18TH INTERNATIONAL CONGRESS ON SOUND & VIBRATION
ICSV18
RIO DE JANEIRO
BRAZIL
10–14 JULY 2011

SENSOR ARRAY OPTIMIZATION FOR SOURCES SEPARATION AND DETECTION IN THE AT-WORST DETERMINED CASE

Patrick Marmaroli, Xavier Falourd, Hervé Lissek

Laboratory of Electromagnetism and Acoustic of EPFL (1015 Lausanne, Switzerland), e-mail : patrick.marmaroli@epfl.ch

Most of the number of sources estimation techniques use the well-known signal-subspace approach in which the number of dominant sources is deduced regarding the multiplicity of the lowest eigenvalues of the correlation matrix. In the at-worst determined case (number of microphones just equals the maximal number of possible radiating sources) such methods are inoperative because the noise subspace could be inexistant. However, a well chosen sensor array geometry permits to achieve source detection using eigenvalues above conditions to some a priori knowledge on the sources. This paper explores some relation between geometry and eigenvalues in order to achieve optimal sources detection and separation. This study yields analytical formulations of both optimisation problem by working on the simple case of two uncorrelated harmonic sources. Theoretical and experimental measurements are presented and discussed.

1. Introduction

The satisfactory of most of the sound sources separation, tracking or localisation algorithms is related to the knowledge of the number of sources to characterize. That's why a large area in signal processing research concerns the number of source estimation. Most of such algorithms includes Support Vector Machine [1], Information Theoretic Criteria [2, 3, 4], Minimum Eigenvalue Varied Rate Criteria [5], Beam Eigenvalue Approaches [6] to list only a few. The main general idea is to study the rank of the covariance matrix of the observations in the light of its eigenvalues which can theoretically be separated in two groups : the eigenvalues which belong to the signal subspace (as many as the number of sources) and the eigenvalues which belong to the noise subspace (all equal and related to the noise power). This is the well known subspace approach theory described in [7]. However, one common necessary condition to all above-mentioned methods is that the number of sensors had to be larger than the number of sources in the sense that a comparison had to be made between signal subspace and noise subspace. In this paper, we are interested in using the properties of subspace approach theory in order to optimize the microphone array geometry in the at-worst determined case, i.e. when the number of microphones at-worst equals the maximal number of sources. In such a case, the noise subspace may not be available so a comparison-based method is not judicious. As eigenvalues of the correlation matrix are closely related to microphone locations, our motivation is to find analytical formulations for optimize them in order to achieve easier detection/separation methods with a low number of sensors.

After a general description of the SVD-based method in Section 2, the optimization problem is described in Section 3. Found analytical formulations are checked by experiments and applied to different contexts in Section 4. A final discussion is given in Section 5.

2. The subspace approach for estimate the number of sources

Consider an array of M omnidirectional sensors m_i with same impulse responses at locations $\mathbf{x}_{mi} \in \mathcal{R}^2$, $i \in [1, 2, ..., M]$ and let N_{max} be the number of maximal mutually independent and isotropic sources which can radiate across the array. Each source is characterized by its location $\mathbf{x}_{sj} \in \mathcal{R}^2$, its wavelength λ_j and its amplitude β_j, $j \in$ [1,...,N_{max}]. Assume that $\mathbf{x}_{sj}$ and λ_j are known for all j and the amplitude β_j variates according to an unknown and different law for each source. Thus, the number N of radiating sources taken at a given instant is unknown and N can fluctuate between 0 and N_{max}.

A simplistic but often sufficient model of sensor array signal processing consider the observations $\mathbf{X} \in \mathbb{C}^{M\times 1}$ as linear combinations of complex sources signal $S \in \mathbb{C}^{N\times 1}$ attenuated and delayed in time through a complex mixing matrix $A \in \mathbb{C}^{M\times N}$ summed with an independent and identically distributed zero mean gaussian noise $W \in \mathbb{C}^{M\times 1}$:

