A New Auditory-Based Index to Evaluate the Blind Separation Performance of Acoustic Mixtures

Juan Manuel Sanchis[1], José Joaquín Rieta[1], Francisco Castells[1], and José Millet[2]

[1]Universidad Politécnica de Valencia,
46730 Gandia, Spain
{jmsanch,jjrieta,fcastells}@eln.upv.es
[2]Universidad Politécnica de Valencia,
46020 Valencia, Spain
jmillet@eln.upv.es

Abstract. A new method to evaluate the performance of convolutive blind signal separation (BSS) algorithms in acoustic mixtures is presented. The method is able to compute the spectral level enhancement, in a frequency-band based fashion, between the estimated and the residual sources, thus combining two previously defined parameters in time and frequency domain: the signal to interference ratio and the spectral preservation index. The new index is able to compute the quality of separation performing the spectral computations over a set of logarithmically spaced frequency bands in a similar way as the human auditory system. Obtained results clearly verify that this methodology is a more realistic approach to evaluate the convolutive BSS results of acoustic mixtures.

1 Introduction

In the blind source separation (BSS) problem the objective is to separate multiple sources, mixed through an unknown mixing system (or channel), using only the system output data (observed signals) and in particular with the absence (or least amount) of information about the sources or the mixing system. Today, it is well known the successful application of BSS techniques in a wide variety of fields including speech processing, data communication, biomedical signal processing, etc.

An extensive part of the available literature has been directed toward the simpler case of instantaneous mixtures; i. e., when the observed signals are generated through a linear combination of the sources and no time-delays are involved in the mixing model [1-3]. A more challenging case arises when dealing with convolutive mixing systems; i.e., where the sources are mixed through a linear filtering operation and the observed signals are linear combinations of the sources and their corresponding scaled and delayed versions [4-6]. A difficult practical example, and possibly one of the most extended convolutive BSS problem, regards the separation of audio signals mixed in a reverberant environment.

To evaluate the results of BSS algorithms in acoustic mixtures two different methods are available in the literature. The first one is based on the impulse response associated to the mixing channels and the separation filters. The quality of separation is calculated by convolving the mixing and separation systems and measuring the ap-

C.G. Puntonet and A. Prieto (Eds.): ICA 2004, LNCS 3195, pp. 1118–1125, 2004.

proximation degree of the result to the unitary system [7], [8]. The second one establishes a comparison between the nuisance sources and the desired source for each separated signal [9], [10]. To perform the comparison, the aforementioned methods need the availability of the original sources and the mixing system, which is a problem for real world situations. Additionally, other method mainly used in real situations, evaluates the signals' subjective quality of separation, because the original sources and the mixing system are unavailable for comparison purposes. Another relevant aspect is to evaluate the spectral preservation of the recovered signals but few methods in the bibliography consider this problem [8].

In the present work, a new index to evaluate the separation performance is presented based on two parameters, previously defined in the bibliography, for measuring the quality of separation from two different points of view: signal to interference ratio (*SIR*) in the computed time interval [10] and the spectral preservation index (*SPI*) [8]. These indexes will be used together for the definition of the new spectral enhancement index (*SEI*).

2 Signal to Interference Ratio

The estimation of *SIR* can be performed both in synthetic and real world situations. In the first case, the original source signals s_j (j = 1,…, N), the mixing matrix $\mathbf{H}^{MxN}$, the observations x_i (i = 1,…,M), the estimated separation matrix $\mathbf{W}^{NxM}$, and the estimated source signals y_j (j = 1,…,N) are available. For real world situations, the only available data are the microphone signals, the estimated signals and the separation matrix.

