Dereverberation of acoustic signals by 
Independent Component Analysis

C. Sistopodi, R. Parisi and A. Uncini

INFOCOM dept. - University of Rome "La Sapienza"
Via Eudossiana 18, 00184 Rome - Italy
email parisi@infocom.uniroma1.it

Abstract. In this paper joint use of array processing techniques and 
Independent Component Analysis is proposed to separate acoustic sig-
nals in the presence of reverberation. The estimated time delay of arrival 
is employed as an index of performance. to show that the described 
approach is effective in performing the dereverberation of the received 
signals, as confirmed by several experimental tests in different environ-
mental conditions.

1 Introduction

Independent Component Analysis (ICA) has raised the interest of many re-
searchers in recent years, due to the high number of potential applications (e.g. 
signal and image enhancement, analysis of medical signals, source separation in 
telecommunications). Generally speaking, the goal of ICA is to recover original 
signals by observation of a mixture, having no knowledge of the mixing func-
tions. In the simplest model, mixing is instantaneous (i.e. no convolutions are considered), thus neglecting any delays, reflections or noises.

An extensive literature on ICA is available. An efficient and popular al-
gorithm was proposed in [1], and subsequently improved by Amari [2]. The 
proposed approach was able to separate up to ten mixed acoustic signals. Un-
fortunately, the instantaneous mixture model is not realistic when dealing with 
the problem of acoustic signal separation in reverberating environments. In this 
case, in fact, signals are convolved with the impulse response of the environment, 
which takes into account the effects of reflections and diffusion. The study of ICA 
in the presence of convolving mixtures thus requires a higher level of complexity.

A recent approach [3] proposed the joint use of array processing techniques 
and ICA to separate multiple acoustic sources in the presence of reverberation. 
In this work the algorithm presented in [3] is properly modified to separate 
signals at different reverberation levels. In particular, the time delay of arrival 
(TDOA) is introduced as a useful performance index in experiments, to show 
that the proposed approach has the effect of actually dereverberating the received 
signals.
2 Separation algorithm

2.1 Signal model

Received signals include ambient noise, directional or not, and reflections from the walls of the room and other objects. Assuming $M$ microphones and $D$ sources, the following model in the frequency domain is commonly adopted:

$$\mathbf{x}(\omega, t) = \mathbf{A}(\omega) \cdot \mathbf{s}(\omega, t) + \mathbf{n}(\omega, t)$$

(1)

where $\mathbf{x}$ is the vector of the short-time Fourier transform (STFT) of the received signals, $\mathbf{s}$ is the STFT vector of the source signals, $\mathbf{A}$ is the mixing matrix and $\mathbf{n}$ is the STFT noise vector.

$\mathbf{A}$ is an $M \times D$ matrix, and its $(m, n)$ element, $A_{m,n}(\omega)$, is the transfer function from the $n$-th source to the $m$-th microphone

$$A_{m,n}(\omega) = H_{m,n}(\omega)e^{-j\omega \tau_{m,n}}$$

(2)

In this equation, $\tau_{m,n}$ is the propagation delay from the $n$-th source to the $m$-th microphone. It can be noted that eq. (1) is formally similar to the case of instantaneous mixtures, if noise components are negligible. So, proper noise reduction techniques can be devised in order to make it possible the application of instantaneous separation algorithms to each frequency bin. A possible approach is the subspace method, quite popular in the array processing field, which is briefly reviewed in the following [3].

2.2 The Subspace method

The subspace method is based on a proper decomposition of the spatial correlation matrix, defined as

$$\mathbf{R}(\omega) = E[\mathbf{x}(\omega, t)\mathbf{x}^H(\omega, t)]$$

(3)

Assuming uncorrelation between signal and noise and omitting $\omega$ for simplicity, from (1) it is possible to obtain

$$\mathbf{R} = \mathbf{AQA}^H + \mathbf{K}$$

(4)

where $\mathbf{Q} = E[\mathbf{s}(t)\mathbf{s}^H(t)]$ and $\mathbf{K} = E[\mathbf{n}(t)\mathbf{n}^H(t)]$.

The hypothesis of uncorrelation between $\mathbf{s}$ and $\mathbf{n}$ is not strictly verified, since $\mathbf{n}(t)$ includes the reflections of $\mathbf{s}(t)$. However, if the STFT window is short enough and the delay time between the direct signal and the first reflection exceeds the window length, it can still be considered valid.

