Proposals for Performance Measurement in Source Separation

Rémi Gribonval, Laurent Benaroya, Emmanuel Vincent, Cédric Févotte

4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

PROPOSALS FOR PERFORMANCE MEASUREMENT IN SOURCE SEPARATION

Rémi Gribonval *Emmanuel Vincent* *Cédric Févotte*
Laurent Benaroya

IRISA, METISS Project
Campus de Beaulieu
F-35042 RENNES CEDEX
FRANCE
remi.gribonval@irisa.fr

IRCAM, Analysis-Synthesis Group
1, place Igor Stravinsky
F-75004 PARIS
FRANCE
emmanuel.vincent@ircam.fr

IRCCyN, ADTS Group
1, rue de la Noë – BP 92 101
F-44321 NANTES CEDEX 03
FRANCE
cedric.fevotte@irccyn.ec-nantes.fr

ABSTRACT

In this paper, we address a few issues related to the evaluation of the performance of source separation algorithms. We propose several measures of distortion that take into account the gain indeterminacies of BSS algorithms. The total distortion includes interference from the other sources as well as noise and algorithmic artifacts, and we define performance criteria that measure separately these contributions. The criteria are valid even in the case of correlated sources. When the sources are estimated from a degenerate set of mixtures by applying a demixing matrix, we prove that there are upper bounds on the achievable Source to Interference Ratio. We propose these bounds as benchmarks to assess how well a (linear or nonlinear) BSS algorithm performs on a set of degenerate mixtures. We demonstrate on an example how to use these figures of merit to evaluate and compare the performance of BSS algorithms.

1. INTRODUCTION

In this paper we address some issues related to the evaluation of the performance of Blind Source Separation algorithms. Source separation is a problem that arises when one or several sensor(s) record data to which can contribute several generating physical processes. Perhaps the most striking example of BSS problem consists in recovering the contributions of several musical instruments to a stereophonic recording. If we denote by $s_n(t)$ the signal emitted by the n-th instrument ($1 \leq n \leq N$) and $x_m(t)$ the data recorded on the m-th channel of the recording (in the stereophonic case $1 \leq m \leq M = 2$), we can make the (simplistic) instantaneous linear mixture model

$$x_m(t) = \sum_n a_{m,n} s_n(t), \qquad 1 \leq m \leq M$$

and try to recover $s_n(t)$ from $x_1(t), x_2(t)$. More generally, Blind Source Separation (BSS) consists in recovering N unknown sources $\{s_n(t)\}_{n=1}^N$ from M instantaneous mixtures $\{x_m(t)\}_{m=1}^M$, and we can use the convenient matrix notation for the instantaneous linear mixture model

$$\begin{bmatrix} x_1(t) \\ \dots \\ x_M(t) \end{bmatrix} = \mathbf{A} \begin{bmatrix} s_1(t) \\ \dots \\ s_N(t) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ \dots \\ n_M(t) \end{bmatrix}$$

where $\mathbf{A}$ is the $M \times N$ mixing matrix and $n_k(t)$ are additive noise signals which will always be assumed to be mutually de-correlated and de-correlated from all sources. Note that as a general notation in this paper, we will use bold letters to denote variables that are "multichannel", such as the vector of observations $\mathbf{x}(t) := [x_1(t), \dots, x_M(t)]^T$, the vector of sources $\mathbf{s}(t) := [s_1(t), \dots, s_N(t)]^T$, or the mixing matrix $\mathbf{A}$, and plain letters to denote variables that correspond to only one channel, such as $s_n(t)$.

Considering the case of discrete signals of T samples ($T \gg \max(M, N)$), with the (unrealistic) assumption that there is no noise, BSS can be seen as a factorisation problem $\mathbf{x} = \mathbf{A}\mathbf{s}$: the $M \times T$ matrix $\mathbf{x} := [\mathbf{x}(1), \dots, \mathbf{x}(T)]$ should be factored into the $M \times N$ matrix $\mathbf{A}$ and the $N \times T$ matrix $\mathbf{s} := [\mathbf{s}(1), \dots, \mathbf{s}(T)]$. This is obviously an ill-posed problem, and its solution cannot be defined without additional assumptions, either on the sources (*e.g.* independence) or the mixing matrix.

