Convolutive BSS of short mixtures by ICA recursively regularized across frequencies

Francesco Nesta*, Piergiorgio Svaizer, and Maurizio Omologo

Abstract—This paper proposes a new method of Frequency-Domain Blind Source Separation (FD-BSS), able to separate acoustic sources in challenging conditions. In frequency-domain BSS, the time-domain signals are transformed in time-frequency series and the separation is generally performed by applying the Independent Component Analysis (ICA) at each frequency envelope. When short signals are observed and long demixing filters are required, the number of time observations for each frequency is limited and the variance of the ICA estimator increases due to the intrinsic statistical bias. Furthermore, common methods used to solve the permutation problem fail especially with sources recorded under highly reverberant conditions. We propose a Recursively Regularized implementation of the ICA (RR-ICA) that overcomes the mentioned problem by exploiting two types of deterministic knowledge: 1) continuity of the demixing matrix across frequencies; 2) continuity of the time-activity of the sources. The recursive regularization propagates the statistics of the sources across frequencies reducing the effect of statistical bias and the occurrence of permutations. Experimental results on real-data show that the algorithm can successfully perform a fast separation of short signals (e.g., 0.5-1s), by estimating long demixing filters to deal with highly reverberant environments (e.g., $T_{90} = 700\,\text{ms}$).

Index Terms—Permutation problem, Blind Source Separation (BSS), Independent Component Analysis (ICA), speech enhancement

I. INTRODUCTION

Blind Source Separation (BSS) is a relatively recent digital signal processing technique that has given rise to a lot of interest in the last two decades [1]. Its main objective is to separate multiple sources mixed through unknown channels using only the observations of their mixtures. Such techniques have been applied in many fields of studies such as biology, biomedical signal processing, digital communication, and speech processing. The term “blind” indicates that separation is performed without using any information about the mixing channels and the sources. The simplest case of BSS regards the separation of instantaneous mixtures when the observed signals are linear combinations of the original source signals. To a large extent, this case is solved by means of the Independent Component Analysis (ICA) [2]. A more challenging case regards sources which are mixed through convolutive channels. This situation is quite common for audio applications where signals are generally recorded in reverberant environments. Separation of acoustic sources in a real environment is not a trivial task. Long reverberation time and non-stationary mixing conditions make the estimation of the original source signals hard. Furthermore, the high computational cost hampers the use of these algorithms in real-time applications.

Based on the frequency-domain approach for the convolutive BSS, each recorded signal is transformed in time-frequency series by a short-time Fourier analysis, and for each frequency the separation is performed independently by means of ICA. However, as pointed out in [3], with the frequency-domain approach when the amount of data is small the global separation performance becomes poor due to the low accuracy of the demixing matrices estimated at each frequency. A way to improve the accuracy is to decrease the number of the mixing matrices estimated with the available data, which means to limit the number of parameters that describe our mixing system. For instance, one may reduce the frame size of the short-time Fourier analysis, but in this way the ICA would be less accurate since an instantaneous mixing model cannot describe exactly the mixtures observed at each frequency. An alternative method to mitigate these drawbacks was proposed in [3]. In order to apply the same ICA adaptation to a group of frequencies, a normalization procedure was proposed based on a mixing system approximated by an anechoic model. This method was shown to provide better performance compared to the traditional Frequency-Domain Blind Source Separation (FD-BSS) as long as the direct to reverberant ratio (DRR) is sufficiently high. However, as the distance between microphones and sources increases (i.e. as the DRR decreases) the anechoic model no longer accurately represents the mixing system and the normalization procedure becomes less effective.

In this paper a novel procedure is proposed which overcomes the limitations of the previous methods and improves the FD-BSS accuracy, since it allows to estimate a large number of demixing matrices (i.e. long demixing filters) even with a short amount of data. In particular, two important issues are addressed: a) in frequency-domain the assumption of independence between sources does not completely hold, especially on a short-time basis [4]; b) the solution of the permutation problem is not trivial.

In this work, FD-BSS is applied by recursively regularizing ICA over the frequencies based on the assumption of a priori knowledge: the demixing matrix and the time-activity of the sources are expected to vary continuously across frequencies. The proposed method represents an extension of the work in [5] where only the continuity of the demixing matrices was considered. The algorithm has the following advantages:

1) since the demixing matrices are not constrained by any anechoic model, more accurate long demixing filters can be estimated to cope with a high reverberation time;
2) if there is no spatial aliasing, the permutation problem

The authors are with Fondazione Bruno Kessler -irst, via Sommarive 18, Povo di Trento 38123, (TN) Italy.
email: {nesta,svaizer,omologo}@fbk.eu
is reduced substantially;
3) it is computationally efficient: it can be implemented for a real-time separation, since the recursive structure drastically decreases the number of ICA iterations necessary at each frequency.

The paper is organized as follows: Section II recalls the frequency-domain approach and the physical interpretation of the demixing matrices. Sections III discusses the drawbacks due to the permutation problem especially when the FDBSS is applied to short signals. Section IV describes how to include a priori knowledge in the ICA and proposes the resulting structure of the Regularized Recursive-ICA (RRICA). In Section V a new performance evaluation criterion is proposed in order to compare the accuracy of RRICA with other well-known algorithms. Finally, experimental results on real-data are shown in Section VI, supported by a comparison with other popular BSS approaches.

II. PHYSICAL INTERPRETATION OF FREQUENCY-DOMAIN BSS

In Frequency-Domain BSS the signals observed at the microphones are modeled using a time-frequency representation. At each frequency the observed data can be considered a linear combination of the time-frequency components associated with each source. Let us consider the case of \( N \) sources and \( M \) microphones. In matrix notation one can write:

\[
x(k, \tau) = H(k)s(k, \tau)
\]

where \( k \) is the frequency bin index, \( \tau \) is a time index related to the analysis frame (e.g. in a short-time Fourier analysis), \( x(k, \tau) \) is a column vector of the observed mixtures (in frequency-domain), \( s(k, \tau) \) is the column vector of the original signals (in frequency-domain) and \( H(k) \) is a \( M \times N \) mixing matrix.

