Two ICA Algorithms Applied to BSS in Non-destructive Vibratory Tests

Juan-José González de-la-Rosa[1], Carlos G. Puntonet[2], R. Piotrkowski, I. Lloret[1], and Juan-Manuel Górriz

[1] University of Cádiz, Research Group TIC168 - Computational Instrumentation and Industrial Electronics, EPSA, Av. Ramón Puyol S/N. 11202, Algeciras-Cádiz, Spain
juanjose.delarosa@uca.es

[2] University of Granada, Department of Architecture and Computers Technology, ESII, C/Periodista Daniel Saucedo. 18071, Granada, Spain
carlos@atc.ugr.es

Abstract. Two independent component analysis (ICA) algorithms have been applied for blind source separation (BSS) in a synthetic, multi-sensor scenario, within a non-destructive pipeline test. The first one, CumICA, is based in the computation of the cross-cumulants of the mixed observed signals, and needs the aid of a digital high-pass filter to achieve the same SNR (up to -40 dB) as the second algorithm, Fast-ICA. Vibratory signals were acquired by a wide frequency range transducer (100-800 kHz) and digitalized by a 2.5 MHz, 8-bit ADC. Different types of commonly observed source signals are linearly mixed, involving acoustic emission (AE) sequences, impulses and other parasitic signals modelling human activity. Both ICA algorithms achieve to separate the impulse-like and the AE events, which often are associated to cracks or sudden non-stationary vibrations.

1 Introduction

Vibratory and acoustic emission (AE) signal processing usually deals with separation of multiple events which sequentially or simultaneously occur in different measurement points during a non-destructive test. In most situations, the tests involve the study of the behavior of secondary events, or reflections, resulting from an excitation (the main event). These echoes carry information related with the medium through which they propagate, as well as surfaces where they reflect [1].

But, in almost every measurement scenario, an acquired sequence contains information regarding not only the AE under study, but also additive noise processes (mainly from the measurement equipment) and other parasitic signals, e.g. originated by human activity or machinery vibrations. As a consequence, in non-favorable SNR cases, BSS should be accomplished before characterization [2], in order to obtain the most reliable spectral *fingerprint* of the AE event.

The purpose of this paper is twofold. First we show how two ICA algorithms separate the true AE event from the parasitics, taking a multi-sensor array of

S. Kollias et al. (Eds.): ICANN 2006, Part II, LNCS 4132, pp. 221–229, 2006.

inputs (SNR=-40 dB). Secondly, we compare performances of Cum-ICA and Fast-ICA, resulting that Cum-ICA needs the aid of a post high-pass filter to achieve the same SNR as Fast-ICA. This comparison could be interesting for a future implementation of the code in an automatic test system.

The paper is structured as follows: in Section 2 we make a brief progress report on the characterization of vibratory emissions. Section 3 summarizes the ICA models and outlines their properties. Results are displayed in section 4. Finally, conclusions and achievements are drawn in section 5.

2 Acoustic Emission Signal Processing

Elastic energy travels through the material as a stress wave and is typically detected using a piezoelectric transducer, which converts the surface displacement (vibrations) to an electrical signal. AE signal processing is used for the detection and characterization of failures in non-destructive testing and identification of low-level biological signals [2]. Most AE signals are non-stationary and they consist of overlapping bursts with unknown amplitude and arrival time. These characteristics can be described by modelling the signal by means of neural networks, and using wavelet transforms [1],[3]. These second-order techniques have been applied in an automatic analysis context of the estimation of the time and amplitude of the bursts. Multiresolution has proven good performance in de-noising (up to SNR=-30 dB, with modelled signals) and estimation of time instances, due to the selectivity of the wavelets filters banks [4].

Higher order statistics (HOS) have enhanced characterization in analyzing biological signals due to the capability for rejecting noise [5]. This is the reason whereby HOS could be used as part of an ICA algorithm.