In a synthetic case, the *SIR* for a source signal s_j can be defined as the difference between the *SIR* of s_j in the estimation y_j, and the *SIR* of s_j in the observation x_i. Hence, the difference between the *SIR* of the estimated (SIR_j^e) and observed (SIR_j^o) signals is the parameter where s_j presents the largest power contribution

$$SIR_j = SIR_j^e - SIR_j^o, \qquad j = 1,...,N \tag{1}$$

where

$$SIR_j^o = 10\log\frac{E\{(h_{jj} * s_j)^2\}}{E\{(\sum_{\substack{k=1\\k\neq j}}^{N} h_{jk} * s_k)^2\}}, \qquad j = 1,...,N \tag{2}$$

being $E\{\cdot\}$ the mathematical expectation operator and h_{ji} the mixing matrix entries. Taking into account that the global transfer function of the mixing-separation system can be defined as $\mathbf{G} = \mathbf{W} * \mathbf{H}$, one can formulate the *SIR* of the estimated signal as

$$SIR_j^e = 10\log\frac{E\{(g_{jj} * s_j)^2\}}{E\{(\sum_{\substack{k=1\\k\neq j}}^{N} g_{jk} * s_k)^2\}}, \qquad j = 1,...,N \tag{3}$$

where g_{jk} are the entries of global matrix.

When an ideal separation is not achievable ($\mathbf{G} \neq \mathbf{I}$), Eq. (**3**) performs the comparison between a modified source signal and the residual sources present in the estimation.

Regarding real world situations, where we have no access to the original signals s_j and the mixing matrix **H**, the measurement of the *SIR* index given in Eq. (1) would not be applicable, unless just one of the sources is active during a certain time interval [11]. In this case, the observed signal by the microphone x_i can be defined as

$$x_i = \sum_{j=1}^{N} h_{ji} * s_j = \sum_{j=1}^{N} x_{ji}, \qquad j = 1,...,N \tag{4}$$

Therefore, we can define the observed *SIR*, with respect to the rest of the signals present in that observation, as

$$SIR_j^o = 10 \log \frac{E\{(x_{ji})^2\}}{E\{(\sum_{\substack{k=1 \\ k \neq j}}^{N} x_{ki})^2\}}, \qquad j = 1,...,N \tag{5}$$

and the estimated *SIR*, given the separation matrix **W**, as the quality of separation of source s_j, with respect to the rest of the signals present in that estimation, as

$$SIR_j^e = 10 \log \frac{E\{(\sum_{i=1}^{M} w_{ij} * x_{ji})^2\}}{E\{(\sum_{i=1}^{M} w_{ij} * \sum_{\substack{k=1 \\ k \neq j}}^{N} x_{ki})^2\}}, \qquad j = 1,...,N \tag{6}$$

3 Spectral Preservation

A complementary way to estimate the performance of a BSS algorithm lies in the analysis of the preservation capability, with respect to the spectral components of the signal, by comparing the spectral content of the estimated sources to the original ones. This kind of measure has special importance when the main goal is focused on recovering the signal with the highest fidelity, in contrast with the situations where the relevance of fidelity is shaded by intelligibility. Hence, the spectral preservation index (*SPI*) can be as [8]

$$SPIj = E\{\left| P_{s_j}(f) - P_{y_j}(f) \right|^2\}, \qquad j = 1,...,N \tag{7}$$

where $P_{sj}(f)$ is the power spectral density (PSD) of the j^{th} original source and $P_{yj}(f)$ the PSD of the j^{th} estimated signal.

Note that the aforementioned defined parameters, *SIR* and *SPI*, present some deficiencies when compared to the human auditory system. Regarding the computation of *SIR*, the mean value of the total signal power is obtained over a concrete time interval, with no consideration of the spectral content of the source signals. Hence, this index gives no information about the real quality of separation because the spectral overlapping degree of the sources is not considered at all. In the case of *SPI*, the spectral content of the sources is analyzed but the computation is performed linearly over the frequency axis, hence, the perceptual behavior of the human auditory system is not considered.