According to the impulse responses between the sources and the microphones, the mixing matrix associated to each frequency bin can be modeled as:

\[
H(k) = \begin{pmatrix}
| h_{11}(k) e^{-j\phi_{11}(k)} & \cdots & | h_{1N}(k) e^{-j\phi_{1N}(k)} \\
| h_{M1}(k) e^{-j\phi_{M1}(k)} & \cdots & | h_{MN}(k) e^{-j\phi_{MN}(k)}
\end{pmatrix}
\]

(2)

\[
\varphi_{mn}(k) = 2\pi f_k T_{mn}(k)
\]

(3)

where \( T_{mn}(k) \) represents the propagation delay from the \( n \)-th source to the \( m \)-th microphone for the \( k \)-th frequency bin, \( |h_{mn}(k)| \) is the magnitude of the frequency response between the \( n \)-th source and the \( m \)-th microphone for the \( k \)-th frequency bin, and \( f_k \) is the real frequency (in Hz) corresponding to the \( k \)-th frequency bin.

Assuming \( N = M \), by applying a complex-valued ICA to the time-series \( x(k, \tau) \), one can retrieve the original components \( y(k, \tau) \) by means of a set of demixing matrices \( W(k) \) that are estimates of the matrices \( H^{-1}(k) \) up to scaling and permutation ambiguities:

\[
y(k, \tau) = W(k)x(k, \tau)
\]

(4)

where the demixing matrix \( W(k) \) can be modeled as:

\[
W(k) = \Lambda(k)\Pi(k)\tilde{H}^{-1}(k).
\]

(5)

Here \( \Lambda(k) \) is an arbitrary diagonal scaling matrix, \( \Pi(k) \) is a permutation matrix and \( \tilde{H}^{-1}(k) \) is an estimate of the inverse of the true mixing matrix \( H(k) \). By applying the inversion to the matrix \( W(k) \) we obtain:

\[
W^{-1}(k) = \tilde{H}(k)\Pi^T(k)\Lambda^{-1}(k)
\]

(6)

The matrix \( \Lambda^{-1}(k) \) is still diagonal and the ratios between the elements of two generic rows \( a \) and \( b \) of the matrix \( W^{-1}(k) \) are scaling invariant, i.e. for the \( n \)-th column:

\[
r_n^{(a,b)}(k) = \frac{|w(k)_{an}^{-1}|}{|w(k)_{bn}^{-1}|} \approx \frac{h_{an}(k)}{h_{bn}(k)}
\]

(7)

Assuming the permutation problem to be solved (i.e., \( \Pi(k) = \mathbf{I} \)), each ratio represents the acoustic propagation of the \( n \)-th source with respect to the microphone pair \( (a, b) \) at the \( k \)-th frequency bin. Indeed, according to the model in (2) we can rewrite (7) as:

\[
r_n^{(a,b)}(k) = |c_n^{(a,b)}(k)| e^{-j2\pi f_k \Delta T_n^{(a,b)}(k)}
\]

(8)

where \( \Delta T_n^{(a,b)}(k) \) is the Time Difference of Arrival (TDOA) for the \( n \)-th source at the chosen microphone pair \( (a, b) \) observed at the \( k \)-th frequency bin. In anechoic conditions, where the reverberation is absent, the magnitude \( |r_n^{(a,b)}(k)| \) and the TDOA \( \Delta T_n^{(a,b)}(k) \) are expected to be invariant with respect to the frequency, and consequently the phase of (8) must vary linearly. Hence, as long as the acoustic waves related to the propagation along the direct paths dominate over the secondary reflections, each ratio can be considered as a state observation of the free-field propagation model of the \( n \)-th source.

Note that, when the permutation problem is not solved, the columns of the matrix \( W^{-1}(k) \) may be permuted leading to a different association between states and sources. Hence, up to the scaling indeterminacy, all the demixing matrices can be fully described by using only the parameters of the anechoic propagation model. Each state \( r_n^{(a,b)}(k) \) may be theoretically represented by the ideal model:

\[
c_n^{(a,b)}(k) = \beta_n^{(a,b)} e^{-j2\pi f_k \Delta \tau_n^{(a,b)}}
\]

(9)

where \( \beta_n^{(a,b)} \) is the inter-microphone attenuation ratio and \( \Delta \tau_n^{(a,b)} \) is the true TDOA related to the \( n \)-th source, at the microphone pair \( (a, b) \). A generic demixing matrix for the \( k \)-th frequency bin can be parameterized as:

\[
\hat{H}(k) = \begin{pmatrix}
1 & \cdots & 1 \\
\cdots & \cdots & \cdots \\
\zeta_1^{(N-1,1)}(k) & \cdots & \zeta_1^{(N-1,1)}(k) \\
\zeta_1^{(N,1)}(k) & \cdots & \zeta_1^{(N,1)}(k)
\end{pmatrix}
\]

(10)

Once defined \( \hat{W}(k) = \hat{H}^{-1}(k) \), the \( n \)-th row of \( \hat{W}(k) \) is equivalent to the steering-vector of an adaptive beamformer which removes all the \( N - 1 \) interfering sources with respect to the \( n \)-th source of interest. Indeed, performing a frequency-domain BSS for \( N \) sources is equivalent to estimate a set of \( N \) null-beamformers blindly adapted at each frequency [6].
Following the model in (1)-(6), $W(k)$ can be estimated in a batch adaptation by a gradient descent as follows:

$$W_{i+1}(k) \leftarrow W_i(k) + \eta \Delta W_i(k)$$  \hspace{1cm} (11)

where $W_i(k)$ is the demixing matrix estimated at $i$-th iteration and $\Delta W_i(k)$ is the gradient which updates the solution according to the step-size $\eta$. Based on the cost function that ICA attempts to minimize, the gradient could assume many forms. The remainder of this work refers to the use of an ICA based on the minimization of the Kullback-Liebler divergence forms. The remainder of this work refers to the use of an ICA that attempts to minimize, the gradient could assume many forms. In such a case, $\Delta W_i(k)$ is expressed at each iteration $i$ as follows:

$$y_i(k) = W_i(k)x(k)$$  \hspace{1cm} (12)

$$\Delta W_i(k) = (I - E[\Phi(y_i(k))y_i(k)^H])W_i(k)$$  \hspace{1cm} (13)

where $y_i(k)$ and $x(k)$ indicates the output and input signal vectors, $\Phi(\cdot)$ is a non-linear function and $E[\cdot]$ is the expectation operator. In a batch implementation, the expectation operator is approximated by averaging the instantaneous generalized covariance matrix over $\tau$:

$$E[\Phi(y_i(k))y_i(k)^H] \approx \langle \Phi(y_i(k, \tau))y_i(k, \tau)^H \rangle_{\tau}$$  \hspace{1cm} (14)

where $y_i(k, \tau)$ indicates the output signal vector at instant $\tau$ and $\langle \cdot \rangle_{\tau}$ is the average operator. However, the statistical bias due to local dependencies between narrow-band signals increases as the number of the time-observations reduces; this occurs especially when the signals $x(k, \tau)$ are obtained by means of a high resolution FFT. Such a bias is the main reason for the poor performance of standard frequency-domain BSS techniques, when long demixing filters are used to separate short signals.

