RESOURCE CONSTRAINED EFFICIENT ACOUSTIC SOURCE LOCALIZATION AND TRACKING USING A DISTRIBUTED NETWORK OF MICROPHONES

G. Valenzise, G. Prandi, M. Tagliasacchi, A. Sarti

Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy
Email: {valenzise,prandi,tagliasa,sarti}@elet.polimi.it

ABSTRACT

In this paper we present an efficient method to perform acoustic source localization and tracking using a distributed network of microphones. In this scenario, there is a trade-off between the localization performance and the expense of resources: in fact, a minimization of the localization error would require to use as many sensors as possible; at the same time, as the number of microphones increases, the cost of the network inevitably tends to grow, while in practical applications only a limited amount of resources is available. Therefore, at each time instant only a subset of the sensors should be enabled in order to meet the cost constraints. We propose a heuristic method for the optimal selection of this subset of microphones, using as distortion metrics the Cramer-Rao Lower Bound (CRLB) and as cost function the total distance between the selected sensors. The heuristic approach has been compared to an optimal algorithm, which searches the best sensor configuration among the full set of microphones, while satisfying the cost constraint. The proposed heuristic algorithm yields similar performance w.r.t. the full-search procedure, but at a much less computational cost. We show that this method can be used effectively in an acoustic source tracking application.

Index Terms— Acoustic Source Localization, Acoustic Source Tracking, Particle Filtering, Distributed Microphones Network

1. INTRODUCTION

The problems of acoustic source localization and tracking have received a great deal of attention in recent years, for their intrinsic usefulness in practical applications such as monitoring and video-surveillance [1]. The performance of acoustic localization depends on the geometry of the microphone array adopted, on the source position and on the SNR. Thus, when the source is moving, a large number of microphones should be used to obtain high-performance localization, with the side-effect that a great computational power is required for the system to produce the position estimates. Recently, an increasing interest towards *sensor networks* has opened the doors to the use of distributed microphone networks [2], [3]. In this scenario, the localization task can be accomplished without using all the microphones simultaneously, but selecting from time to time the best subset of microphones that allows to produce a good estimate of the source position. In order to do this, some data have to be exchanged between microphones. Communication between the nodes of the network is generally costly: for example, the total power consumption increases as the average distance between the selected sensors grows. Therefore, to select the optimal subset of microphones out of the full set of sensors, not only does the chosen configuration need to maximize some performance metrics (e.g. the localization precision), but also the cost of the resulting network must be cheap enough so that the resource constraints are met.

Most of the approaches used in distributed acoustic source localization and tracking may be borrowed by the sensor-network literature. In [2], a decentralized, dynamic clustering algorithm for target tracking in a wireless sensor network is proposed. The full set of sensors is partitioned in clusters, and the optimization is carried out in such a way that only the cluster nearest to the target is activated at each time instant. In this way, the algorithm implicitly performs a trade-off optimization between the sensing performance and the communication cost. In [3] the constrained selection of the sensors is made explicit, and a sensor-scheduling optimization for target tracking, which takes into account sensor usage, is devised. The work proposed in [4] applies the sensor network ideas to the acoustic tracking of vehicles, using a Bayesian approach. A sensor selection method is also proposed, which tries to choose the most "informative" microphones, thus minimizing the waste of resources. A theoretical analysis of the performance of sensor selection for audio source tracking is presented in [5], where dense circular clusters of uniformly distributed sensors are considered. It is shown that, using a localization algorithm based on range-differences (e.g. spherical interpolation [6]), the performance of the distributed clusters depend on the radii of the arrays and on the distance of the source from the cluster reference microphone.

This paper describes a heuristic microphone selection algorithm for acoustic source tracking, when a resource constraint has to be fulfilled. The goal is to dynamically build a microphone array, composed by a subset of the full collection of available sensors, so that the error on the source position estimate (measured by MSE) is minimized. In the tracking problem, at each time instant a dynamic model of the source trajectory gives the expected source location. Using this information, we can compute the Cramer-Rao Lower Bound (CRLB) of the source position estimate for each sensor configuration (see Section 2), which can be used as cost function for array selection. This procedure can be very costly as the number of microphones increases; for this reason, we present a recursive procedure which noticeably reduces the number of computations. This heuristic technique gives quasi-optimal results and can be employed for real time computations.