3 The ICA Model and Algorithms

3.1 Outline of ICA

BSS by ICA is receiving attention because of its applications in many fields such as speech recognition, medicine and telecommunications [6]. Statistical methods in BSS are based in the probability distributions and the cumulants of the mixtures. The recovered signals (the source estimators) have to satisfy a condition which is modelled by a contrast function. The underlying assumptions are the mutual independence among sources and the non-singularity of the mixing matrix [2],[7].

Let $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_m(t)]^T$ be the transposed vector of sources (statistically independent). The mixture of the sources is modelled via

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{s}(t) \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_m(t)]^T$ is the available vector of observations and $\mathbf{A} = [a_{ij}] \in \Re^{m \times n}$ is the unknown mixing matrix, modelling the environment in which signals are mixed, transmitted and measured [8]. We assume that $\mathbf{A}$ is a non-singular n×n square matrix. The goal of ICA is to find a non-singular n×m separating matrix $\mathbf{B}$ such that extracts sources via

$$\hat{\mathbf{s}}(t) = \mathbf{y}(t) = \mathbf{B} \cdot \mathbf{x}(t) = \mathbf{B} \cdot \mathbf{A} \cdot \mathbf{s}(t) \tag{2}$$

where $\mathbf{y}(t) = [y_1(t), y_2(t), \ldots, y_m(t)]^T$ is an estimator of the sources. The separating matrix has a scaling freedom on each row because the relative amplitudes of sources in $\mathbf{s}(t)$ and columns of $\mathbf{A}$ are unknown [7]. The transfer matrix $\mathbf{G} \equiv \mathbf{BA}$ relates the vector of independent (original) signals to its estimators.

3.2 CumICA

High order statistics, known as cumulants, are used to infer new properties about the data of non-Gaussian processes. Before, such processes had to be treated as if they were Gaussian, but second order statistics are phase-blind. The relationship among the cumulant of r stochastic signals and their moments of order $p, p \leq r$, can be calculated by using the *Leonov-Shiryayev* formula [9]:

$$\begin{aligned} Cum(x_1, ..., x_r) = \sum (-1)^k \cdot (k-1)! \cdot E\{\prod_{i \in v_1} x_i\} \\ \cdot E\{\prod_{j \in v_2} x_j\} \cdots E\{\prod_{k \in v_p} x_k\} \end{aligned} \tag{3}$$

where the addition operator is extended over all the set of v_i $(1 \leq i \leq p \leq r)$ and v_i compose a partition of *1,. . . ,r*.

A set of random variables are statistically independent if their cross-cumulants are zero. This is used to define a contrast function, by minimizing the distance between the cumulants of the sources $\mathbf{s}$(t) and the outputs $\mathbf{y}$(t). As sources are unknown, it is necessary to involve the observed signals. Separation is developed using the following contrast function based on the entropy of the outputs [2]:

$$H(\mathbf{z}) = H(\mathbf{s}) + log[det(\mathbf{G})] - \sum \frac{\mathbf{C}_{1+\beta, y_i}}{1+\beta} \tag{4}$$

where $\mathbf{C}_{1+\beta, y_i}$ is the $1+\beta$th-order cumulant of the ith output, $\mathbf{z}$ is a non-linear function of the outputs y_i, $\mathbf{s}$ is the source vector, $\mathbf{G}$ is the global transfer matrix of the ICA model and $\beta > 1$ is an integer verifying that $\beta + 1$-order cumulants are non-zero.

Using equation 4, the separating matrix can be obtained by means of the following recurrent equation [8]

$$\mathbf{B}^{(h+1)} = [\mathbf{I} + \mu^{(h)}(\mathbf{C}_{y,y}^{1,\beta}\mathbf{S}_y^{\beta} - I)]\mathbf{B}^{(h)} \tag{5}$$

where $\mathbf{S}_y^{\beta}$ is the matrix of the signs of the output cumulants. Equation 5 is interpreted as a quasi-Newton algorithm of the cumulant matrix $\mathbf{C}_{y,y}^{1,\beta}$. The learning rate parameters $\mu^{(h)}$ and η are related by:

$$\mu^{(h)} = \min(\frac{2\eta}{1+\eta\beta}, \frac{\eta}{1+\eta\|\mathbf{C}_{y,y}^{1,\beta}\|_p}) \tag{6}$$

with $\eta < 1$ to avoid $\mathbf{B}^{(h+1)}$ being singular; $\|.\|_p$ denotes de p-norm of a matrix. The adaptative equation 5 converges, if the matrix $\mathbf{C}^{1,\beta}_{y,y}\mathbf{S}^{\beta}_{y}$ tends to the identity.