The most widely studied BSS situation is the non degenerate case where there is at least as many mixtures as there are sources, *i.e.* $M \geq N$. In this case, estimating the mixing matrix $\mathbf{A}$ is sufficient to get an estimate of the sources, and the standard methods (see [1] and the references therein) have essentially the following structure : an

This work is part of a Junior Researchers Project funded by GdR ISIS (CNRS). See `http://www.ircam.fr/anasyn/ISIS/` for some insights into the Project.

estimate $\widehat{\mathbf{A}}$ of the mixing matrix is obtained, by optimising some contrast function which is generally highly nonlinear; the (pseudo)inverse $\widehat{\mathbf{B}} := \widehat{\mathbf{A}}^{\dagger} = (\widehat{\mathbf{A}}^H \widehat{\mathbf{A}})^{-1} \widehat{\mathbf{A}}^H$ of the estimated mixing matrix is applied to the mixtures to estimate the sources as $\widehat{\mathbf{s}} := \widehat{\mathbf{B}}\mathbf{x}$. If there is no noise ($\mathbf{n} = 0$) and a perfect estimate of $\mathbf{A}$ is available, these methods provide perfect recovery of the sources. In general, there are however intrinsic limitations to the accuracy of the estimation of the mixing matrix $\mathbf{A}$ [2].

In this paper we are particularly interested in the degenerate case $M < N$. In this case, estimating the mixing matrix $\mathbf{A}$ is not sufficient to estimate the sources, because (as noted in [3]) the equation $\mathbf{x} = \widehat{\mathbf{A}}\mathbf{s}$ has an affine set of solutions. A preferred solution $\widehat{\mathbf{s}}$ is selected in this affine set using (probabilistic) prior models of the sources. The selection usually involves nonlinear steps that can introduce nonlinear distortions in the estimate of the sources.

In degenerate demixing, the accuracy of a BSS algorithm cannot be assessed only from its ability to estimate the mixing matrix. It becomes of particular importance to measure how well BSS algorithms estimate the sources with adequate criteria. In Section 2 we address this simple topic that seems to have been a bit overlooked in the literature. We define several measures of distortion of the sources that take into account the well known gain indetermination. Besides a global measure of distortion, we also introduce the notions of *Source to Interference Ratio*, of *Source to Noise Ratio* and *Source to Artifacts Ratio*.

Based on the above distortion measures and an appropriate database, a detailed evaluation of the performance of a given (linear or nonlinear) BSS algorithm can be assessed, and different algorithms can be compared. While "good" estimators of unknown sources from degenerate mixtures are necessarily nonlinear, the performance of linear estimators can indicate how difficult is a given degenerate problem. In Section 3, we study linear separation, *i.e.* separation performed by applying a separation matrix (which may be computed by nonlinear methods) to the mixtures. Such a separation consists in estimating the sources by simple linear combinations $\widehat{s}_n(t) = \sum_m b_{n,m} x_m(t)$ of the mixtures. We derive upper bounds on the Source to Interference Ratio that can be achieved with linear separation in the degenerate case, and propose these bounds as benchmarks to assess how well a (linear or nonlinear) BSS algorithm performs on a given set of mixtures. We will see some examples in Section 4.

2. MEASURES OF DISTORTION

How to measure the distortion between a source $s_n(t)$ and its estimate $\widehat{s}_n(t)$ (provided by some BSS algorithm) is a simple but not completely trivial topic that seems to have been a bit overlooked in the literature. We try to address it in this section, without presuming whether $\widehat{\mathbf{s}} = \mathbf{B}\mathbf{x}$ for some matrix $\mathbf{B}$ or not. We will denote $\langle f, g \rangle := \sum_t f(t)\overline{g}(t)$. the standard inner product of two signals $f(t)$ and $g(t)$, and $\|f\|^2 = \langle f, f \rangle$ is the squared norm of f, *i.e.* its energy. We use the convention that each source is normalised, *i.e.* $\|s_n\| = 1$. Let us note that when BSS is considered in a statistical framework such as Independent Component Analysis (ICA), the *inner product* between sources is called their *correlation* and the notion of *orthogonality* is replaced by that of *de-correlation*. More precisely, when the sources are stationary ergodic, the inner product is, up to a factor $1/T$, the empirical correlation between the sources.