4 Spectral Enhancement Index

In order to define an auditory-based index, capable of measuring the quality of separation in a similar way as the human auditory system, the spectral computations should be performed over a logarithmic frequency axis, more specifically, in critical bands. Considering this fact, is it possible to define the spectral enhancement index *SEI* that can be obtained following this steps: firstly, the contribution of each independent source s_j is obtained separately on the observation point. This will give us the contribution of the j^{th} source into the i^{th} observation in the same way as has been defined in Eq. (4). The second step consists of obtaining the difference between the PSD of the source with largest power contribution into the i^{th} observation (x_{ji}) with respect to the PSD of the rest of the sources in that observation, which can be defined as

$$SEI_{ji}^{o}(f) = P_{x_{ji}}(f) - \sum_{\substack{k=1 \\ k\neq 1}}^{N} P_{x_{ki}}(f), \qquad \begin{matrix} j = 1,...,N \\ i = 1,...,M \end{matrix} \tag{8}$$

The next step deals with the application of the BSS algorithm and the evaluation of the quality of separation over each estimated source. Then, the difference between the PSD of the j^{th} source with largest power contribution into the I^{th} estimation (y_{jl}) with respect to the PSD of the rest of the sources in that estimation, can then be computed as

$$SEI_{jl}^{e}(f) = P_{y_{jl}}(f) - \sum_{\substack{k=1 \\ k\neq j}}^{N} P_{y_{kl}}(f), \qquad j,l = 1,...,N \tag{9}$$

Finally, the *SEI* for the j^{th} source can be defined as

$$SEIj = E\{SEI_{jl}^{e}(f)\big|_{1/3} - SEI_{ji}^{o}(f)\big|_{1/3}\}, \qquad j = 1,...,N \tag{10}$$

where $SEI_{jl}^{e}(f)\big|_{1/3}$ and $SEI_{jl}^{o}(f)\big|_{1/3}$ are the spectral enhancement indexes for the estimated and observed sources evaluated in 1/3 octave bands, respectively.

5 Results

The proposed methodology has been applied over real recordings corresponding to an acoustic guitar (s_1) and a male speaker (s_2) in the recording studio of Gandia Higher School of Technology. The resulting impulse response between each source and two microphones are plotted in Fig. 1.

The signals were 4 seconds in length and sampled at 44.1 kHz. Fig. 2 plots the PSD of source s_1 and s_2 in the observation x_1. After the application of the estimated separation matrix **W** (Fig. 3) limited to 16384 taps, it is possible to obtain a set of two estimated sources. Fig. 4 plots the PSD of the estimated source s_1 and the interfering source s_2 in the estimation y_1.

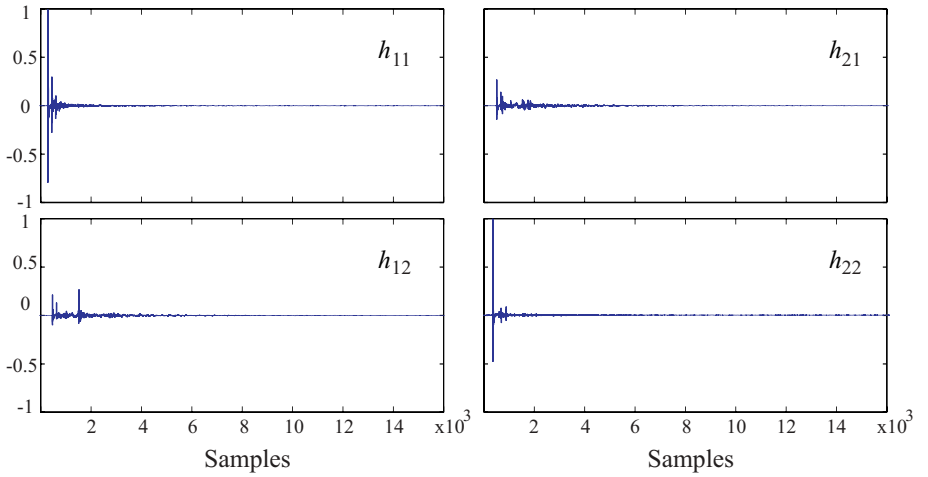

Fig. 1. Mixing matrix **H,** composed by the impulse response among the position of two sources and two microphones located in a recording study.