### III. Permutation for Short Signals

Permutation is still an open problem which mostly affects the global performance of FD-BSS. The main idea behind these methods is that searching for the optimal solution of demixing filters can be speeded up if each frequency is processed by an independent ICA stage. However, the cost function that is minimized by ICA has a number of equivalent minima corresponding to all the possible permutations of the output signals. If no connections are established between ICAs related to different frequencies, the same order in the source output is not guaranteed in frequency. Therefore, the separated frequency components related to the same source have to be grouped a posteriori.

Many methods of permutation correction were proposed, which can be assigned to two main categories:

1) The methods belonging to the first category, based on TDOA (or Direction of Arrivals (DOA)), exploit the inherent phase information of the demixing matrices by assuming a coherence across frequencies [9]. More advanced approaches exploit the coherence of the demixing matrices across frequencies [10], or estimate the entire propagation model [11].

2) The methods belonging to the second category are based on the inter-frequency magnitude correlation, and exploit the non-stationarity of sources as well as the spectral continuity between adjacent frequencies, which is typical of acoustic signals such as speech [12],[13]. Although less sensitive to the environmental conditions, these methods lack of reliability if very short signals are analyzed [14].

Hybrid approaches have also been investigated [15] [16], to overcome the limitations of the methods of each category. Rather than solving the permutation a posteriori, such indeterminacy can be overcome by means of constraints which interconnect ICA adaptations related to different frequencies. An extension of ICA which explicitly includes inter-frequency dependencies is the Independent Vector Analysis (IVA) [17]. In spite of the promising results obtained with a correct convergence, the IVA robustness is limited by the presence of local minima, similarly to what happens for time-domain approaches.

Another mechanism to impose an inter-frequency coupling is binding each demixing matrix to a frequency dependent model. Popular algorithms constrain, or initialize, the optimization by means of geometrical information such as source and microphone locations [18]. A drawback of these procedures is that the adaptation is strictly constrained by the model or, in the case of initialization, the geometrical information needs to be known in advance. Imposing continuity only between the demixing matrices of adjacent frequencies is a strategy which partially relaxes the constraint on the geometrical model. The latter constraint is justified by the physical interpretation of Section II, and it also corresponds to continuity in the minima of ICA cost function [5]. This evidence was exploited in [19] and [20] by methods that use a recursive initialization of the demixing matrix $W_0(k)$ for the $k$-th frequency bin, starting from the solution $W(k - 1)$ obtained at the previous adjacent frequency bin. In fact, if the variation of two subsequent iterations in (11) is small enough, the convergence point and the permutations depend only on the initialization of the matrix $W_0(k)$, which derives from the steepest-descent behavior of the Natural Gradient. Since recursive approaches can lack of reliability when ICA converges towards poor solutions introducing discontinuities across adjacent frequencies, the robustness of these systems can be improved by considering two main cues:

1) the late reverberation introduces discontinuities between adjacent estimated demixing matrices, which affect the reliability of the recursion. Therefore, the recursive approach can be improved if the initializing matrix is a filtered version of the estimated demixing matrix; the latter one roughly represents the propagation of the acoustic waves over the direct path;

2) the occurrence of poor solutions can be minimized also by reducing the high statistical bias of ICA by means of a priori information regarding the time activity of the sources.

Following the above insights, the next section discusses how deterministic knowledge can be included in the ICA in order to regularize the convergence, thus reducing both the statistical bias and the occurrence of permutation.
IV. RECURSIVELY REGULARIZED ICA

In the traditional FD-BSS approach, when performing the separation independently across frequencies two main kinds of information are lost: the inter-dependencies between frequency components of each source, and the physical meaning of the demixing filters. First, in many types of acoustic processes as those related to speech and music, inter-frequency dependencies are often observable. In particular, the time-activity of the sources is expected to be coherent across frequencies. Second, as long as the acoustic waves related to the direct propagation paths are relatively strong compared with the reflection paths, the demixing matrices are expected to vary continuously across frequencies according to (10).

Such properties are commonly exploited a posteriori in order to reduce the permutation ambiguity. Here, we analyze how to directly include this knowledge in the ICA adaptation, in order to regularize its convergence behavior.

A. Wiener-like weighting

In standard ICA-based BSS methods, it is implicitly assumed that the sources are always overlapping in time. On the other hand, acoustic sources have a sparse representation in time-frequency domain which implies that only one source in each time-frequency point has a non-negligible energy. The sparseness was previously exploited in the Sparse Component Analysis (SCA) [21], [22]. Similarly to the ICA which seeks for the most independent source representation, the SCA seeks for the sparsest source representation. In other techniques the sparse representation is exploited to estimate time-frequency masks used to directly separate the sources [23], [24], [25] or alternatively to improve the solution of the ICA [26], [27]. Here, we show how to exploit the sparseness and the continuity across frequency of the source time-activity in order to improve the ICA accuracy.

First of all, we need a reformulation of the optimization in (11) and (13) in terms of mixing matrix $\mathbf{H}(k)$; the reason is explained in the last part of this section. Still according to the Natural Gradient optimization, the mixing matrix is updated [28] as follows:

$$y_{(i)}(k) = [\mathbf{H}_{(i)}(k)]^{-1}x(k)$$

(15)

$$\Delta \mathbf{H}_{(i)}(k) = \mathbf{H}_{(i)}(k)(I - E[\Phi(y_{(i)}(k))y_{(i)}(k)^H])$$

(16)

$$\mathbf{H}_{(i+1)}(k) \leftarrow \mathbf{H}_{(i)}(k) - \eta \Delta \mathbf{H}_{(i)}(k)$$

(17)

During the adaptation all the coefficients of the matrix $\mathbf{H}_{(i+1)}(k)$ are updated with the estimated gradient, i.e., computed by means of the expectation of the generalized covariance matrix $E[\Phi(y_{(i)}(k))y_{(i)}(k)^H]$. From a theoretical point of view, according to (16) the convergence is reached when the gradient becomes null, i.e., when the generalized covariance matrix $E[\Phi(y_{(i)}(k))y_{(i)}(k)^H]$ becomes an identity matrix. However, the diagonality of this matrix depends not only on the estimated $\mathbf{H}(k)$ but even on the intrinsic correlation between the source signals which is not necessarily null in a short-time.