The rest of the paper is organized as follows: in the next section the microphone selection problem is set up by defining the fundamental parameters used in the heuristic resource-constrained optimization, which is detailed in Section 3. Section 4 presents some experimental results about the heuristic selection performance with static and moving sources. Finally, Section 5 draws some concluding remarks.

2. PROBLEM FORMULATION

Consider a set of M omnidirectional microphones, distributed onto a finite 2D environment. To perform the localization and tracking of an acoustic source that lies in the same environment, we want to select an optimal subset $\mathcal{S}_K^*$ of K microphones, $3 \leq K \leq M$, by solving the following combinatorial constrained optimization problem:

$$\min_{\mathcal{M}} D(\mathcal{S}_K, \mathbf{p}) \quad \text{s.t.} \quad R(\mathcal{S}_K) \leq R_T \tag{1}$$

where $\mathcal{M}$ denotes all the possible combinations of the M available sensors and R_T is a total resource constraint. The functions D and R give, respectively, an estimate of the localization error and of the expense of resources, for a given sensor configuration $\mathcal{S}_K$ and a source position $\mathbf{p}$. These two functions can be detailed as follows:

Array Cost $R(\mathcal{S}_K)$. We consider as *array cost* the total distance between the K microphones of the subset. When a maximum array cost constraint is established, we discard all the solutions (arrays) that do not satisfy such constraint. Let d_{ij} denote the Euclidean distance between two microphones i and j belonging to the subset $\mathcal{S}_K$. The array cost is computed as follows:

$$R(\mathcal{S}_K) = \frac{1}{2} \sum_{i,j \in \mathcal{S}_K} d_{ij} \tag{2}$$

Distortion $D(\mathcal{S}_K, \mathbf{p})$. We consider as *distortion* metrics the Mean Square Error of the source localization estimate, given $\mathcal{S}_K$ and $\mathbf{p}$. To have a simple expression for the MSE, we approximate it by computing the Cramer-Rao Lower Bound (CRLB) [7], which gives a bound on the variance of the estimated source coordinates. Therefore we postulate that the selected array adopts an unbiased and efficient estimator, which attains the CRLB. Actually, such an estimator does not exist [8]. However, we employ an MLE estimator [6], which is *asymptotically* efficient. As shown in [9], the variance of the MLE is very close to the theoretical CRLB also when a small number of microphones is used.

The Cramer-Rao Lower Bound for the estimation variance of each source coordinate is computed from the inverse of the Fisher Information Matrix (FIM) $\mathbf{J}$:

$$\mathbf{J} = \mathbf{G}\mathbf{C}^{-1}\mathbf{G}^T \tag{3}$$

The columns of the $2 \times \frac{K(K-1)}{2}$ matrix $\mathbf{G}$ are difference vectors between unit-norm vectors $\mathbf{g}_i$, which are directed from the source to the i-th microphone:

$$\mathbf{G} = [\ldots, \mathbf{g}_{ij}, \ldots] \tag{4}$$

$$\mathbf{g}_{ij} = \mathbf{g}_i - \mathbf{g}_j \tag{5}$$

$$\mathbf{g}_i = \frac{\mathbf{p} - \mathbf{q}_i}{\|\mathbf{p} - \mathbf{q}_i\|}, i \in \mathcal{S}_K \tag{6}$$

In equation (6), the vector $\mathbf{q}_i$ represents the position of the i-th microphone that belongs to the subset $\mathcal{S}_K$. $\mathbf{C}$ is a $\frac{K(K-1)}{2} \times \frac{K(K-1)}{2}$ diagonal matrix that contains the variances of the TDOA estimates for each couple of microphones. In our work we take into account the influence of the SNR on TDOA estimation variance [10], considering only high SNRs. Using an exponential acoustic attenuation model [11] on source-microphone distance we obtain the following inverse diagonal matrix:

$$\mathbf{C}^{-1}_{(t,t)} = \frac{SNR}{d_i^\alpha + d_j^\alpha}, \, i, j \in \mathcal{S}_K, \, 0 \leq t \leq \frac{K(K-1)}{2} \tag{7}$$