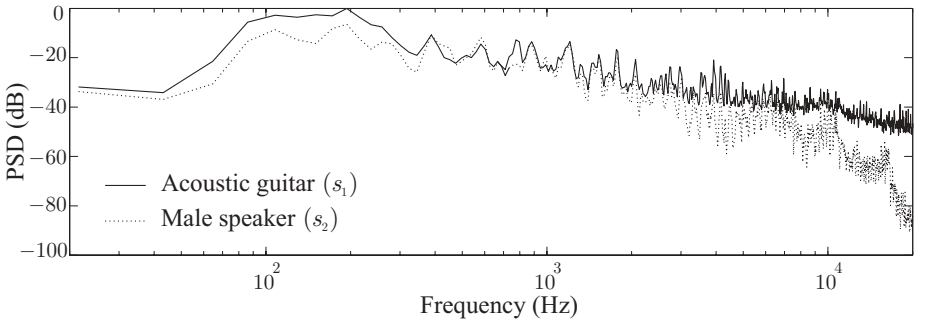

Fig. 2. PSD of the sources s_1 and s_2 in the observation x_1.

Finally, the difference between $SEI_{11}^{e}(f)$ (Eq. (9)) and $SEI_{11}^{o}(f)$ (Eq. (8)), in 1/3 octave bands, can be observed in Fig. 5. As shown in this Fig. the quality of the separation decays substantially in the mid-range frequencies, indicating a poor separation performance in this frequency band that would be masked with the use of the previously mentioned indexes *SIR* and *SPI*, thus corroborating that *SEI* is a much more realistic approach to compute the BSS separation.

When a global separation index is required, it is possible to obtain the SEI_1 index as the mean value of the partial band indexes across the whole bandwidth shown in Fig. 5.

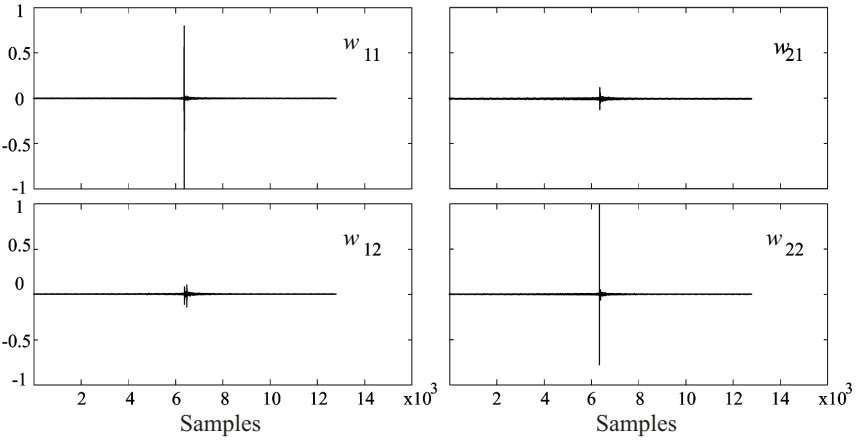

Fig. 3. Separation matrix **W**, obtained from the 16384 taps Hamming window of the mixing matrix inverse ($\mathbf{H^{-1}}$).

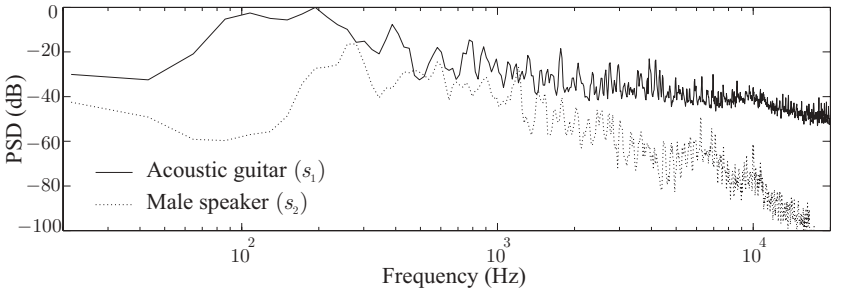

Fig. 4. PSD of the sources s_1 and s_2 in the estimation y_1.