In a batch approach, the generalized covariance matrix as well as the gradient $\Delta \mathbf{H}_{(i)}(k)$ are estimated by replacing the expectation operator with average over time. If a sufficient number of observations of the output signals $y(k)$ is available, the time average is a good estimator of the gradient since it is assumed that the sources are always overlapping in time. However, when the signals are observed in a short-time, the ICA becomes less accurate due to a high statistical bias in the averaged gradient, which is determined by the intrinsic correlation between the source signals. Thus a better gradient estimation technique is required.

According to the mixing model in (1), each column of the matrix $\mathbf{H}_{(i+1)}(k)$ is an estimation (up to a scaling ambiguity) of the mixing coefficients related to one of the sources. Hence, under the assumption of sparseness, the estimation can be improved if the $n$-th column of the gradient is averaged over the time-instants in which the corresponding source is active. Thus, the generalized covariance is not evaluated on the basis of the time-instants in which the source is likely to be silent, so reducing the bias introduced by the presence of interfering sources. The estimation of the gradient expectation can then be improved by weighting the instantaneous gradient as follows:

$$\Delta \mathbf{H}_{(i)}(k) = \mathbf{H}_{(i)}(k)(I - E[\Phi(y_{(i)}(k))y_{(i)}(k)^H]) = E_{\mathbf{H}(k)}(k)(I - \Phi(y_{(i)}(k))y_{(i)}(k)^H) \simeq \sum_{\tau} \mathbf{\Psi}(k, \tau) \odot [\mathbf{H}_{(i)}(k)](I - \Phi(y_{(i)}(k), \tau)y_{(i)}(k, \tau)^H)$$

(18)

where $\odot$ is the Hadamard product (i.e. element-wise), and $\mathbf{\Psi}(k, \tau)$ is a weighting matrix constructed as

$$\mathbf{\Psi}(k, \tau) = [\psi_1(k, \tau), \psi_2(k, \tau), \ldots, \psi_N(k, \tau)]$$

(19)

and the generic weighting column vector $\psi_n(k, \tau)$ is defined as:

$$\psi_n(k, \tau) = \frac{\psi_n(k, \tau)}{N_T}[1, 1, \ldots, 1]^T_{1 \times N}$$

(20)

Here $T$ indicates the vector transpose, $N_T$ is the number of time frames on which the gradient is averaged and each $\psi_n(k, \tau)$ is a weight indicating the plausibility that the $n$-th source is active in the given time-frequency point. If the weights $\psi_n(k, \tau)$ are constrained in the range between 0 and 1, according to the definition in (20) the resulting gradient in (18) is equivalent to a weighted average over time. However, if the activity of the sources is not uniformly distributed over time, for some sources the gradient may be underestimated. To solve this drawback, the weighting vector can be improved by normalizing the weights as:

$$\psi_n(k, \tau) = \frac{\psi_n(k, \tau)}{N_T \sum_{\tau} \psi_n(k, \tau)}[1, 1, \ldots, 1]^T_{1 \times N}$$

(21)

Under the assumption of ideal sparseness, for each point $(k, \tau)$ all the weights $\psi_n(k, \tau)$ but one must be equal to 0, which means that only one source is assumed to be active. If the ideal assumption of sparseness is relaxed, a more accurate model is obtained imposing $\psi_n(k, \tau)$ to be a continuous weight which indicates the dominance of the $n$-th source over the others. An estimate of $\psi_n(k, \tau)$ can be obtained through a Bayesian estimation framework (see Appendix), by means of the gain used to compute the minimum mean-square-error estimation of $y_n(k)$ in the sum $\sum_q y_q(k, \tau)$ [29].
Assuming the sources to be normally distributed, for the sake of simplicity, the weights are obtained as:

$$\psi_n(k, \tau) = \frac{E[|y_n(k, \tau)|^2]}{E[|y_n(k, \tau)|^2] + \sum_{q \neq n} E[|y_q(k, \tau)|^2]}$$  (22)

One can note that the weights defined by (22) are similar to the Wiener-like scaling weights used in Acoustic Echo Cancellation (AEC) to regularize the Least Mean Square (LMS) adaptation in presence of local noise [30]. In fact, also in this case the weighting reduces the error in the gradient estimation, which is introduced by noisy instantaneous observations related to the presence of interfering sources.

Assuming no permutation occurs, and by exploiting the spectral smoothness of acoustic sources, the source energy can be recursively estimated across frequencies. For instance, $E[|y(k, \tau)|^2]$ can be estimated from the highest to the lowest frequency bins, based on an autoregressive moving average, as:

$$E[|y(k, \tau)|^2] = \alpha E[|y(k + 1, \tau)|^2] + (1 - \alpha)|y(k, \tau)|^2$$  (23)

where $\alpha$ is a smoothing factor (e.g. $\alpha = 0.9$).

The described weighting procedure can be applied only if the ICA adaptation is formulated in terms of estimated mixing matrix $\mathbf{H}(k)$ as in (17). Actually, it makes sense to associate a different weight to each column of the gradient since different columns update the mixing coefficients of different sources. On the other hand, this procedure cannot be applied if the adaptation is formulated as in (11). In the general case, the coefficients of the demixing matrix $\mathbf{W}(k)$ are a function of the coefficients of $\mathbf{H}(k)$ and the matrix can not be partitioned to isolate specific sub-matrices (i.e. columns or rows) selectively related to a single source. A particular case in which the weighting can be applied directly to (11) is $N = 2$. In this case it can be shown [5] that each row of $\mathbf{W}(k)$ is associated to the mixing parameters of a different source, and the described weighting procedure can be modified by substituting the matrix in (19) with its transpose.

It is worth noting that the proposed weighting procedure does not attempt to solve the permutation problem, although it was shown that the correlation between weights as $\psi_n(k, \tau)$ can represent a good dependence measure of bin-wise separated signals [31]. Under the ideal assumption of sparseness and of continuity of the source time-activity across the frequency, the columns of the estimated demixing matrices $\mathbf{H}(k)$ are likely to be re-ordered according to the same permutation if the weights are recursively estimated and propagated in frequency. However, the accuracy of the weights can not always be guaranteed since it depends on the output energy estimation as well as on the choice of the parameter $\alpha$. Therefore, the linking between the frequencies provided by the weighting procedure can not be considered as a reliable solution to the permutation problem. Differently, in the proposed algorithm the permutations are effectively reduced by an alternative strategy, described in the subsection IV-C, which exploits the continuity of the demixing matrices. It will be shown in section V that the two strategies complement each other and, when combined together, allow one to achieve the best performance.