$$\mathbf{X} = \mathbf{AS} + \mathbf{W} \tag{1}$$

According to the signal-subspace theory, $\mathbf{X}$ can be seen as a vector in M dimensional space where each line x_i is a linear combination of source signals s_j through the complex coefficients a_{ij} of $\mathbf{A}$ [7]. Hence, in presence of N sources located so as to avoid the type I ambiguity (i.e. spatial ambiguity, e.g. two sources symmetrically located with respect to the axe of a line array of microphone), the rank of the correlation matrix $\mathbf{R}$ is equal to N where $\mathbf{R}$ is defined as :

$$\mathbf{R} = \mathbb{E}\left\{\mathbf{X}\mathbf{X}^H\right\} \tag{2}$$

and $(.)^H$ is the transpose hermitian operator. Consequently, estimating the number of sources is equivalent to estimate the rank of $\mathbf{R}$. By definition, this is achieved by studying the eigenstructure of $\mathbf{R}$. Using the definition of the mathematical expectation the expression (2) may be expanded as below :

$$\mathbf{R} = \mathbf{A}\psi\mathbf{A}^H + \sigma^2 I_{N\times N} \tag{3}$$

where ψ is the signal correlation matrix and $\sigma^2 I_{N\times N}$ is the noise correlation matrix. The M eigenvalues Λ_i of $\mathbf{R}$ respect the following relations [8, 9] :

$$\begin{array}{lcll} \Lambda_i & = & \mu_i + \sigma^2 & \forall i \in [1, 2, \dots, N] \text{ and } \mu_i \in \mathbb{R}^+ \\ \Lambda_i & = & \sigma^2 & \forall i \in [N+1, \dots, M] \end{array} \tag{4}$$

Hence, if $M > N$, the eigenvectors V_i associated to the eigenvalues Λ_i can be separated in two groups :

$$\begin{array}{rcl} E_S & = & [V_1, V_2, \dots, V_N] \text{ the signal subspace associated to the } N \text{ biggest eigenvalues.} \\ E_N & = & [V_{N+1}, V_{N+2}, \dots, V_M] \text{ the noise subspace associated to the } M-N \text{ smallest eigenvalues.} \end{array}$$

The rank of $\mathbf{R}$ is so deduced regarding the multiplicity of its smallest eigenvalues as illustrated on Fig. 1.

If the theory seems very attractive because of its simplicity, in practice, the smallest eigenvalues are never equal with probability one because of the finite size of the observations [2] and it frequently happens that the signal and noise related eigenvalues are difficultly distinguishable. That's why several methods have been proposed for help to decision. The most popular include the Information

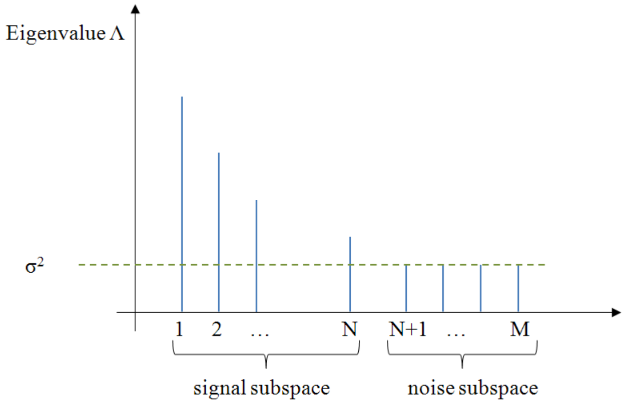

FIGURE 1. A theoretical set of eigenvalues of $\mathbf{R}$ in presence of N sources and $M>N$ microphones, the $M-N$ smallest eigenvalue are equal

Theoretic Criteria : the idea is to test a family of P hypothesis, where the hypothesis p traduces the equality between the $M-p$ smallest eigenvalues, and see which hypothesis best fits the data (i.e. which hypothesis has the maximum likelihood). As maximum likelihood estimators are generally biased, penalty functions are introduced to make correction. Most well-known of them are the AIC (Akaike Information Criterion) [10], MDL (Maximum Description Length) [4], EDC (Efficient Detection Criterion) [3], MDL-BSS [11] and so on. But all above-mentioned methods are inoperative in the at-worst determined case because of possible inexistant noise subspaces. That's why the presented strategy consists in optimizing the sensor array geometry as so to predict the behavior of the eigenvalues in function of the number of radiating sources N and, for a given N, optimize the sensor array geometry to separate them.