Computation of the generalized eigenvalue decomposition of $\mathbf{R}$ [4] leads to

$$\mathbf{R} = \mathbf{KEAE}^{-1}$$

(5)

where $\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_M]$ is the eigenvector matrix, and $\mathbf{A} = \text{diag}(\lambda_1, \ldots, \lambda_M)$ is the diagonal matrix of the eigenvalues. Assuming that the signal power is high compared with the noise power, eigenvectors and eigenvalues have the following properties, that we recall from [3]:

- The eigenvectors are orthogonal to each other.
- The eigenvalues are real and non-negative.
- The eigenvalues are sorted in descending order.
- The first eigenvalue is the largest and corresponds to the signal component.
- The remaining eigenvalues correspond to the noise components.

These properties are exploited in the subspace method to separate the different sources.
1. The energy of the $D$ directional components of the signal $s(t)$ is concentrated in the $D$ dominant eigenvalues.
2. The noise energy is uniformly distributed on all eigenvalues.
3. The eigenvectors $(e_1, \ldots, e_D)$, corresponding to the dominant eigenvalues, constitute an orthonormal basis of $\mathcal{R}(A)$, where $\mathcal{R}(A)$ is the column space of $A$ [4].
4. Eigenvectors $(e_{D+1}, \ldots, e_M)$ constitute a basis of $\mathcal{R}(A)^\perp$, where $\mathcal{R}(A)^\perp$ is the orthogonal complement of $\mathcal{R}(A)$.

Subspaces $\mathcal{R}(A) = \mathcal{R}(E_s)$ and $\mathcal{R}(A)^\perp = \mathcal{R}(E_n)$ are called the signal subspace and the noise subspace respectively, being $E_s = [e_1, \ldots, e_D]$ and $E_n = [e_{D+1}, \ldots, e_M]$.

Properties 1 e 3 correspond to assume that directional components lie into the signal subspace, while noise is fairly distributed over all frequencies.

The subspace filter is finally defined as

$$W = A^{-1/2}E_s^H$$

and processed signals are given by

$$y(\omega, t) = W(\omega)x(\omega, t)$$

### 2.3 Scaling and permutation

Since the separation algorithm is applied separately to every frequency bin, it is fundamental the determination of the correct amplitude and the correct permutation of the separated signals in order to correctly reconstruct the signals. In [3] the amplitude problem is solved by multiplying the output signals $u_\omega(t_s)$ by the pseudoinverse of the mixing matrix. If

$$u(\omega, t) = B(\omega)x(\omega, t)$$

is the unmixing equation (where matrix $B(\omega)$ takes into account the effects of both the subspace filtering and the ICA algorithm), the amplitude ambiguity can be solved by use of a proper scaling matrix $B^\sharp_m(\omega)$ [3], where symbol $\sharp$ indicates pseudoinversion [4].

In order to solve the permutation problem, in [3] the Inter-Frequency Coherence (IFC) method is proposed. This technique is based on the consistency of the mixing matrix in adjacent frequencies. Assuming for the mixing matrix $A(\omega)$ the model (1) and assuming for simplicity $H_{m,n}(\omega) = 1$, the steering vector $a_n$ is given by

$$a_n(\omega) = [e^{-j\omega \tau 1_n}, \ldots, e^{-j\omega \tau Mn}]^T$$

At the adjacent frequency $\omega_0 = \omega - \Delta \omega$ it is

$$a_n(\omega_0) = [e^{-j(\omega - \Delta \omega) \tau 1_n}, \ldots, e^{-j(\omega - \Delta \omega) \tau Mn}]^T$$
It can be seen that the location vector \( \mathbf{a}_n(\omega) \) is equal to \( \mathbf{a}_n(\omega_0) \) rotated by an angle \( \theta_n \). So, matrix \( \mathbf{A}(\omega) \) can be written as

\[
\mathbf{A}(\omega) = \mathbf{T}(\omega, \omega_0) \cdot \mathbf{A}(\omega_0)
\]  

where \( \mathbf{T} \) is a rotation matrix. If the STFT frequency resolution \( \Delta \omega \) is small enough, also angles \( \theta_n \) are small and \( \mathbf{T}(\omega, \omega_0) \approx \mathbf{I} \) holds.