**B. Relationship between Wiener-like weighted ICA and Adaptive BeamFormers**

It is interesting to observe that under the ideal assumption of sparse acoustic sources, the Wiener weighting leads to an improved approximation of FD-BSS based on ICA to a set of independent Adaptive BeamFormers (ABF). According to the optimization in (13), the ICA adaptation converges when $E[\Phi(y(k))y(k)^H]$ is diagonalized. The generalized covariance matrix can be approximated with a sum of high-order cross-correlations, in conformity with the Taylor expansion of the non-linearity $\Phi(\cdot)$.

Therefore, for zero-mean random variables the adaptation converges if all the high-order covariance matrices are diagonalized. In particular, considering only second-order-statistics (SOS), for $\alpha = 1$ one should obtain:

$$E[y(k)y(k)^H] = E[\mathbf{W}(k)x(k)x(k)^H\mathbf{W}(k)^H] = \mathbf{W}(k)\mathbf{H}(k)E[\mathbf{R}_{ss}(k)]\mathbf{H}(k)^H\mathbf{W}(k)^H = \mathbf{D}$$  (24)

where $\mathbf{D}$ is a diagonal matrix. It has been shown [4] that the optimization in (24) is equivalent to compute the weights of a set of $N$ adaptive beamformers steered to the direction of each source. Such an equivalence is mainly based on the assumption that the correlation between each pair of source signals is null, and therefore the expectation $E[\mathbf{R}_{ss}(k)]$ is diagonal. However, a longer frame size used for the frequency analysis and a limited amount of data make this hypothesis unlikely; for this reason, the performance of the ICA is drastically reduced with respect to an equivalent adaptive beamformer [4].

Under the assumption of sparseness, the source vector $s(k, \tau) = [s_1(k, \tau), \ldots, s_N(k, \tau)]$ can be approximated with:

$$s(k, \tau) \simeq [0, \ldots, s_{n(\tau)}(k, \tau), \ldots, 0]$$  (25)

where $n(\tau)$ indicates the index of the unique active source, at instant $\tau$, which means that the instantaneous covariance matrix $\mathbf{R}_{ss}(k, \tau)$ has all the off-diagonal elements equal to zero.

In this limit case, according to (22), the weights $\psi_n(k, \tau)$ are equal to 1 in the frames $\tau$ when only the $n$-th source is active. By means of the above-described weighting procedure, the $n$-th column of the gradient $\Delta \mathbf{W}(k)$, which updates the steering weights for the $n$-th source, is computed by averaging the instantaneous gradient over the time instants $\tau$ when the source is active. Therefore, the expectation of $\mathbf{R}_{ss}(k, \tau)$, obtained averaging $\mathbf{R}_{ss}(k, \tau)$ over all the time instants, is a diagonal matrix. This means that the adaptation for the $n$-th source is equivalent to that of an adaptive beamformer when only the $n$-th source is present.

Furthermore, it is worth noting that the gradient is computed by involving all the high-order covariance matrices and thus, compared with a Second Order Statistics (SOS) based ABF, the Wiener-like weighted ICA includes more information in the adaptation.

**C. Least Mean Square tracking of the demixing matrix**

When few data are used, the estimated matrix $\mathbf{W}(k)$ can be modeled as a noisy version of the true demixing matrix $\mathbf{W}(k)$:

$$\mathbf{W}(k) = \mathbf{W}(k) \odot \mathbf{N}(k)$$  (26)
where $\odot$ is the Hadamard product (i.e. element-wise) and $N(k)$ is a noise term representing the uncertainty of the ICA estimator. According to the model (9), we assume that the noise affects more heavily the phase than the magnitude of the generic element of $W(k)$. The noise $N(k)$ and the intrinsic phase distortion introduced by the reverberation in the matrices $W(k)$ are responsible of the local discontinuities, which make a simple recursive initialization not able to avoid local minima and, as a consequence, permutations.

An accurate estimation of $\tilde{W}(k)$ is not possible, due both to the difficulty in modeling statistics of $N(k)$ and to the highly non linear phase distortion introduced by reverberation. However, the risk to converge to local minima can be reduced since the propagation model of the acoustic wave along the direct path exhibits a local continuity across frequencies. In practice, each ICA can be initialized with tracked demixing matrices which vary smoothly with the frequency. By this strategy, a direct estimation of $\tilde{W}(k)$ is avoided, while we assume that the stochastic optimization performed by ICA converges towards a solution very close to the true $\tilde{W}(k)$. Thus, the knowledge about the continuity of the demixing matrices can be included in the ICA adaptation, performing the separation recursively from the highest to the lowest frequency. At each frequency a matrix $W_{\text{tracked}}(k)$ is estimated as a smooth extension from the noisy matrices $W(k)$ observed at previous frequencies, and it is used to initialize the matrix $W_0(k)$ for the ICA of the new frequency. Many methods can be adopted to track a continuous demixing matrix. In this work we propose to estimate $W_{\text{tracked}}$ by constraining its determinant to vary smoothly across frequencies. In fact, the determinant of $W(k)$ is strictly related to the spatial diversity of the sources and varies slowly with frequency under anechoic conditions. However, if for some frequencies the ICA converged to a poor solution, discontinuities both in $W(k)$ and in its determinant would be generated. The initialization of ICA with a smooth estimate of the demixing matrix acts as a regularization procedure which avoids that discontinuities in some $W(k)$ compromise the global recursion. Furthermore, the effect of the discontinuities introduced by the reverberation is reduced.

The tracking can be performed by a simple and computationally inexpensive $\varepsilon$-Normalized Least Mean Square ($\varepsilon$-NLMS) predictor. Figure 1 shows a typical plot of the determinant values, computed for the observed noisy $W(k)$ and the $\varepsilon$-NLMS smoothed versions. In order to reduce the steady-state error, the LMS filter has been implemented with a variable $\varepsilon$ approach as proposed in [32]. For a given frequency bin $k$, in order to remove the scaling ambiguity the observed matrix $W(k)$ is normalized as:

\[
W(k) = U(k)W(k)
\]

where $U(k)$ is defined as:

\[
U(k) = \begin{pmatrix}
u_1(k) & 0 & \cdots & 0 \\
0 & u_2(k) & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & u_N(k)
\end{pmatrix}
\]

The following pseudo code summarizes the main steps of the proposed RR-ICA procedure:

**initialization***

$W_{\text{tracked}(\text{highest_frequency})} = I$

$\Psi(\text{highest_frequency}, \tau) = I, \forall \tau$

**iteration***

for $k = \text{highest_frequency} \text{ to } l$

***wieners_like_weights_estimation***

if $(k < \text{highest_frequency})$

The normalizations in (27)-(29) is necessary to perform a filtering that only accounts for spatial diversity of the matrices, regardless of the presence of different scaling across frequencies. In other terms, the error in (32) only depends on the physical interpretation of the matrix $W(k)$ which is not related to the intrinsic scaling ambiguity of the ICA. Moreover, thanks to this unit norm normalization the adaptation does not get unstable because of outliers that have high magnitude in the coefficients of $W(k)$. The length $L$ of the filter $h(k)$ must be increased according to the degree of discontinuity of the demixing matrices. This discontinuity depends both on the noise $N(k)$ (which is related to the length of the observed signals) and on the DRR conditions. For a sampling frequency $f_s = 16 \text{kHz}$ and an FFT size in the range $2048 - 8192$ points, it was found empirically that a length around $40 - 80$ taps and a step-size around $\mu = 0.01$ are good choices to separate real-word signals of $0.5 - 1$s. Experimental results reported in section V shows that the method is quite stable against variation of the parameters $L$ and $\mu$.

**D. Proposed algorithm**

The following pseudo code summarizes the main steps of the proposed RR-ICA procedure:

**initialization***

$W_{\text{tracked}(\text{highest_frequency})} = I$

$\Psi(\text{highest_frequency}, \tau) = I, \forall \tau$

**iteration***

for $k = \text{highest_frequency} \text{ to } l$

***wieners_like_weights_estimation***

if $(k < \text{highest_frequency})$
estimate \( E[|y(k, \tau)|^2] \) as in (23) and compute \( \Psi(k, \tau) \)
as in (19)

***ICA ADAPTATION***
\[ \Theta_0(k) = W_{\text{tracked}}(k), \quad y(k, \tau) = W_{\text{tracked}}(k)x(k, \tau) \]
compute \( \Theta(k) \) as in (15)-(18), from \( \Theta_0(k) \)

***LMS DE-MIXING MATRIX TRACKING***
\[ W(k) = \Theta(k) \]
if \( k = \text{highest}_\text{frequency} \)
\[ W_{\text{tracked}}(k) = W(k) \]
else
normalize \( W(k) \) as in (27)-(29)
\[ L = \min(L_{\text{max}}, \text{highest}_\text{frequency} - k + 1) \]
estimate \( W_{\text{tracked}}(k) \) as in (27)-(33)

end

At the first (i.e. highest) frequency the ICA iteration is initialized with the identity matrix \( I \) and the resulting demixing matrix is used to initialize \( W_{\text{tracked}}(k) \). At the subsequent frequencies, the smooth matrix \( W_{\text{tracked}}(k) \) is estimated as in (27)-(33), where the length of the filters \( h(k) \) progressively increases from 1 to a maximum value \( L_{\text{max}} \). At each frequency, the weighting matrix \( \Psi(k, \tau) \) is computed as in (19) with the output energy estimated as in (23) by means of the output energy of the previous separated frequencies.

Note that by exploiting the continuity of the demixing matrices across frequency, the recursive approach implicitly avoids permutations when there are no intersections among trajectories of the propagation models which are associated to different sources. In the following experiments, an array of two microphones is used, which leads to trajectories based on monodimensional states (i.e. one phase observation for each frequency). With this geometry, a safe condition to avoid intersections is to limit the distance between the microphones, so preventing spatial aliasing for all the frequencies.

V. EVALUATING THE DEMIXING MATRICES ESTIMATION ACCURACY

This section reports on simulation results that demonstrate how the proposed approach can drastically improve the accuracy of the estimated demixing matrices when short signals are observed. Since we are not interested in the evaluation of the overall performance of the BSS algorithm, a new performance measure is here proposed which is invariant to the permutation and to the scaling ambiguity.

Impulse responses have been generated by the Lehmann & Johansson’s image source method [33] according to the setup shown in Figure 2, where sources and microphones are assumed to be omnidirectional. Microphones are spaced apart of 0.02 m to avoid spatial aliasing for any source position and frequency, given a sampling frequency of \( f_s = 16 \text{kHz} \). Two clean speech utterances are convolved with the resulting impulse responses in order to generate the signals received at the two microphones. Then the mixture signals are obtained as sum, over each channel, of the signals generated for all the sources.

For each frequency bin, the demixing matrices are obtained by applying a Natural Gradient ICA to the time-series produced by a short-time Fourier analysis with windows of 4096 points and overlapping factor of 75%. To better show the robustness of the proposed method, the time observations used by ICA were obtained analyzing only a signal portion of 1 second. In the following tests, the algorithm is parameterized with \( \eta = 0.1, L=40, \mu = 0.01 \) and \( \alpha = 0.95 \), where no other values are specified.

The polar diagram in Figure 3 shows the phase of the states \( r_n^{(1,2)}(k) \) associated to the matrix \( W(k) \) computed without any recursion. The state trajectories related to the matrices \( W(k) \) (points) and \( W_{\text{tracked}}(k) \) (solid lines).

Fig. 3. State trajectories associated to \( W(k) \).

(a) Matrix \( W(k) \) computed without any recursion. (b) Matrix \( W(k) \) computed with the recursion.

Fig. 4. Histogram of the observed \( \Delta r_n^{(1,2)}(k) \) associated to \( W(k) \). The red dotted lines are the true expected TDOAs.

The dispersion of the final states is smaller when the
recursion is applied than in the case where ICA is performed independently without any recursion.

A numerical evaluation of this experimental evidence can be obtained by defining a Propagation Model Error (PME), which measures the mean square error between estimated and true propagation models. Since we are considering closely-spaced microphones, the attenuation ratios of the acoustic wave can be neglected and the observed propagation models are normalized as follows:

\[
\hat{r}_{(1,2)}(n)(k) = \frac{r_{(1,2)}(n)(k)}{|r_{(1,2)}(n)(k)|}\] (35)

In the same way, one can compute the true normalized propagation models:

\[
\sigma_{(1,2)}(n)(k) = \frac{o_{(1,2)}(n)(k)}{|o_{(1,2)}(n)(k)|}\] (36)

where \(o_{(1,2)}(n)(k)\) are the models obtained by the simulated impulse responses \(h_{mn}(k)\) as:

\[
o_{(1,2)}(n)(k) = \frac{h_{1n}(k)}{h_{2n}(k)}\] (37)