3. Proposed method of optimization for the case $M = N_{max} = 2$

From Eq. (1), all the information about i) sensors locations in relation to ii) the sources locations and iii) the sources wavelength is comprised in $\mathbf{A}$. For the sake of simplicity, let's consider the case where $M = N_{max} = 2$. In such a situation, the mixing matrix $\mathbf{A}$ has the following form :

$$\mathbf{A} = \begin{pmatrix} \gamma_{11}e^{-j2\pi a} & \gamma_{12}e^{-j2\pi b} \\ \gamma_{21}e^{-j2\pi c} & \gamma_{22}e^{-j2\pi d} \end{pmatrix} \tag{5}$$

where

$$a = \frac{\|\mathbf{x}_{m1}-\mathbf{x}_{s1}\|_2}{\lambda_1}, \quad b = \frac{\|\mathbf{x}_{m1}-\mathbf{x}_{s2}\|_2}{\lambda_2}, \quad c = \frac{\|\mathbf{x}_{m2}-\mathbf{x}_{s1}\|_2}{\lambda_1}, \quad d = \frac{\|\mathbf{x}_{m2}-\mathbf{x}_{s2}\|_2}{\lambda_2} \tag{6}$$

and $\gamma_{ij} = \beta_j/(4\pi \|\mathbf{x}_{mi}-\mathbf{x}_{sj}\|_2^2)$. In the compact and far-field sensor array context, distance between each sensor is low in comparison with distances between microphones and sources, hence the above model can be simplified by letting $\gamma_{ij} = \alpha_j$ where α_j is one positive constant which represents the initial intensity level of the source j. Under assumption of mutually uncorrelated sources and i.i.d noise, $rank\{\mathbf{R}\} = rank\{\mathbb{E}\{\mathbf{AA}^H\}\}$. The idea is so to act on the microphone array geometry in order to achieve desired eigenvalues of $\mathbf{AA}^H$ depending on the application : source number estimation or source separation. Let Λ be such an eigenvalue, then Λ obeys to $P(\Lambda) = 0$ with :

$$\begin{aligned} P(\Lambda) &= det\left(AA^H - \Lambda I_{M\times M}\right) \\ &= \Lambda^2 - 2(\alpha_1^2+\alpha_2^2)\Lambda + 4\alpha_1^2\alpha_2^2 sin^2(\pi(a-b+c-d)) \end{aligned} \tag{7}$$

Which gives two solutions :

$$\Lambda_1 = \frac{2(\alpha_1^2 + \alpha_2^2) + \sqrt{\Delta}}{2}, \Lambda_2 = \frac{2(\alpha_1^2 + \alpha_2^2) - \sqrt{\Delta}}{2} \tag{8}$$

Where Δ equals :

$$\Delta = 4(\alpha_1^2 + \alpha_2^2)^2 - 16\alpha_1^2\alpha_2^2 sin^2(\pi(a - b + c - d)) \tag{9}$$

By working on Λ_1 and Λ_2, some different relations between the optimal position $\mathbf{x}_{m2}$ and other parameters in **A** can be found according to the application context : source separation or number of source estimation. This is discussed in the following.

3.1 Source separation context

For a source separation application, the array geometry has to set the observations as independent (in a second order sense) as possible and so to set the associated eigenvectors as orthogonals as possible. Mathematically speaking, knowing $\mathbf{x}_{m1}$, $\mathbf{x}_{s1}$, $\mathbf{x}_{s2}$ λ_1 and λ_2, this amounts in find the optimal $\mathbf{x}_{m2}$ permitting to have a geometrical multiplicity of $\mathbf{AA}^H$ equal to N_{max}, that is to say :

$$\text{find } \mathbf{x}_{m2} \quad \text{such that} \quad dim[Ker(\mathbf{AA}^H - \lambda\mathbf{I})] \quad = N_{max} \tag{10}$$

As AA^H is an hermitian matrix, algebraic and geometric multiplicity are similar. Thus, a sufficient condition to verify (10) is to equalize both of its eigenvalues. From (8) and (9), it comes :

$$\Lambda_1 = \Lambda_2 \Leftrightarrow \Delta = 0 \tag{11}$$

i.e. :

$$a - b + c - d = \pm\frac{1}{\pi}Arcsin\left(\frac{\alpha_1^2 + \alpha_2^2}{2\alpha_1\alpha_2}\right) \tag{12}$$