$$d_i = \|\mathbf{p} - \mathbf{q}_i\| \tag{8}$$

SNR indicates the Signal-To-Noise Ratio when the source-microphone distance is zero, and t is the column-index of the vector $\mathbf{g}_{ij}$ in $\mathbf{G}$. We assume that $\alpha = 2$. To derive an MSE distortion metrics from the CRLB, we use the trace of $\mathbf{J}^{-1}$, which is the sum of the variances of each estimated source coordinate:

$$D(\mathcal{S}_K, \mathbf{p}) = \text{tr}(\mathbf{J}^{-1}) \tag{9}$$

We notice that by formulation, $\mathbf{G}$ and $\mathbf{C}$ matrices depend only on the geometry of the array and on the acoustic source position.

3. HEURISTIC MICROPHONE SELECTION FOR SOURCE LOCALIZATION

We aim at finding a minimum distortion subset $\mathcal{S}_K^*$ of K microphones, given the source position and a maximum array cost constraint, as formalized in the previous section. The most effective algorithm we can use to solve the optimization problem specified by (1) is based on a *full-search* over all the possible solutions $\mathcal{S}_K \in \mathcal{M}$. Unfortunately, the complexity of such approach grows as 2^M. We can dramatically diminish such complexity by exploiting suboptimal heuristic techniques. In this work we use two heuristic algorithms which adopt a general approach composed of the following two main steps:

Array initialization The array is initialized by inserting the first k microphones using heuristic rules based on the array cost and distortion functions; this step leads to an initial, possibly incomplete, microphone subset $\mathcal{S}_k$.

Array completion The array is updated using adding microphones, until we reach the target array cost, exploiting distortion and array cost functions; this step leads to the final $\hat{\mathcal{S}}_K^*$ subset.

There is no guarantee that $\mathcal{S}_K^* = \hat{\mathcal{S}}_K^*$, i.e. the heuristic approach is in general suboptimal w.r.t. the full search algorithm. In Section 4 we will see that in practice, if the resource constraints are not too tight, the distortions obtained with the two methods are almost equal. In the following, we describe in detail the initialization and completion steps.

3.1. Array initialization

The initialization step builds a first feasible solution, which is used as starting point for the subsequent completion phase. We have used two different initialization heuristics: a k Nearest Neighbor (k-NN) technique and a Greedy Randomized Adaptive Search Procedure (GRASP) [12]. The k-NN initialization is based on the choice of the k microphones nearest to the source: this technique requires $M \log M$ computations. The GRASP initialization has the same computational complexity, but in this case the first k microphones are chosen to resemble a Uniform Angular Array (UAA) [8], i.e. the sensor should be uniformly spaced along a circle centered in the source. The GRASP method can be summarized by the following two steps:

Construction The algorithm is initialized by finding the L microphones nearest to the source. We call this subset $\mathcal{N}_L$. For each microphone $l \in \mathcal{N}_L$ we try to build a UAA of k microphones that contains l:

1. The theoretical positions of the other $k-1$ microphones are established.

2. For each position, we select the nearest microphone which has not been selected yet.

Local search Given the previous L solutions, we choose the one that exhibits the minimum distortion.

3.2. Array Completion

It is possible that heuristic algorithms used in the initialization phase do not exploit all the resources available, so other microphones may be added to the subset. Adding a new microphone allows to decrease the distortion, but at the same time it increases the array cost. The goal is to reach the maximum decrease in distortion with the minimum increase in array cost. Our approach is based on the evaluation of the following ratio:

$$\rho(\mathcal{S}_k, \mathcal{S}_{k+1}, \mathbf{p}) = -\frac{D(\mathcal{S}_{k+1}, \mathbf{p}) - D(\mathcal{S}_k, \mathbf{p})}{R(\mathcal{S}_{k+1}) - R(\mathcal{S}_k)} \quad (10)$$

where D identifies the distortion terms and R the array cost terms. The subscripts k and $k+1$ are used to identify values respectively before and after the insertion of the new microphone. In our heuristic approach, we add the microphone that maximizes ρ, since this produces the best trade-off between the distortion gain and array cost increase. Each time a new microphone is inserted, a check on the array cost constraint R_T is made. The array completion phase requires $O(M)$ computations; therefore, the whole heuristic selection has a computational complexity of $M \log M + M$.