Then the error for the \(n\)-th source is computed as follows:

\[
PME_n = 10 \cdot \log_{10} \left( \sum_{k=1}^{N_{bin}} \frac{\left| \hat{r}_{(1,2)}(n)(k) - \sigma_{(1,2)}(n)(k) \right|^2}{N_{bin}} \right)\] (38)

where \(N_{bin}\) is the number of frequency bins of the FFT analysis. Note that PME is similar to the formulation of the frequency-domain misalignment used to measure the accuracy of AEC systems. It is important to underline that applying a direct distance measure to estimated and true impulse responses would not be meaningful in the BSS case. In fact, ICA is still affected by a scaling ambiguity that cannot be completely removed and does not allow to estimate the true impulse responses. On the other hand, the ratios \(r_{(1,2)}(n)(k)\) are scaling invariant, and thus PME allows a better evaluation of the accuracy of the estimated demixing filters. The performance is evaluated in two cases:

1) with an ideal permutation correction. The matrices \(W(k)\) are permuted with the optimal permutation which minimizes the total error:

\[
\Pi(k) = \arg \min_{\Pi} \sum_{n} |r_{n}(\Pi)(k) - \sigma_{n}(k)|^2\] (39)

where \(r_{n}(\Pi)(k)\) are the states obtained from the matrix \(W(k)\) after the permutation \(\Pi\);

2) without any permutation correction.

The performance is evaluated averaging the results over 15 different mixtures, which are generated by considering all the possible combinations of 6 sentences uttered by italian
speakers (3 males + 3 females).

The benefit of the recursive regularization can be better observed by computing PME with a different maximum number of iterations in the ICA adaptation. This result is shown in Figure 5(a), for PME (averaged over both sources) obtained with each of the following ICA strategies: a) Natural Gradient (NG) b) Natural Gradient with the recursive regularization c) Scaled Natural Gradient [34] d) Scaled Natural Gradient with the recursive regularization. The Scaled Natural gradient is an efficient improvement of the Natural Gradient which imposes a posteriori unit norm constraint in the generalized covariance matrix $E[\Phi(y_{(k)}(i))y_{(k)}(i)\mathbf{H}]$. Such a normalization improves and stabilizes the convergence of the Natural Gradient. It is implemented in the algorithm in section IV-A by modifying the equations (16) and (17) as:

$$\Delta \mathbf{H}_{(i)}(k) = \mathbf{H}_{(i)}(k)(I - d_{(i)}(k)^{-1}E[\Phi(y_{(k)}(i))y_{(k)}(i)\mathbf{H}])$$

$$\mathbf{H}_{(i+1)}(k) = c_{(i)}(k)^{-1}\mathbf{H}_{(i)}(k) - \eta\Delta \mathbf{H}_{(i)}(k)$$

where the scaling factors $d_{(i)}(k)$ and $c_{(i)}(k)$ are computed as in [34] with respect to the chosen nonlinear function $\Phi(\cdot)$.

A first analysis considers the PME results when the ideal permutation correction is applied. Figure 5(a) shows that the recursive approach always reduces PME for any value of Max iterations which means that the deterministic constraint improves the accuracy of the demixing matrices estimation. Moreover, both of the strategies b) and d) are able to obtain a good solution with a very small number of iterations. This confirms that the smoothed demixing matrix estimated across the frequencies is a good approximation of the optimal demixing matrix; thus, few iterations are sufficient at each frequency to refine the solution. The strategy d) is the best in terms of convergence speed, since it also exploits the advantages of the gradient normalization provided by Scaled Natural Gradient. Therefore, the solution d) is a convenient method to reduce significantly the computational cost of the global adaptation and it makes a corresponding real-time implementation feasible.

A second analysis concerns the PME results obtained without any ideal permutation correction. Figure 5(b) shows that PME is low only when the recursive strategy is applied, which indicates that the permutation problem is less critical when the deterministic knowledge is propagated through the recursion. Nevertheless, it can be observed that PME is about 2dB higher than in the previous case, which means that the permutations are not completely solved. In fact, the deterministic constraints imposed in the recursion are based on approximated source and mixing models. These constraints increase the robustness of the demixing matrices estimation but, on the other hand, they also limit the optimality of the solution when the model
does not fit well the observed data.

Figure 6 shows the separated effects of the Wiener-like weighting (method in section IV-A) and of the LMS tracking of the demixing matrix (method in section IV-C). Both the strategies reduce PME when compared with standard ICA. However, without any ideal permutation correction only the LMS tracking of the demixing matrix gives a low PME while the Wiener-like weighting has a poor performance, which means that there are still many wrong permutations. As previously discussed, the Wiener-like weighting alone can not be effective in reducing the permutation problem and its performance can change considerably with the parameter $\alpha$ (see figure 7). On the other hand the permutations are reduced by the LMS tracking of the demixing matrix (see figure 8) which is also more stable against variation of its parameters ($\mu$ and $\alpha$).

An increased number of iterations does not necessarily correspond to a lower PME; in fact, the minimum is obtained with 10-20 iterations. Note that the ICA optimization is still unconstrained at each frequency, and thus the final solution depends only on the observed data. When the number of iterations is limited, the final solution of ICA is more dependent on the initialization provided by $W_{\text{tracked}}$ and the filters are induced to comply with continuity across frequency. Therefore, the effect of discontinuities due to poor solutions in the ICA optimization is reduced.

The recursive regularization with the Wiener weighting (section IV-C) reduces PME, since the accuracy of the ICA solutions is increased by the limited statistical bias. However, since each ICA is always initialized with an identity matrix, no deterministic knowledge about the structure of the filters is exploited and the convergence speed is slow. Combining both strategies, the deterministic knowledge is better exploited and PME further diminishes to values less dependent on the number of the ICA iterations.

VI. EXPERIMENTAL RESULTS ON REAL DATA

The proposed algorithm has been tested in Matlab and implemented in C++. On a standard laptop (Pentium IV 2.0 GHz), it is able to estimate in real-time demixing filters of more than 8192 taps to separate two sources in adverse conditions. The block diagram in Figure 9 summarizes the main steps of the algorithm.

As a first step, a short-time Fourier analysis is performed in order to obtain a time-frequency representation of the observed mixtures $x(t)$. For each frequency bin, the demixing matrices $W(k)$ are obtained by using the Scaled Natural Gradient [34]. ICA is applied recursively from the highest to the lowest frequency according to the estimation/initialization method explained in Section IV-D.