A direct definition of the relative distortion as $D_1 := \|\widehat{s}_n - s_n\|^2 / \|\widehat{s}_n\|^2$ does not take into account one of the well-known aspects of BSS. Indeed, BSS algorithms can in general recover the sources only up to (a permutation and) a gain factor α, *i.e.* the estimate has the form $\widehat{s}_n = \alpha s_n + e$ where e is an error term. The limitations of this definition of relative distortion can be seen if we consider the case of a "perfect estimate" $\widehat{s}_n = \alpha s_n$: the measure of distortion $D_1 = |1 - \alpha^{-1}|^2$ is generally nonzero. To some extent, the gain indetermination is taken into account by the relative distortion $D_2 := \min_{\epsilon=\pm 1} \left\| \widehat{s}_n / \|\widehat{s}_n\| - \epsilon s_n \right\|^2$ [4, 5, 6, 7]. Consider however the worst case where there is no contribution of the true source to its estimate, *i.e.* $\widehat{s}_n = e$ ($\alpha = 0$) : one would likely desire a measure of distortion $D_2 = \infty$; however, if the error e is orthogonal to the true source, then the above measure takes at most the value $D_2 = 2$.

2.1. Total relative distortion

Given an estimate $\widehat{s}_n$ of a (normalised) source s_n, we propose to define the **total relative distortion** as

$$\mathtt{D}_{\mathtt{total}} := \frac{\|\widehat{s}_n\|^2 - |\langle \widehat{s}_n, s_n \rangle|^2}{|\langle \widehat{s}_n, s_n \rangle|^2}. \qquad (1)$$

This is only a slight modification of the measure D_2 (see above), indeed $\mathtt{D}_{\mathtt{total}} = D_2(4 - D_2)/(2 - D_2)^2$. However, when the estimated source is orthogonal to the true source, we have $\langle \widehat{s}_n, s_n \rangle = 0$ and $D_2 = 2$ while $\mathtt{D}_{\mathtt{total}} = +\infty$. We believe this makes $\mathtt{D}_{\mathtt{total}}$ a more relevant distortion measure than D_2.

The definition of $\mathtt{D}_{\mathtt{total}}$ corresponds to the ratio of the energy of the two terms in the decomposition

$$\widehat{s}_n = \langle \widehat{s}_n, s_n \rangle s_n + e_{\mathtt{total}}$$

where the error term $e_{\mathtt{total}}$ is orthogonal to the contribution of the true source. In fact, we have $\|e_{\mathtt{total}}\|^2 = \|\widehat{s}_n\|^2 - |\langle \widehat{s}_n, s_n \rangle|^2$ by Pythagore theorem.

2.2. Interferences, Noise and Artifacts

The error term $e_{\mathtt{total}}$ includes contributions of the other sources (interferences), of the noise $\mathbf{n}$, as well as "artifacts"

of the separation algorithm. In some BSS problems, such as Audio Source Separation, the nature of the distortion is as important as its relative energy level. For example, a distortion due to artifacts of the separation algorithm (which may come, *e.g*, from un-natural zeroes in the Short Time Fourier Transform (STFT) of $\widehat{s}_n$ as in [8]) may be more annoying than interference from the other sources or additive Gaussian noise. We propose to define several measures of distortion instead of just a global one, by decomposing $e_{\mathtt{total}}$ into three terms.

Let us assume for a moment that the source signals $s_n(t)$ are mutually orthogonal (remember that the noise signals $n_k(t)$ are always assumed mutually orthogonal and orthogonal to all sources). Then, the estimated source has an orthogonal decomposition

$$\widehat{s}_n = \langle \widehat{s}_n, s_n \rangle s_n + e_{\mathtt{interf}} + e_{\mathtt{noise}} + e_{\mathtt{artif}} \qquad (2)$$

where $\langle \widehat{s}_n, s_n \rangle s_n$ is the contribution of the true source, $e_{\mathtt{interf}} := \sum_{l \neq n} \langle \widehat{s}_n, s_l \rangle s_l$ is the error term due to interference of the other sources, $e_{\mathtt{noise}} := \sum_{k=1}^{N} \langle \widehat{s}_n, n_k \rangle n_k$ is the error term due to contribution of the additive noise, and