Because of parameters a, b, c and d are real, initial intensities α_1 and α_2 of the sources had to respect the following contraint to give a physical solution :

$$\left|\frac{\alpha_1^2 + \alpha_2^2}{2\alpha_1\alpha_2}\right| \leq 1 \tag{13}$$

Without loss of generality, let set α_2 = $k\alpha_1$ where $k \in \Re^+$, it comes :

$$1 + k^2 \leq 2k \tag{14}$$

Hence, only k = 1 gives a physical solution, letting to think that the optimal $\mathbf{x}_{m2}$ can be found only where both sources have the same initial radiating intensity. In other cases, only suboptimal separation can be achieved with two microphones and more evolutive methods had to be deployed. If k = 1, Eq. (12) yields the final equality constraint h that $\mathbf{x}_{m2}$ had to verify with respect to $\mathbf{x}_{m1}$, $\mathbf{x}_{s1}$ and $\mathbf{x}_{s2}$:

$$h(\mathbf{x}_{m2}) = \frac{1}{\lambda_1}\left(\|\mathbf{x}_{s1} - \mathbf{x}_{m2}\|_2 - \|\mathbf{x}_{s1} - \mathbf{x}_{m1}\|_2\right) + \frac{1}{\lambda_2}\left(\|\mathbf{x}_{s2} - \mathbf{x}_{m1}\|_2 - \|\mathbf{x}_{s2} - \mathbf{x}_{m2}\|_2\right) \pm \frac{1}{2} = 0 \tag{15}$$

3.2 Detection context

In the context of source number estimation where N_{max} = 2, the goal is to discriminate three cases, case a) : both sources radiate, case b) : one source radiates, and case c) : no sources radiate. Because of the ignorance of the initial intensity of sources, the eigenvalues are difficultly predictable, but their ratio $r = \Lambda_2/\Lambda_1$ is much more easier to predict. Supposing that $\mathbf{x}_{m2}$ respects the source separation constraint (15), r equals one in the case a) because of the independence of the two recordings for this geometry. But by definition of an i.i.d noise $\mathbf{W}$, r also equals one into the case c). So if $\mathbf{x}_{m2}$ respects Eq. (15), case a) and c) can't be dissociated. That's why an other optimal $\mathbf{x}_{m2}$ (this time in the sense of the number of source estimation context) has to be found.
Suppose to be in the case a), the conditions to respect are $r \neq 1$ in order to avoid ambiguity with case c) and $r \neq 0$ in order to avoid ambiguity with the case b). After calculations, we found :

$$\text{choose } \mathbf{x}_{m2} \quad \text{such that} \quad a - b + c - d \neq \begin{cases} \pm\frac{1}{2} & \text{if } \alpha_1 = \alpha_2 \\ \mathbb{Z} & \text{otherwise} \end{cases} \tag{16}$$

For example, in the specific case where $k = 1$ and we want $r = 0.5$, it comes :

$$h(\mathbf{x}_{m2}) = \frac{1}{\lambda_1}\left(\|\mathbf{x}_{s1} - \mathbf{x}_{m2}\|_2 - \|\mathbf{x}_{s1} - \mathbf{x}_{m1}\|_2\right) + \frac{1}{\lambda_2}\left(\|\mathbf{x}_{s2} - \mathbf{x}_{m1}\|_2 - \|\mathbf{x}_{s2} - \mathbf{x}_{m2}\|_2\right) \pm \frac{1}{\pi} Arccos\left(\frac{1}{3}\right) = 0 \tag{17}$$

