a 10cm error (the target grid step is

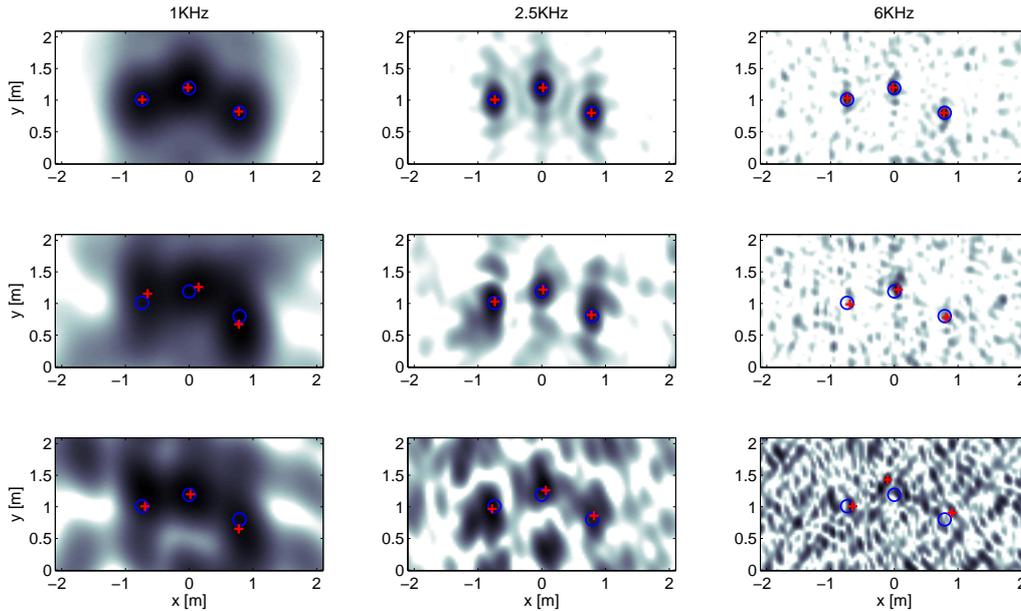


Figure 2. Comparison of source localization processes at three frequencies : Background SBF sound maps (15dB dynamics grey scale) - Sparse reconstruction using OMP (red crosses) - Actual sources location (blue circles) - a) Numerical simulation b) Experimental results using the complete 120 microphones array c) Experimental results using a CS 30 microphones array

4.2cm). The poorer results occur at 1KHz with the complete array and at 6KHz using the CS array. Note that reducing the number of microphones (1kHz - 30 microphones) reduces the error. Further experiments have to be led in order to understand the relevant parameters of these results. At 6 kHz the CS array is probably too highly under-sampled to keep a good accuracy. Nevertheless it shows an acceptable accuracy.

4. Source identification

4.1 Multipoles description

In the near-field, a large array of sensors can be sensitive to the directivity of the sources, and it can be expected to identify their nature more precisely. However, for complex sources, an other more realistic model has to be defined. As suggested in the literature [8] we propose to use a multipole radiation model. Deriving the wave equation in spherical coordinates, it can be shown that at any point in the free field, the sound pressure radiated from a source can be described as a sums of multipoles. This writes:

$$p(r, \theta, \phi, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n C_{mn}(k) h_n^{(1)}(kr) Y_n^m(\theta, \phi) \quad (12)$$

The $Y_n^m(\theta, \phi)$ denotes the spherical harmonic of order n and degree m . $h_n^{(1)}(kr)$ is the associated outgoing radial Hankel function of the first kind of order n . The coefficient $C_{mn}(k)$ is the corresponding multipole component strength. This approach has already been used to extend the SBF method to multipolar sources [9]. However it has not shown to be successful in terms of dynamics. In the following we propose an other method based on sparsity.

4.2 Source identification algorithm

The algorithm described in section 3.2 is modified to take into account the multipole components. For a target point l , at position r_l , the radiation pattern of degree m and order n is denoted $J_{lnm} = h_n^{(1)}(kr_l)Y_n^m(r_l)$. The multipole component strength C_{lmn} is to be identified. When the spherical harmonics identification scheme is restricted to order N , the multipoles sub-space associated to the target point l expresses $\mathbf{A}_l = [\mathbf{J}_{l11} \dots \mathbf{J}_{l1m} \dots \mathbf{J}_{lNm}]$. The previous OMP algorithm is improved so as to integrate orthogonalized multipole functions. The new ‘‘Group OMP’’ algorithm processes according to the following steps :

- Initialize the residual $\mathbf{r}_0 = \mathbf{y}$ and set the iteration counter $l = 1$
- Select the target point l which verify $l = \arg \max_i \|\mathbf{A}_i^H \mathbf{r}_1\|$
- Remove the orthogonal projection of \mathbf{r}_1 on every elements of the sub-space \mathbf{A}_l
- Iterate $l = l + 1$ until the number L of sources is reached

Finally, the estimated sources position are the selected target points $l = 1 \dots L$. The contribution of each multipole, element of the sub-spaces \mathbf{A}_l ($l = 1 \dots L$) are listed in the vector $\mathbf{x} = \mathbf{A}_l^H \mathbf{y}$.