$$e_{\mathtt{artif}} := \widehat{s}_n - \langle \widehat{s}_n, s_n \rangle s_n - e_{\mathtt{interf}} - e_{\mathtt{noise}}$$

is the error term attributed to numerical artifacts of the separation algorithm. In the general case where the various sources may be correlated but are still linearly independent, we consider $P_{\mathbf{s}}$ the orthogonal projector onto their span, and $P_{\mathbf{s},\mathbf{n}}$ the orthogonal projector onto the subspace spanned by the source signals $\{s_n(t)\}_{n=1}^{N}$ together with the noise signals $\{n_k(t)\}_{k=1}^{M}$. The decomposition (2) still holds with

$$e_{\mathtt{interf}} := P_{\mathbf{s}} \widehat{s}_n - \langle \widehat{s}_n, s_n \rangle s_n, \qquad (3)$$

$$e_{\mathtt{noise}} := P_{\mathbf{s},\mathbf{n}} \widehat{s}_n - P_{\mathbf{s}} \widehat{s}_n, \qquad (4)$$

$$e_{\mathtt{artif}} := \widehat{s}_n - P_{\mathbf{s},\mathbf{n}} \widehat{s}_n. \qquad (5)$$

We define the **relative distortion due to interferences**

$$\mathtt{D}_{\mathtt{interf}} := \frac{\|e_{\mathtt{interf}}\|^2}{|\langle \widehat{s}_n, s_n \rangle|^2},$$

the **relative distortion due to additive noise**

$$\mathtt{D}_{\mathtt{noise}} := \frac{\|e_{\mathtt{noise}}\|^2}{\|\langle \widehat{s}_n, s_n \rangle s_n + e_{\mathtt{interf}}\|^2},$$

and the **relative distortion due to algorithmic artifacts**

$$\mathtt{D}_{\mathtt{artif}} := \frac{\|e_{\mathtt{artif}}\|^2}{\|\langle \widehat{s}_n, s_n \rangle s_n + e_{\mathtt{interf}} + e_{\mathtt{noise}}\|^2}.$$

Based on $\mathtt{D}_{\mathtt{total}}$ (resp. $\mathtt{D}_{\mathtt{interf}}$, $\mathtt{D}_{\mathtt{noise}}$, $\mathtt{D}_{\mathtt{artif}}$) we also define the Source to Distortion Ratio (SDR)

$$\mathtt{SDR} := 10 \log_{10} \mathtt{D}_{\mathtt{total}}^{-1},$$

(resp. Source to Interference Ratio (SIR), Source to Noise Ratio (SNR), Source to Artifacts Ratio (SAR)) expressed in decibels.

Note that the definition of $\mathtt{D}_{\mathtt{noise}}$ aims at making it independent of $\mathtt{D}_{\mathtt{interf}}$: consider for example, in Audio Source Separation, an estimate $\widehat{s}_n = \epsilon s_n + s_l + e_{\mathtt{noise}}$ where ϵ, $|\langle s_l, s_n \rangle|$ and $\|e_{\mathtt{noise}}\|$ are small. Such an estimate of s_n is perceived as (almost) noiseless, but with a lot of interference from the source s_l. This is consistent with $\mathtt{D}_{\mathtt{interf}} \approx 1/\epsilon^2 \gg 1$ and $\mathtt{D}_{\mathtt{noise}} \approx \|e_{\mathtt{noise}}\|^2 \ll 1$. Similarly, the definition of $\mathtt{D}_{\mathtt{artif}}$ makes it independent of $\mathtt{D}_{\mathtt{noise}}$ and $\mathtt{D}_{\mathtt{interf}}$.

2.3. Computation of the distortion measures

By definition of the distortion measures, their computation involves computing $\langle \widehat{s}_n, s_n \rangle$ as well as the orthogonal projections $P_{\mathbf{s}} \widehat{s}_n$ and $P_{\mathbf{s},\mathbf{n}} \widehat{s}_n$. In the case of mutually orthogonal sources, we have $P_{\mathbf{s}} \widehat{s}_n = \sum_{l=1}^{N} \langle \widehat{s}_n, s_l \rangle s_l$. For possibly correlated sources, computing $P_{\mathbf{s}} \widehat{s}_n$ is a least squares problem that corresponds to finding the vector $\mathbf{c}^T = [c_1, \ldots, c_N]$ such that $P_{\mathbf{s}} \widehat{s}_n = \sum_{l=1}^{N} c_l s_l = \mathbf{c}^T \mathbf{s}$. It involves the Gram matrix