3.3 Optimization procedure

From Eq. (15) and (17), the optimal position of $\mathbf{x}_{m2}$ can be found using a standard optimization method formulated as :

$$\min_{\mathbf{x}_{m2} \in \mathcal{R}^2} \quad f(\mathbf{x}_{m2}) \tag{18}$$

$$\text{subject to} \quad h(\mathbf{x}_{m2}) = 0 \tag{19}$$

Where the function to minimize here is the distance between the first microphone according to assumptions made at the beginning of the section :

$$f(\mathbf{x}_{m2}) = \|\mathbf{x}_{m2} - \mathbf{x}_{m1}\|_2 \tag{20}$$

A standard method for solve such a non linear convex optimization problem is the Local Sequential Quadratic Programming method (Local-SQP). For a complete description of Local-SQP, see for example [12] page 465. As an example of results, Fig. 2 illustrates a simulation where $\mathbf{x}_{s1} = [0, 0]^T$ m, f_1 = 600 Hz, $\mathbf{x}_{s2} = [0.6, 0.3]^T$ m, f_2 = 500 Hz and $\mathbf{x}_{m1} = [0, -1]^T$ m. Both sources radiated with the same intensity level (k = 1). For the source separation context, the found optimal position $\mathbf{x}_{m2}$ equals $[0.62, -1.03]^T$ m and for the number of source estimation context, the found optimal $\mathbf{x}_{m2}$ equals $[0.48, -1.08]^T$ m.

4. Experimental measurements (anechoïcal conditions)

Let's consider two sound sources s_1 and s_2, each radiates an harmonic sound of respective frequency $f_1 = 2000$ Hz and $f_2 = 3000$ Hz with the same intensity level (k = 1). Positions of sources and first microphone are : $\mathbf{x}_{s1} = [0, 0]^T$ m, $\mathbf{x}_{s2} = [0.5, 0]^T$ m and $\mathbf{x}_{m1} = [0, -4]^T$ m. The ordinate of microphone m_2 is set equal to -4 m and different abscissas are tested :

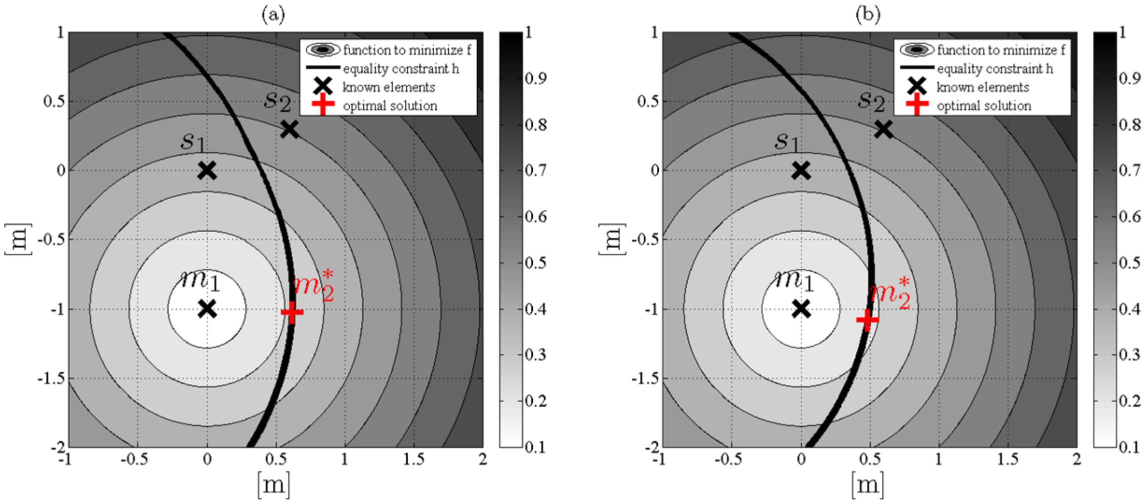

FIGURE 2. Optimum solution for location of microphone m_2 in the a) source separation context and b) source detection context for the same situation.

Position Number	1	2	3	4	5	6	7	8	9
Abscissa of m_2 [m]	0.125	0.245	0.36	0.48	0.59	0.725	0.845	0.96	1.075

Position Number	10	11	12	13	14	15	16	17
Abscissa of m_2 [m]	1.2	1.32	1.445	1.565	1.68	1.8	1.925	2.045

For each location of m_2, the ratio r defined in Section 3.2 is computed. It can be seen on Fig. 3 that theoretical and experimental values of r match well for small distances between m_1 and m_2 (until 1.6 m).