3.3. Fast Fisher Information Matrix Updating

When a new microphone is inserted into the incomplete subset $\mathcal{S}_k$ we have to manage the re-computation of the Fisher Information Matrix $\mathbf{J}$ for the evaluation of the new localization distortion $D(\mathcal{S}_{k+1}, \mathbf{p})$. Of course, the new value of $\mathbf{J}$ can be evaluated by computing the matrices $\mathbf{G}$ and $\mathbf{C}^{-1}$ in (3): however, this procedure may be too complex on the computational side, especially for real-time applications. In this section we present a less expensive Fisher Information Matrix updating algorithm, useful for real-time localization and tracking tasks.

Using the notation given in [8], we can rewrite equation (3) in the following form:

$$\mathbf{J}_k = \mathbf{H}_k \mathbf{T}_k \mathbf{C}_k^{-1} \mathbf{T}_k^T \mathbf{H}_k^T \quad (11)$$

$$\mathbf{H}_k = [\ldots, \mathbf{g}_i, \ldots], i \in \mathcal{S}_k \quad (12)$$

where $\mathbf{H}_k$ is a $2 \times k$ matrix that contains the versors $\mathbf{g}_i$, which point from the source to the microphones; $\mathbf{C}_k^{-1}$ is a $\frac{k(k-1)}{2} \times \frac{k(k-1)}{2}$ matrix; finally, $\mathbf{T}_k$ is a $k \times \frac{k(k-1)}{2}$ transform matrix for which we have:

$$\mathbf{G}_k = \mathbf{H}_k \mathbf{T}_k \quad (13)$$

For example, for $k = 4$ we have:

$$\mathbf{T}_4 = \begin{bmatrix} -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \quad (14)$$

For each column we have two non-zero entries, so that each column of $\mathbf{H}_k \mathbf{T}_k$ is a difference vector $\mathbf{g}_{ij}$. When we add a new microphone, a new subset $\mathcal{S}_{k+1}$ of $k+1$ microphones is generated, and a new vector $\overline{\mathbf{g}}$ has to be added to $\mathbf{H}_k$, generating a new matrix $\mathbf{H}_{k+1}$.

The new FIM (related to the new subset $\mathcal{S}_{k+1}$) can be rewritten as follows:

$$\mathbf{J}_{k+1} = \mathbf{H}_{k+1} \mathbf{T}_{k+1} \mathbf{C}_{k+1}^{-1} \mathbf{T}_{k+1}^T \mathbf{H}_{k+1}^T \quad (15)$$

If we define

$$\mathbf{H}_{k+1} = [\overline{\mathbf{g}} | \mathbf{H}_k] \quad (16)$$

we can write the new $\mathbf{T}_{k+1}$ as:

$$\mathbf{T}_{k+1} = \left[\begin{array}{c|c} -\mathbf{1}^T & \mathbf{0} \\ \hline \mathbf{I}_k & \mathbf{T}_k \end{array}\right] \quad (17)$$

Where $\mathbf{1} = [1, 1, \ldots, 1]^T$ is a $k \times 1$ vector of ones, and $\mathbf{I}_k$ is the $k \times k$ identity matrix. The new $\mathbf{C}_{k+1}^{-1}$ matrix is then defined as:

$$\mathbf{C}_{k+1}^{-1} = \left[\begin{array}{c|c} \overline{\mathbf{C}}^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{C}_k^{-1} \end{array}\right] \quad (18)$$

$\mathbf{C_n}$ is a $k \times k$ diagonal matrix that contains the variances on TDOA estimates between the added microphone and each of the microphones that belong to subset $\mathcal{S}_k$. Rewriting now the FIM matrix (15) gives the update equation:

$$\begin{aligned} \mathbf{J}_{k+1} &= \mathbf{J}_k + \overline{\mathbf{g}} \mathbf{1}^T \overline{\mathbf{C}}^{-1} \mathbf{1} \overline{\mathbf{g}}^T + \mathbf{H}_k \overline{\mathbf{C}}^{-1} \mathbf{H}_k^T + \\ &\quad - \overline{\mathbf{g}} \mathbf{1}^T \overline{\mathbf{C}}^{-1} \mathbf{H}_k^T - \mathbf{H}_k \overline{\mathbf{C}}^{-1} \mathbf{1} \overline{\mathbf{g}}^T \end{aligned} \quad (19)$$

Note that the computation of (15) requires $O(k^4)$ multiplications, while (19) can be computed with just $O(k^2)$ multiplications, thus yielding a noticeable complexity reduction.