$$\mathbf{G} := [\langle s_l, s_k \rangle]_{l,k} = \mathbf{s}\mathbf{s}^H$$

of the sources, and we get $\mathbf{c} := \mathrm{conj}(\mathbf{G})^{-1} \mathbf{d}_m$ where $\mathbf{d}_m := (\langle \widehat{s}_n, s_k \rangle)_{k=1}^{N}$ and $\mathrm{conj}(.)$ denotes complex conjugation. Note that if the source signals are considered as realizations of zero mean ergodic stationary stochastic processes, then the Gram matrix $\mathbf{G}$ is also, up to a factor $1/T$, the estimate of their covariance matrix.

In general, one can compute $P_{\mathbf{s},\mathbf{n}} \widehat{s}_n$ in a similar fashion, however we can generally make the assumption that the additive noise signals are mutually orthogonal and orthogonal to each source, so $P_{\mathbf{s},\mathbf{n}} \widehat{s}_n = P_{\mathbf{s}} \widehat{s}_n + \sum_{k=1}^{M} \langle \widehat{s}_n, n_k \rangle n_k / \|n_k\|^2$. Numerical routines to compute the distortion measures are available online [9] and we will see in Section 4 how to use them to compare the performance of BSS algorithms on test examples.

3. LINEAR SEPARATION

In this section we consider the problem of determining a "best" $N \times M$ separation matrix $\mathbf{B}$, in the ideal situation where the true mixing matrix $\mathbf{A}$ is known, for a degenerate mixture ($M < N$). We assume that the sources $s_n(t)$ are mutually orthogonal and that the noise is orthogonal to the sources, and we look for the matrix $\mathbf{B}$ such that $\widehat{\mathbf{s}} := \mathbf{B}\mathbf{x}$ achieves the maximum SIR.

We choose to define the "best" matrix in terms of the SIR for several reasons. First, because it can be checked that with linear separation we always have SAR$= \infty$ (up to numerical inaccuracies and quantisation errors) for all sources. But mainly because the interference from the other sources

is often a more annoying distortion than additive noise (especially from a perceptive point of view in Audio Source Separation) and because de-noising techniques are available [10] that can help remove additive noise after separation. A similar analysis could be held when the matrix is optimised in terms of SDR.

The resulting matrix turns out to be simply the pseudo-inverse $\mathbf{B} = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}$ of the mixing matrix, which is not a real surprise (but was not completely obvious due to the definition of the distortion measure). It shows that the best (in terms of SIR) linear estimate of the sources is nothing but the Maximum a Posteriori (MAP) estimate under the assumption that the sources are Gaussian (see [3]).

The performance of the so obtained "best linear separation" will serve as an upper bound when $\mathbf{B}$ is replaced with an estimated version $\widehat{\mathbf{B}}$. We will see that this upper bound can serve as a difficulty measure for the separation problem.

3.1. Computation of the distortion

Denoting by $\mathbf{a}_n$ the n-th column of $\mathbf{A}$ and $\mathbf{b}_n^H$ the n-th row of $\mathbf{B}$, we have $\widehat{s}_n := \mathbf{b}_n^H\mathbf{A}\mathbf{s} + \mathbf{b}_n^H\mathbf{n}$, thus we have

$$\widehat{s}_n = \mathbf{b}_n^H\mathbf{a}_n s_n + \sum_{l\neq n}\mathbf{b}_n^H\mathbf{a}_l s_l + \mathbf{b}_n^H\mathbf{n}.$$

For the n-th source we only need to know $\mathbf{A}$ and $\mathbf{B}$ to compute the relative distortion due to interferences

$$\mathtt{D}^n_{\mathtt{interf}} = \frac{\sum_{l\neq n}|\mathbf{b}_n^H\mathbf{a}_l|^2}{|\mathbf{b}_n^H\mathbf{a}_n|^2}. \tag{6}$$

One should notice from Eq. (6) that, for orthogonal sources, it is possible to compute the level $\mathtt{D}^n_{\mathtt{interf}}$ of interference *without knowing the original sources*, using only the mixing-demixing pair $(\mathbf{A}, \mathbf{B})$. Eq. (6) can also serve to define a measure of quality for an estimated separation matrix.