Indicating with \( \tilde{\mathbf{a}}_n(\omega) \) the column vector of the estimated mixing matrix \( \tilde{\mathbf{A}}(\omega) \), it is

\[
\cos \theta_n = \frac{\tilde{\mathbf{a}}_n^H(\omega) \cdot \tilde{\mathbf{a}}_n(\omega_0)}{||\tilde{\mathbf{a}}_n(\omega)|| \cdot ||\tilde{\mathbf{a}}_n^H(\omega_0)||}
\]  

The following cost function \( F(\mathbf{P}) \) is introduced

\[
F(\mathbf{P}) = \frac{1}{D} \sum_{n=1}^{D} \cos \theta_n
\]  

In order to avoid that an error in a frequency bin induces errors on the choice of the permutation for next frequencies, the reference frequency \( \omega_0 \) is extended to an interval \( \omega_0 = \omega - k \cdot \Delta \omega \) for \( k = 1, \ldots, K \). The value of the cost function at \( \omega_0 \) is denoted by \( F(\mathbf{P}, k) \) and is properly exploited to determine the best possible permutation. More details can be found in [3].

3 Experimental results

Two sources were considered in the experiments. The image method [8] was used to simulate the room response for different reverberation times [7]. The room size was 5.45m \( \times \) 4.15m \( \times \) 2.80m and 6 microphones were supposed to be placed on a wall, 20 apart from each other, at a height of 1.7m. Sources were placed in front of the microphone array, at a distance of one meter.

Various experiments were performed, with different kinds of sources. The sampling rate was 16kHz.

Fig. 1 shows the overall structure of the proposed separation algorithm.

![Fig. 1. Algorithm flowchart](image)

Input signals were first centered and whitened. STFT was then computed, using a Hamming window of 32ms and fifty percent overlapping. The FFT length was 512, as a compromise between performance and complexity.
The subspace filter was then applied, to remove noises and reflections and to reduce the number of signals to be processed. Finally ICA was performed, by use of the well-known Amari’s algorithm [2]. Amari’s algorithm was adopted in this work, since it does not require any matrix inversion and it yields a fast convergence rate. In particular, Amari’s rule was properly adapted to take into account the complex nature of quantities of interest. In fact, Hermitian transposition was used instead of simple matrix transposition and the following activation function $f(\cdot)$ was adopted [5]

$$f(z) = \tanh(\text{Re}(z)) + j\tanh(\text{Im}(z))$$ (14)

The filtering matrix was updated by use of a learning coefficient $\eta$, whose choice is critical for the success of the algorithm. In this application the value $\eta = 10^{-4}$ was chosen.

After solving the amplitude and permutation ambiguities, the inverse STFT was applied to yield the unmixed signals.

Several experiments were performed, at different reverberation levels. Specifically, gaussian noise, speech signals and music instruments were considered as sources, in various combinations. The reverberation time $T_R$ was varied from 0 to 0.5 seconds.

As a performance index, the relative time delay of arrival (TDOA) between the separated signals on different pairs of microphones was adopted. It is well-known that in the presence of reverberation cross correlation methods often fail in determining the correct TDOA [9]. In this case generalized cross correlation (GCC) is usually employed [10]. In the present work, in order to analyze the performance of the proposed solution in terms of its dereverberation effects, the delay times estimated by the PHAT-GCC algorithm [10] applied on separated signals and on single reverberated signals were compared.

Figures 2 show the histograms of the estimated TDOA for a single microphone pair in the presence of two speech signals and $T_R = 0.3$. The vertical line indicates the true TDOA. Figure on the left shows the TDOA estimates obtained in the presence of the first signal alone, while figure on the right shows the TDOA histogram obtained from the same signal after separation. The improvement is clearly visible. Figure 3 shows the TDOA statistics in terms of

![Fig. 2. TDOA histograms in the case of two speech signals and $T_R = 0.3$](image-url)
number of anomalies, bias and standard deviation of the estimate, in the case of
separation of one gaussian and one speech signals. The TDOAs are estimated for
the gaussian signal, alone and after separation, at different reverberation times.
More specifically, a TDOA estimate is an anomaly (or outlier) when it exceeds
a prespecified threshold [9]. The number of anomalies is usually referred to in
TDOA estimation problems as a measure of an algorithm's robustness. Bias and
development standard are evaluated on all non anomaly estimates. Also in this case
the dereverberating effects of the proposed algorithm are evident.

**Fig. 3.** TDOA statistics for gaussian and speech signals. Left to right: no. of anomalies,
bias and standard deviation. Dashed line: gaussian signal alone, solid line: gaussian
signal after separation

**References**

[1] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind sep-

separation. In MIT Press, editor, *Advances in Neural Information Processing*, vol-

array processing and independent component analysis for blind separation of


Int. Workshop on Independent Component Analysis and Blind Signal Separation


tion in the presence of room reverberation. *IEEE Trans. on Speech and Audio

24, no. 4, Aug. 1976.