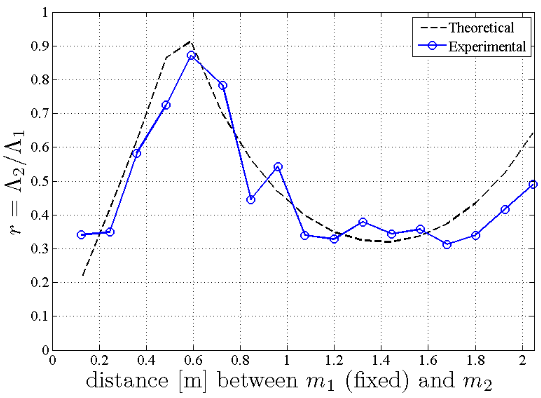

FIGURE 3. Theoretical and experimental ratio $r = \Lambda_2/\Lambda_1$ according to the abscissa of the second microphone.

In such a situation, it clearly appears that the maximal independence between both recordings is obtained for the fifth position of m_2. It can be said this position is the optimal one for a source separation context towards all tested positions (or positioning constraints). According to the statistical signal processing theory, an efficient [1] estimator of $\mathbf{S}$ is $\hat{\mathbf{S}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{X}$ [13]. The computation of such an estimator is done for two positions : the fifth (optimal : , i.e $m_2 = [0.59, -4]^T$) and the sixth (i.e $m_2 = [0.725, -4]^T$). Both results are represented on Fig. 4. As expected the estimation of the original sources is conclusive when $\mathbf{x}_{m2}$ is the optimal position and much less when $\mathbf{x}_{m2}$ is a few centimeters side the optimal. This proving the influence of the microphone array geometry in the final performance of a source separation application.

1. an estimator is said efficient if it is unbiased and if it attains the Cramer-Rao Lower Bound

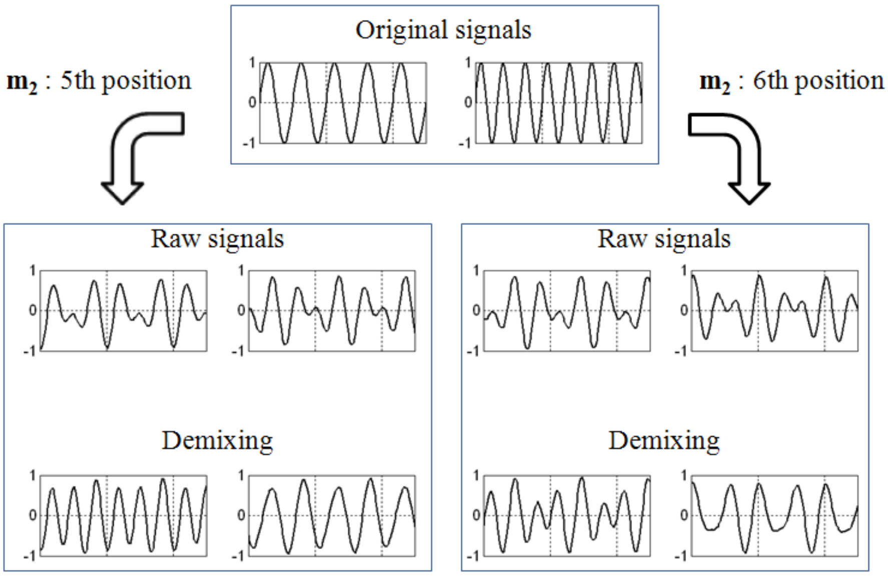

FIGURE 4. Comparison between original signals and source separation results from raw recorded signals for the optimal position of m_2 and another one (13.6 cm side). The source separation is well performed for the theoretically found optimal solution in comparison with another position.

In the same manner, the smallest microphone array which give r equal to 0.5 is obtained for $\mathbf{x}_{m2}$ comprised between the second and third position. For the abscissa of m_2 equal to 0.31 m, a recording has been done where s_1 and s_2 radiate randomly (Fig. 5-a). The three different cases : no signal, one signal and two signals, are clearly distinguishable and conform to the theory as illustrated on Fig. 5-b. Using the fifth position would have not permitted to differentiate no signal case from two signals case. This confirms that an optimal microphone array in the source separation context is not necessary optimal for a detection context and vice-versa.