3.2. Best separation matrix

The "best" matrix $\mathbf{B}^\star$ corresponds to $\mathbf{b}_n^\star := \arg\min_{\mathbf{b}_n}\mathtt{D}^m_{\mathtt{interf}}$. Standard linear algebra shows that

$$\min_{\mathbf{b}_n}\mathtt{D}^n_{\mathtt{interf}} = \frac{1-\lambda_n}{\lambda_n} \tag{7}$$

where λ_n is the largest eigen-value of $(\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{a}_n\mathbf{a}_n^H$ and that $\mathbf{b}_n^\star$ is a corresponding eigen-vector. That is to say,

$$\lambda_m = \mathtt{tr}\Big\{(\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{a}_n\mathbf{a}_n^H\Big\}, \tag{8}$$

$$(\mathbf{b}_n^\star)^H = \mathbf{a}_n^H(\mathbf{A}\mathbf{A}^H)^{-1}, \tag{9}$$

where $\mathtt{tr}\{.\}$ denotes the trace of a matrix, hence $\mathbf{B}^\star = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H)^{-1}$.

Remark 1 *In the case where the sources are not supposed orthogonal, but their Gram matrix* $\mathbf{G}$ *is known, a similar analysis can be carried out and the result involves* $\mathbf{G}$.

3.3. Fundamental limit of linear separation

The following lemma is an interesting consequence of the above analysis.

Lemma 1 *Consider any degenerate source separation problem with* M *mixtures,* $N > M$ *sources and additive noise where the source and the noise signals are mutually de-correlated. Assume the sources* $\widehat{s}_n$ *are estimated using some separation matrix* $\mathbf{B}$, i.e. $\widehat{\mathbf{s}} = \mathbf{B}\mathbf{x}$. *Then*

$$\max_n \mathtt{D}^n_{\mathtt{interf}} \geq \frac{N}{M} - 1 \tag{10}$$

Proof. We simply observe that, if $\mathbf{I}_M$ denotes the $M \times M$ identity matrix,

$$\begin{aligned}\sum_{n=1}^N \lambda_n &= \mathtt{tr}\Big\{(\mathbf{A}\mathbf{A}^H)^{-1}\sum_{n=1}^N \mathbf{a}_n\mathbf{a}_n^H\Big\} \\ &= \mathtt{tr}\Big\{(\mathbf{A}\mathbf{A}^H)^{-1}\mathbf{A}\mathbf{A}^H\Big\} = \mathtt{tr}\{\mathbf{I}_M\} = M.\end{aligned}$$

We conclude that $\max_n \lambda_n \geq M/N$. Combining with (7) we get the result. □

What is expressed in Lemma 1 is that for any degenerate separation problem with M mixtures and $N > M$ (de-correlated) sources, if we try to perform a separation by some linear algorithm, there is *at least* one source that will be poorly estimated: the corresponding SIR will be at most $10\log_{10} M/(N-M) < \infty$.

3.4. Difficulty measures for degenerate separation

The above analysis gives bounds on the performance of linear source separation. These lower bounds on the relative distortion due to interference are upper bounds in terms of SIR. They can serve as difficulty measures [11] and/or benchmarks for the evaluation [4] of actual BSS algorithms on test databases of sources and mixing matrices.

The bounds given by Equations (7) and (8) can be computed as soon as the mixing matrix $\mathbf{A}$ is known, without assuming the sources themselves are available. These bounds can then serve as a difficulty measure [11] for the separation problem, in order to "calibrate" the range of difficulties of separation problems that a given BSS algorithm can address.

Lemma 1 gives another type of bound, which can be computed based on the sole knowledge of the relative number of sources and mixtures. As such, the figure $M/(N-M)$ can serve as a global measure to compare the difficulty of degenerate separation problems of different sizes (M, N).

4. EXAMPLES

A very common degenerate separation problem is the separation of $N \geq 3$ different instruments from a stereophonic

recording ($M = 2$). For $N = 3$ instruments, Lemma 1 shows that at least one of the instruments will be recovered with a relative distortion $\mathtt{D}_{\mathtt{interf}}$ at least $1/2$, hence the worst SIR is no better than $+3$ dB. In the case of $N = 4$ instruments, the worst SIR becomes at best 0 dB.