In this evaluation, two sources were recorded at $f_s = 16$kHz with two microphones spaced of $d = 0.02$ m. In such conditions, no frequencies are affected by spatial aliasing and the trajectory intersections are avoided, thus minimizing the risk of permutations. In the more general case of a larger microphone spacing, however, the permutations problem can be solved for higher frequencies by using the State Coherence Transform as explained in [35]. The Minimal Distortion Principle (MDP) [36] is applied to solve the scaling ambiguity and the smoothing method proposed in [37] is adopted in order to reduce the spikes due to the circularity effect of FFT. By means of the inverse Fourier transform, the resulting demixing matrices are finally used to obtain the demixing filters in the time-domain.

According to the FFT window size, the length of the filter used by the $\epsilon$-NLMS procedure was chosen between 40 and 80 taps, while the step-size was set to 0.01. As for the Scaled Natural Gradient, the fixed step-size was set to 0.1.

In the following, the given performance regards a real case of two loudspeakers playing clean speech utterances. Two tests have been realized by recording the sources according to the configurations showed in Figure 10. The proposed algorithm was compared with other popular BSS ones, namely:

- ALG1: the Independent Vector Analysis (IVA)[17], parameterized with step size 0.1 and a maximum number of 1000 iterations;
- ALG2: the time-domain Parra’s method [38], applied by using time-domain filters whose size is half of the chosen FFT frame size, a number of matrices to diagonalize equal to 5, and a maximum number of 1000 iterations;
- ALG3: the frequency-domain Pham’s algorithm based on [12] and [20], parameterized with a FFT overlapping factor equal to 75% and a window size equal to 5.

Performance is evaluated with the BSS_EVAL toolbox, which
is based on the criteria proposed in [39], using time-invariant filters of 1024 taps to represent the family of allowed distortions. The Source-to-Interferences Ratio (SIR) and Source-to-Distortion Ratio (SDR) are evaluated using the whole separated signals (whose length is about 9s); however, the demixing filters are computed using time observations within portions of different length ranging from 0.5 to 9 s. Performance is averaged over 15 speaker combinations (3 males + 3 females) and for all the methods. The experiments were conducted by using FFT analysis of different window size. All the related results are reported in Tables (I-VIII) in terms of average performance and standard deviation over all the separated sources. In case of divergence of the given algorithm, the table reports it by means of div. The corresponding audio signals are available in http://shine.itc.it/people/nesta/testBSS.htm.

To show the given experimental evidence in a more compact representation, a summary analysis is reported based on: 1) best performance over the FFT size; 2) average performance over FFT size and signal length. In all the cases the results do not take into account output signals for which the algorithms diverged. The best performance reported in Figures 11 and 12, for Test1 and Test2 configurations, respectively, shows that the proposed method performs well for any signal length. It is also interesting to observe that with just 500ms the separated sources have SIR and SDR almost equal to half of the optimal values obtained with 9 s of data. Hence, the proposed algorithm produces separated signals of very good quality and characterized by no audible distortions, even when a small amount of data has been observed. Moreover, the standard deviation (indicated by the black lines around the bars) indicates that the proposed method is stable and the performance does not vary a lot with different signal lengths. The IVA method (ALG1), which is an attractive extension of ICA for multivariate distributions, provides good performance for some test files when the observed signals are long enough. The IVA optimization directly exploits inter-frequency dependencies to avoid permutations and its optimization is jointly performed for all the frequencies. However, the convergence to the optimal solution is hampered by the presence of local minima which reduce its robustness, similarly to time-domain approaches. Indeed its performance is not stable and, as shown by the high standard deviation, varies a lot with different speaker combinations. The time-domain method ALG2 is more stable but did not produce an acceptable average performance. The frequency-domain method ALG3 produced better results for Test2 and for longer signals. However, the inter-frequency correlation method used in ALG3 to solve the permutation problem produces (similarly to the IVA) an unstable behavior across different speaker combinations (see the high standard deviation in all the tests) and does not provide a good average performance when applied to separate short signals (less than 2 seconds in Test2).

Based on the average performance, one can observe that the proposed method is stable against variation of parameters and speaker combinations. On the other hand, the performance obtained in Test1 for longer signals indicates that the method is stable (i.e. small standard deviation) but also suboptimal (i.e. small average performance). This is due to the regularization introduced to increase the overall robustness, which also hampers the convergence of ICA toward the optimal solution when the underlying mixing model does not fit the true conditions. Such a sub-optimality is strictly related to the remaining permutation errors which cannot be completely avoided by the regularization (see discussion in section V). When a
sufficient amount of data is available, BSS algorithms without any constraint on the mixing model (as IVA and Pham’s methods) can better approach the optimality, though with a less stable behavior. With this regard, future investigation may concern the combination of the proposed method with traditional unconstrained FD-BSS to achieve both stability and optimality.

Figures 13(a) and 13(b) show the average performance for any FFT sizes and signal lengths. Results confirm that the proposed method provides the best overall performance and the best stability across different parameter settings. Based on an increased overall robustness, this improvement is even more evident in the case of Test2. This is allowed by the stability properties of the source separation algorithm when long demixing filters are adopted to tackle a higher reverberation time.

The proposed algorithm has been used to separate some of the test files proposed in the last Signal Separation Evaluation Campaign (SISEC 2008 [40], see http://sassec.gforge.inria.fr/SiSEC_determined), without any further permutation correction. In fact, the propagation model trajectories do not intersect and no further permutation correction is needed for the case of tests referred as "Room4" and "Room5". The room used in both the tests is a chamber with cushion walls and has size of L3.55m × W4.45m × H2.5m. The sources were located at around 1m for "Room4" and 1.8m for "Room5". Figure 14 shows the performance comparison with other algorithms benchmarked by that evaluation campaign, which represents a further demonstration of the validity of the proposed method.
ACKNOWLEDGMENT
The authors want to thank the anonymous reviewers for providing many valuable suggestions.

REFERENCES


<table>
<thead>
<tr>
<th>Table I</th>
<th>Performance Proposed (Test)</th>
<th>SR/B (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table II</th>
<th>Performance ALGO Algorithm (Test 1)</th>
<th>SR/B (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table III</th>
<th>Performance Proposed (Test 2)</th>
<th>SR/B (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table IV</th>
<th>Performance ALGO Algorithm (Test 2)</th>
<th>SR/B (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table V</th>
<th>Performance Proposed (Test 3)</th>
<th>SR/B (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table VI</th>
<th>Performance ALGO Algorithm (Test 3)</th>
<th>SR/B (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table VII</th>
<th>Performance Proposed (Test 4)</th>
<th>SR/B (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table VIII</th>
<th>Performance ALGO Algorithm (Test 4)</th>
<th>SR/B (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16384</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>