4. EXPERIMENTAL RESULTS

We have tested the heuristic approach described in Section 3 on two localization/tracking tasks. In the first simulation, we compare the MSE obtained with arrays of different size K (which correspond to different resource constraints) selected through the heuristic algorithms with the MSE produced by arrays built with the full search procedure. In the second experiment, instead, we consider a more realistic application, in which the array selection is performed dynamically to track a moving source. We have considered as test environment a 10×5 m rectangular 2D space for both simulations.

4.1. Heuristic Microphone Selection

We tested heuristic algorithms for microphone selection using 100 different random configurations of $M = 16$ uniformly distributed microphones. The source is positioned at the coordinates $\mathbf{p} = [5, 2.5]^T$, and the SNR is 10 dB. For each configuration, we set 6 different maximum rate constraints, based on the maximum rate R_{MAX} available for each configuration (the sum of all the microphone distances), and we measured the distortion associated to the best subset found from each heuristic algorithm. In equation (20) the vector R_T that contains the six maximum rate thresholds is defined.

$$R_T = R_{MAX} \begin{bmatrix} \frac{1}{100} & \frac{1}{50} & \frac{1}{20} & \frac{1}{10} & \frac{1}{5} & \frac{1}{3} \end{bmatrix} \quad (20)$$

Figure 1 shows the mean array cost/distortion curves obtained using our different initialization algorithms. As a reference performance index, also the curve related to the full-search technique is drawn. The GRASP approach overcomes the result of the simple k-NN especially for very low rate thresholds. The highest distortion gains on kNN are located when the array cost constraint is $R_T(1) = R_{MAX} \frac{1}{100}$ and $R_T(2) = R_{MAX} \frac{1}{50}$.

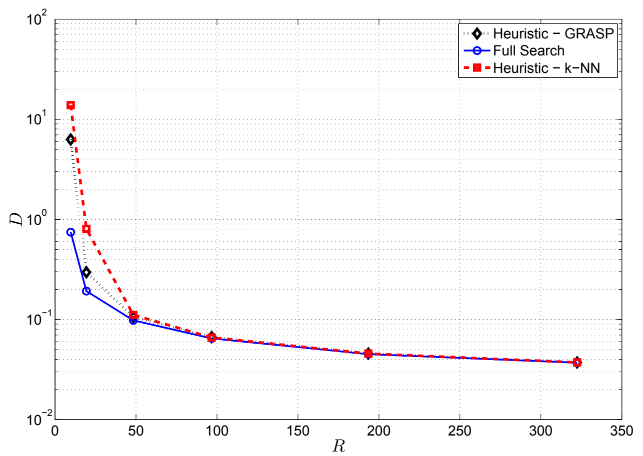

Fig. 1. Array Cost/Distortion mean curves over 100 different random sensor configurations.

SNR [dB]	0	5	10	15
MSE $R_T = 10$	1.56	0.41	0.30	0.12
MSE $R_T = 20$	0.79	0.32	0.13	0.05
MSE $R_T = 50$	0.40	0.33	0.06	0.05

Table 1. Mean RMSE for source tracking for different SNRs.

4.2. Source Tracking

During tracking, a particle filtering technique [13] [14] is adopted to estimate the trajectory of the source. At each time instant t, the next estimated position of the source $\hat{\mathbf{p}}(t)$ is used to choose the array configuration during the evolution of the system, exploiting the GRASP-based heuristic microphone selection and the fast FIM update algorithm presented in Section 3.3. The acoustic source follows a pseudo-casual trajectory computed through the Langevin model [13], over 200 time instants $t = [1, 2, \ldots, 200]$ (see Figure 2).