4.3 Numerical simulations

For our numerical experiment, we use the following directivity patterns: a monopole denoted M, two dipoles: D1 in the x direction (see Fig. 1 for the coordinate system), D2 in the y direction, and a quadripole Q1 of degree 1 (order 2) in the xy plane. For the simulation the sources consists of two point sources with complex radiation pattern emitting harmonically at 1KHz. The first source is made of a dipole D1 with a 0dB amplitude, and a dipole D2 with a -1dB amplitude. The second source is made of a monopole M with a -3dB amplitude and a quadripole Q1 with a -1dB amplitude. Fig. 3 exhibits the sources localization with their respective identification terms. The localization is very accurate for both sources and the identification is correct as well showing a dynamic over 20dB between the true components and the fake ones.

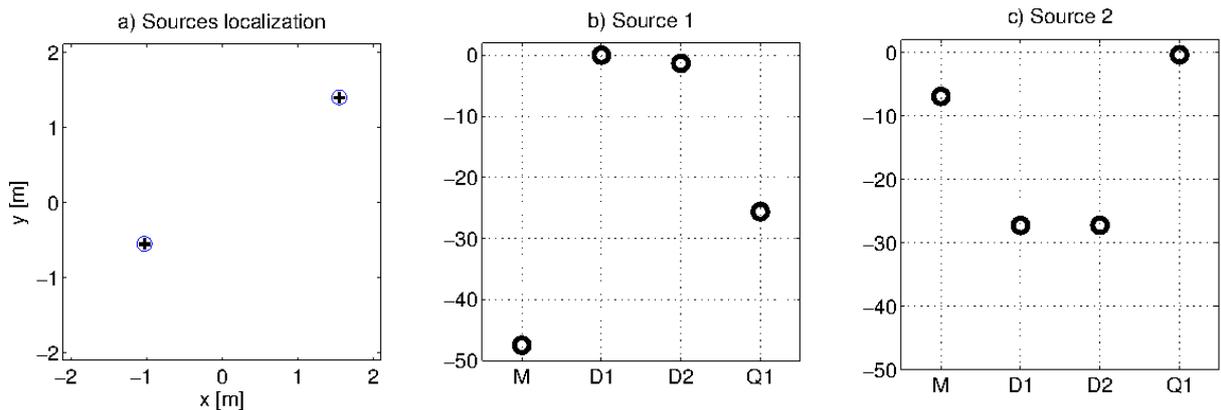


Figure 3. Simulation at 1KHz a). Localization map (blue circles pointing at the actual sources positions) b). Multi-pole coefficient estimation for the 1st source at [-1.03,-0.55] c). Multi-pole coefficient estimation for the 2nd source at [1.55,1.40]

However, the ability of the algorithm to identify accurately a source is strongly dependant on the grid fineness. If the grid is too rough, no grid point fits the actual sources position and the identification fails. Since the computational cost increases significantly with the number of points

in the target domain, the grid fineness has to be limited. The idea is thus to refine the grid around the source position estimated at the first iteration, and refine the grid in that way at each iteration until some stabilization arises. Moreover the “Group OMP” algorithm appears to be very sensitive to noise. The first attempts to apply it to the previous measurements were not promising.

5. Conclusion

In our objective of source localization, preliminary results have shown that using sparsity-promoting algorithm associated to compressive sampling with a low number of randomly distributed microphones, good results were obtained. Further leads are to be explored in order to evaluate the limitations of the method and appreciate the results with smaller-aperture arrays. Concerning the source identification scheme, the adopted strategy produces good results only for ideal cases. Therefore alternative algorithms have to be investigated to make the estimation more robust to noise. A trail could be a “grouped sparsity” algorithm using jointly ℓ_1 - and ℓ_2 - norm minimisations.

6. Acknowledgments

This work was supported by the French Agence Nationale de la Recherche (ANR), project ECHANGE (ANR-08-EMER-006). We wish to thank the “Laboratoire National d’Essais” for the lending of their anechoic facility.

REFERENCES

- ¹ T.J. Mueller (Ed.) (Chapter 2, Chapter Author: R.P. Dougherty), *Aeroacoustics Measurements*, Springer, New York (2002).
- ² R.O. Schmidt, Multiple emitter location and signal parameter estimation, *IEEE Transactions on Antennas and Propagation*, **34** 276–280 (1986).
- ³ Z. Wang, J. Li, P. Stoica, T. Nishida and M. Sheplak, Constant-beamwidth and constant-powerwidth wideband robust Capon beamformers for acoustic imaging, *Journal of the Acoustical Society of America* **116**, 1621–1631 (2004).
- ⁴ E.J. Candès and M.B. Wakin, An introduction to compressive sampling, *IEEE Signal Processing Magazine* **21**, 21–30 (2008).
- ⁵ S. Chen, D.L. Donoho and M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM Journal on Scientific Computing* **20**, 33-61 (1999).
- ⁶ Y.C. Pati, R. Rezaifar, Y.C.P.R. Rezaifar and P.S. Krishnaprasad, Orthogonal Matching Pursuit: Recursive function approximation with applications to wavelet decomposition, *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers*, California, 40–44 (1993).
- ⁷ P.A.G. Zavala, W. De Roeck, K. Janssens, J.R.F. Arruda, P. Sas and W. Desmet, Generalized inverse beamforming with optimized regularization strategy, *Mechanical Systems and Signal Processing*, doi:10.1019/j.ymsp.2010.09.012 (2010).
- ⁸ E. Williams, *Fourier Acoustics*, Academic Press, London(1999).
- ⁹ C. Bouchard, D.I. Havelock and M. Bouchard, Beamforming with microphone arrys for directional sources, *Journal of the Acoustical Society of America* **125**, 2098–2104 (2009).