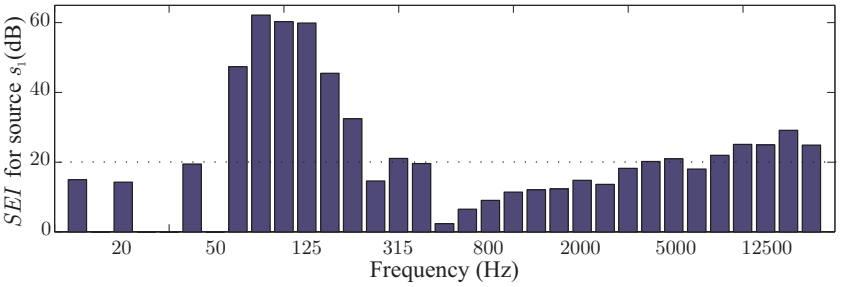

Fig. 5. SEI_1 evaluated in 1/3 octave bands in estimated signal y_1.

Table 1. *SIR* and *SEI* values obtained in the you are considered signal y_1 by lineal average and by average of the value in 1/3 octave bands, respectively.

SIR_1	SEI_1
30.1 dB	21.1 dB

SIR values at different frequencies are linearly averaged in order to obtain the global *SIR*. Using this performance measurement, the *SIR* values at high frequencies are considerably overweighted in comparison to the *SIR* values at low and medium frequencies, and have a significant impact on the global value. Therefore, the performance index provided by *SIR* is far from the performance index that would be obtained if the logarithmic characteristic of the human auditory system had been considered.

This divergence between the values given by *SIR* and *SEI* is even more significant in the cases where the involved audio signals present limited bandwidth. In those situations, both *SIR* and *SPI* will quantify with much lower accuracy the separation degree of the estimated source in contrast to the real perception of the human auditory system.

6 Conclusion

A new index for measuring the separation performance of convolutive BSS algorithms for acoustic mixtures has been defined. The method considers the quality of separation between the estimated source and the secondary sources as a function of frequency . Moreover, the measurement is performed over a set of frequency bands logarithmically spaced similar to the human auditory system.

The most widely used methods to evaluate the quality of separation have been compared to the new proposed methodology in order to show that the results obtained with the latter are much more approximated to reality. This observation can be corroborated through the introduction of the logarithmically spaced frequency-band based computations. Therefore, this methodology establishes a new measurement procedure closer to human perception and, though more spectral and statistic parameters could be defined, this new index can be used as a more realistic basis to evaluate the separation performance in acoustic mixtures of present and future BSS algorithms.

References

1. P. Comon, ”Independent component analysis, A new concept?,” *Signal Processing*, vol. 36, pp. 287-314, 1994.
2. A. Bell and T. J. Sejnowski, “An information maximization approach to blind separation and blind deconvolution,” *Neural Computation*, no. 7, pp. 1129-1159, 1995.
3. J. Cardoso and B. Laheld, “Equivariant adaptative source separation,” *IEEE Transaction on Signal Processing*, vol. 44, pp. 3017-3030, Dec. 1996.

4. H. Sahlin and H. Broman, "MIMO signal separation for FIR channels: a criterion and performance analysis," *IEEE Transaction on Signal Processing*, vol. 48, pp. 642-649, March. 2000.
5. E. Weinstein, M. Feder, and A. V. Oppenheim, Multi-channel signal separation by decorrelation", *IEEE Transaction on Signal Processing*, vol. 1, pp. 404-413, Oct. 1993.
6. J. K. Tugnait, "Adaptative blind separation of convoluitve mixtures of independent signals," *Signal Processing*, vol. 73, pp. 139-152, 1999.
7. S. Cruces. *An unified view of Blind Source Separation Algoritmhs*. Ph. D. Thesis, University of Vigo, 1999.
8. K. Kokkinakis, V. Zarzoso and A. Nandi, "Blind separation of acoustic mixtures based on linear prediction analysis," *International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*, vol. 4, pp. 59-64, Nara, Japan, April 2003.
9. S. Ikeda and N. Murata, "A method of blind separation based on temporal structure of signals," *Proceedings of the International Conference on Neural Information Processing, (ICONIP'98)*, pp.737-742, Kitakyushu, Japan, October 1998.
10. S. Araki, R. Mukai, S. Makino, T. Nishikawa, H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixture of speech," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no 2, pp. 109-116, March 2003.
11. D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods," *International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*, vol. 1, pp. 261-266, Aussois, France, January 1999.