In the ($M = 2, N = 3$) case, consider for example the two following mixing matrices

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

One can easily compute that for the first matrix $\lambda_1 = \lambda_2 = 1/2$, $\lambda_3 = 1$, hence $\mathrm{SIR}_1 = \mathrm{SIR}_2 = 0$ dB and $\mathrm{SIR}_3 = +\infty$ dB. This coincides with the obvious observation that the third source, which is the only one present on the second channel, can be perfectly recovered, while the first two sources cannot be linearly separated. For the second matrix, we compute $\lambda_1 = \lambda_2 = \lambda_3 = 2/3$ and $\mathrm{SIR}_1 = \mathrm{SIR}_2 = \mathrm{SIR}_3 = +3$ dB. In a sense, this corresponds to the "worst case" of Lemma 1, because *all* sources are poorly estimated, with *all* SIR equal to the bound given by the Lemma.

Let us now give a more concrete example. We consider a noiseless mixture of three normalised sources (s_1=cello, s_2=drums, s_3=piano) on two channels, where the mixing matrix and the Gram matrix of the sources are respectively

$$\mathbf{A} \approx \begin{bmatrix} 1.8 & 2.8 & 2.1 \\ 0.8 & 2.8 & 5 \end{bmatrix},$$

and

$$\mathbf{G} \approx \begin{bmatrix} 1 & -0.0055 & -0.0016 \\ -0.0055 & 1 & 0.0063 \\ -0.0016 & 0.0063 & 1 \end{bmatrix},$$

The Gram matrix shows that the sources are essentially de-correlated, so we used the results of Section 3 to get the SIR figures of the best linear separation indicated in the first column of Table 1. These figures illustrate well the fact that some very common problems of source separation can only get an acceptable solution by relying on BSS algorithms that are not linear. By comparison, we show in the second column of Table 1 the performance figures (on the same example) obtained with a nonlinear separation algorithm based on Matching Pursuit and clustering [12]. The nonlinear separation algorithms improves the SIR figures by at least about 20 dB. However, it introduces artifacts, which is indicated by the SAR figures. In fact, the SDR figures reflect that the distortion due to artifacts completely dominates the interference of the other sources. Informal listening tests with this example and some others confirmed the good correlation between the perceived nature of the distortion and the SIR/SAR figures. It may depend on the target application whether it is preferable to have few artifacts or few interference from the other sources.

Method		linear	nonlinear
SIR (dB)	s_1	0	19.3
	s_2	1.1	26.6
	s_3	11.5	31.2
SAR (dB)	s_1	78.2	3.7
	s_2	77.5	8.1
	s_3	82.4	14.9
SDR (dB)	s_1	0	3.5
	s_2	1.1	8.1
	s_3	11.5	14.8

Table 1. Comparison of the performance of the best linear separation and of a nonlinear separation algorithm based on Matching Pursuit and clustering, on an example with three sources and two mixtures. The SDR figures reflect that the total distortion is dominated respectively by the interference of the other sources (linear separation) and by the artifacts of the separation algorithm (Matching Pursuit).

5. CONCLUSION

In this paper we addressed the simple, yet crucial question of how to measure the performance of source separation algorithms in terms of distortion, by taking properly into account the gain indeterminacies of BSS. We pointed out that the main issue is to decompose each estimated source into a contribution due to the true source and a distortion term. We proposed to further decompose the distortion term into interference of the other sources, noise and algorithmic artifacts, and we defined the Source to Interference Ratio (SIR), the Source to Noise Ratio (SNR), the Source to Artifact Ratio (SAR) and the Source to Distortion Ratio (SDR).

Our second main contribution is specific to degenerate BSS. We showed that for de-correlated sources, the SIR of linear separation algorithms can be computed based on the sole knowledge of the (mixing matrix, de-mixing matrix) pair. We characterised the limits of such algorithms and proposed to use the derived bounds as difficulty measures for degenerate BSS problems.

We demonstrated with an example how the proposed distortion measures can be used to evaluate and compare the performance of source separation algorithms. Using a database of sources and mixtures, these figures of merit could be used to assess the relative performance of various algorithms on different tasks [13]. Besides the test mixtures, the database should obviously contain the original sources. In the case of noisy mixtures, the database should also contain the realization of additive noise that was added (in the case of synthetically added noise of course).