5. CONCLUSIONS AND FUTURE WORKS

In this paper an efficient algorithm to perform source localization and tracking using a distributed network of microphones has been described. The localization and tracking tasks are performed under a resource constraint that influences the number of active microphones. Future work involves the investigation of the performance of algorithms on configurations in which the microphones are grouped in clusters: the idea is to exploit array fusion techniques when microphones of different clusters have to be involved in the localization task.

6. REFERENCES

[1] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, and A. Sarti, “Scream and gunshot detection and localization for audio-surveillance systems,” in *Proc. IEEE Int. Conf. Advanced Video and Signal-based Surveillance Systems*, London, September 2007.

[2] W.P. Chen, J.C. Hou, and L.Sha, “Dynamic clustering for acoustic target tracking in wireless sensor networks,” in *Proceedings of the IEEE International Conference on Network Protocols*, Atlanta, GA, November 2003.

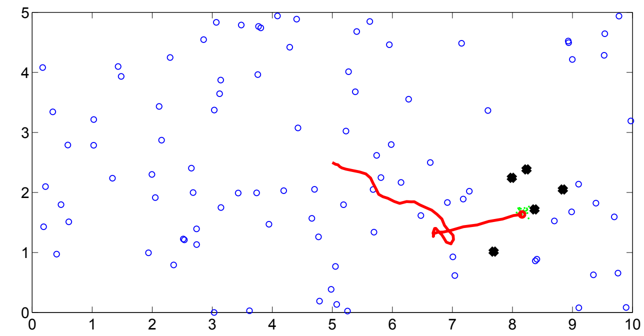

Fig. 2. Tracking of a sound source. The circles are the sensors, the bold crosses are the selected microphones, the small dots are the particles. The continuous line is the source trajectory.

[3] Y. He and E.K.P. Chong, “Sensor scheduling for target tracking in sensor networks,” in *Proceedings of the IEEE Conference on Decision and Control*, Atlantis, Paradise Island, Bahamas, December 2004.

[4] J. Liu, J. Reich, and F. Zhao, “The fusion of distributed microphone arrays for sound localization,” *EURASIP Journal on Applied Signal Processing*, vol. 2003, pp. 378–391, 2003.

[5] C. Meesookho, U. Mitra, and S. Narayanan, “An analysis of range difference based target localization in uniformly distributed sensor field,” in *Proceedings of the ACM and IEEE International Symposium on Information Processing in Sensor Networks*, Los Angeles, CA, April 2005.

[6] Y. Huang, J. Benesty, and G. W. Elko, *Audio Signal Processing for the Next-Generation Multimedia Communication Systems*, chapter 9, Source Localization, Kluwer Academic Publishers, 2004.

[7] S.M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1993.

[8] B. Yang and J. Scheuing, “Cramer-rao bound and optimum sensor array for source localization from time differences of arrival,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia, PA, March 2005.

[9] T. Ajdler, I. Kozintsev, R. Lienhart, and M. Vetterli, “Acoustic source localization in distributed sensor networks,” *Conference Record of the 38th Asilomar Conference on Signals, Systems and Computers*, vol. 2, 2004.

[10] B.M. Sadler and R.J. Kozick, “A survey of time delay estimation performance bounds,” in *Proceedings of the IEEE Workshop on Sensor Array and Multichannel Processing*, Waltham, MA, July 2006.

[11] C. Chang and A. Sahai, “Cramer-Rao-Type Bounds for Localization,” *EURASIP Journal on Applied Signal Processing*, vol. 2006.

[12] T.A. Feo and M.G.C. Resende, “Greedy Randomized Adaptive Search Procedures,” *Journal of Global Optimization*, vol. 6, no. 2, pp. 109–133, 1995.

[13] D.B. Ward, E.A. Lehmann, and R.C. Williamson, “Particle filtering algorithms for tracking an acoustic source in a reverberant environment,” *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 826–836, 2003.

[14] F. Antonacci, D. Riva, D. Saiu, A. Sarti, M. Tagliasacchi, and S. Tubaro, “Tracking multiple acoustic sources using particle filtering,” in *European Signal Processing Conference*, Florence, Italy, September 2006.