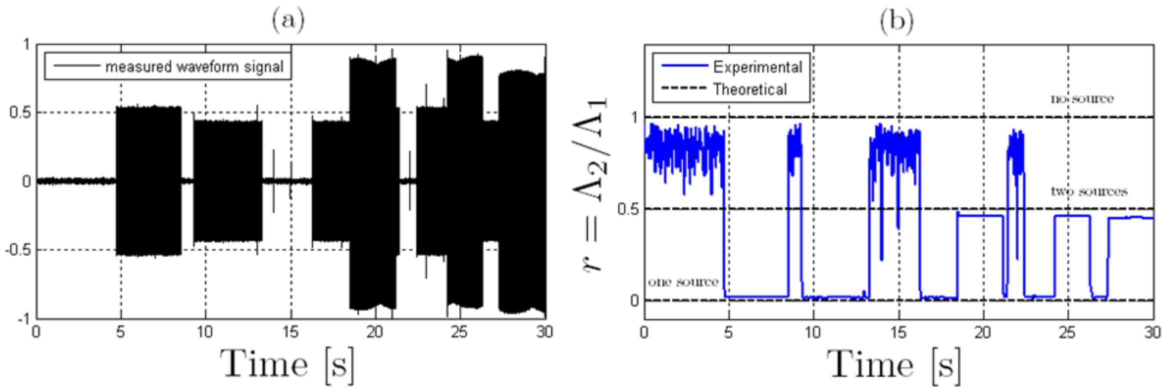

FIGURE 5. (a) waveform of received signal on one microphone, (b) $r = \Lambda_2/\Lambda_1$ in time, as expected 0 value is obtained when 1 source is radiating, 0.5 when 2 sources are radiating and 1 in case of noise. Theoretical expected value for the ratio in the 1 and 2 source case are represented by dash lines.

5. Conclusion

Based on the eigenvalues of the spatial correlation matrix of the observations, some relations to find the optimal sensor array geometry in the sense of sources separation and sources detection have been exposed. The theoretical development and the experimental examples have been proposed for the case of two harmonic and uncorrelated sources. In particular, it was demonstrated and confirmed by measurements that the best sensor array geometry could be found by solving a simple optimization problem, but the result of this optimization is not the same for both applications. Forthcoming works will consist in extending this work for larger bandwidth and much more sound sources.

6. Aknowledgments

The authors would like to thanks Prof. Dacorogna (EPFL, Switzerland) for his valuable help about calculations and interpretations of the eigenvalues.

Références

[1] K. Yamamoto, F. Asano, W.F.G. van Rooijen, E.Y.L. Ling, T. Yamada, and N. Kitawaki. Estimation of the number of sound sources using support vector machines and its application to sound source separation. volume 5, pages V – 485–8 vol.5, apr. 2003.

[2] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 33(2) :387 – 392, apr. 1985.

[3] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai. On detection of the number of signals in presence of white noise. *Journal of Multivariate Analysis*, 20(1) :1–25, October 1986.

[4] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5) :465 – 471, 1978.

[5] Y. Liu, J.J. Soraghan, and T.S. Durrani. Detection of number of harmonics by maximum eigenvalue varied rate criteria. pages 2543 –2546 vol.5, apr. 1990.

[6] Jiang Lei, Cai Ping, and Yang Juan. The source number estimation based on the beam eigenvalue method. In *Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on*, pages 2727 –2731, May 2007.

[7] Ralph O. Schmidt. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3) :276–280, march 1986.

[8] G. Bienvenu and L. Kopp. Optimality of high resolution array processing using the eigensystem approach. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 31(5) :1235 – 1248, oct. 1983.

[9] J.-J. Fuchs. Estimating the number of sinusoids in additive white noise. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(12) :1846 –1853, Dec 1988.

[10] H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6) :716 – 723, dec. 1974.

[11] Radu Balan. Estimator for number of sources using minimum description length criterion for blind sparse source mixtures. In *Proceedings of the 7th international conference on Independent component analysis and signal separation*, ICA'07, pages 333–340, Berlin, Heidelberg, 2007. Springer-Verlag.

[12] Michel Bierlaire. *Introduction à l'optimisation différentiable*. Presses Polytechniques et Universitaires Romandes, 2006.

[13] Steven M.Kay. *Fundamentals of Statistical Signal Processing : Estimation Theory*. Prentice Hall PTR, 2010.