6. FUTURE WORK

The distortion measures that we have introduced in this paper are defined on the whole temporal signal for each source $s_n(t), 1 \leq t \leq T$. Naturally, it is also possible to consider these measures on "pieces" of the sources, such as windowed temporal frames and/or frequency sub-bands. In such cases, one should check that the assumption of linear independence between the different sources is still valid, in particular that $\max(N, M)$ is (much) smaller than the size of the frames or the bandwidth of the sub-bands. In order to get usable performance measures on large signals (with many frames and/or many sub-bands, it will be necessary to summarise the distortion figures on each piece by a few appropriate global statistics.

Here we have showed how to decompose an estimated source into a contribution of the true source and various error terms that correspond respectively to the interference of the other sources, the contribution of additive noise and algorithmic artifacts. We have measured the distortion due to each of these error terms based on their relative energy. However for Audio Source Separation, perceptual effects should be taken into account when measuring the level of distortion [14]. Another issue that we are currently considering is the extension of the approach to define measures of distortion for convolutive BSS problems.

7. ACKNOWLEDGEMENTS

This work has been performed within a Junior Researchers Project "Resources for Audio Signal Separation" funded by GdR ISIS (CNRS). The goal of the project is to identify typical "tasks" that are specific to audio signal separation, suggest relevant numerical criteria, and gather test signals of calibrated difficulty level, in order to evaluate the performance of existing and future algorithms. Some insights, numerical routines and databases of audio signals can be found on the web site [9].

8. REFERENCES

[1] Jean-François Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE. Special issue on blind identification and estimation*, vol. 9, no. 10, pp. 2009–2025, Oct. 1998.

[2] Jean-François Cardoso, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO*, Edinburgh, Sept. 1994, pp. 776–779.

[3] Olivier Bermond and Jean-François Cardoso, "Méthodes de séparation de sources dans le cas sous-déterminé," in *Proc. GRETSI, Vannes, France*, 1999, pp. 749–752.

[4] D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods," in *Proceedings Int. Workshop Independent Component Analysis and Blind Signal Separation*, Aussois, France, Jan. 1999, pp. 261–266.

[5] M. Van Hulle, "Clustering approach to square and non-square blind source separation," in *IEEE Workshop on Neural Networks for Signal Processing (NNSP99)*, Aug. 1999, pp. 315–323.

[6] M. Zibulevsky and B.A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural Computations*, vol. 13, no. 4, pp. 863–882, 2001.

[7] P. Kisilev, M. Zibulevsky, Y. Y. Zeevi, and B. A. Pearlmutter, "Multiresolution framework for blind source separation," Tech. Rep. CCIT Report # 317, Technion University, June 2001.

[8] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing n sources from 2 mixtures," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP'00)*, Istanbul, Turkey, June 2000, vol. 5, pp. 2985–2988.

[9] Action Jeunes Chercheurs du GDR ISIS (CNRS), "Ressources pour la séparation de signaux audiophoniques," `http://www.ircam.fr/anasyn/ISIS/`.

[10] D.L. Donoho and I.M. Johnstone, "Ideal denoising in an orthonormal basis chosen from a library of bases," *Comptes-Rendus Acad. Sci. Paris Série I*, vol. 319, pp. 1317–1322, 1994.

[11] R.H. Lambert, "Difficulty measures and figures of merit for source separation," in *Proc. Int. Workshop on ICA and BSS (ICA'99)*, Aussois, France, 1999, pp. 133–138.

[12] R. Gribonval, "Sparse decomposition of stereo signals with matching pursuit and application to blind separation of more than two sources from a stereo mixture," in *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP'02)*, Orlando, Florida, May 2002.

[13] E. Vincent, X. Rodet, A. Roëbel, C. Févotte, R. Gribonval, L. Benaroya, and F. Bimbot, "Typical tasks in audio source separation," in *Proc. Int. Workshop on ICA and BSS (ICA'03)*, 2003, submitted.

[14] C. Colomes, C. Schmidmer, T. Thiede, and W.C. Treurniet, "Perceptual quality assessment for digital audio (PEAQ) : the proposed ITU standard for objective measurement of perceived audio quality," in *Proc. AES Conf.*, 1999.