3.3 FastICA

One of the independent components is estimated by $y = \mathbf{b}^T\mathbf{x}$. The goal of FastICA is to take the vector $\mathbf{b}$ that maximizes the non-Gaussianity (independence)of y, by finding the maxima of its negentropy [7]. The algorithm scheme is an approximative Newton iteration, resulting from the application of the *Kuhn-Tucker* conditions. This leads to the equation 7

$$E\{\mathbf{x}g(\mathbf{b}^T\mathbf{x}) - \beta\mathbf{b} = 0\} \tag{7}$$

where g is a non-quadratic function and β is an iteration parameter.

Provided with the mathematical foundations the experimental results are outlined.

4 Experimental Results

The inputs of the ICA algorithms comprise synthetics (laboratory mixtures), which have been obtained by mixing real AE events (the ones we are interested in getting the spectral *track*), impulse-like events, noise processes and damping sinusoids. The sensor used to capture the AE events was attached to the outer surface of the pipeline, which is under mechanical excitation.

A number of 20 AE events were captured. One of these vibratory signals is depicted in Fig. 1, where we can observe the main AE event and the secondary reflections or echoes.

Each digitalized sequence comprises 2502 points (sampling frequency of 2.5 MHz and 8 bits of resolution), and assembles the main AE event and the subsequent reflections (echoes).

Four types of sources have been considered and linearly mixed in the synthetics. These subsequent mixtures constitute the inputs of the algorithm: A real AE event, an uniform white noise (SNR=-40 dB), a damped sine wave and an impulse-like event. The damping sine wave models a mechanical vibration which may occur, e.g. as a consequence of a maintenance action. It has a damping factor of 2000 and a frequency of 8000 Hz. Finally, the impulse is included as a very common signal registered in vibration monitoring. Fig. 2 shows one possible input quartet.

One of the 20 results (output quartet) of CumICA is depicted in Fig. 3. The damping sinusoid is considered as a frequency component of the impulse-like event because IC3 and IC4 are almost the same. The final independent components are obtained filtering the independent components by a 5th-order *Butterworth* high-pass digital filter (20 kHz).

The resulting separated sources resulting from one of the Fast-ICA processing are depicted in Fig. 4.

Finally, to test the independence of the independent components, some relevant joint distributions have been included in Fig. 5 and in Fig. 6. The left

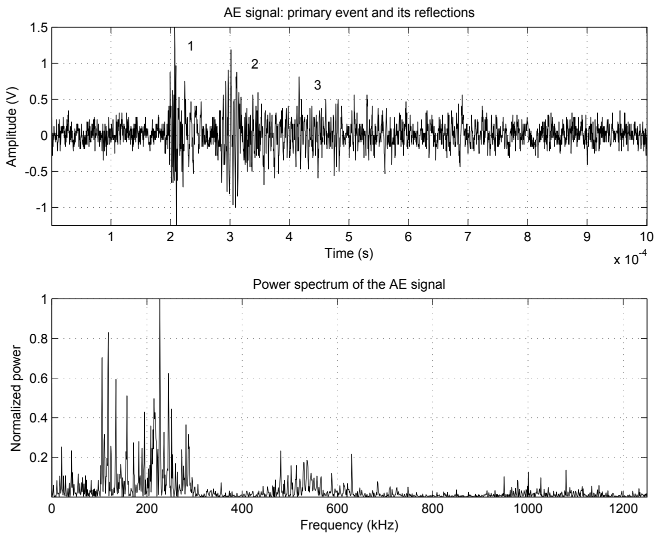

Fig. 1. One of the 20 AE events and its associated spectrum. Usually, these are the signals under study which constitute a main perturbation and its associated reflections. The main event (1) and two reflections (2,3) can be seen.

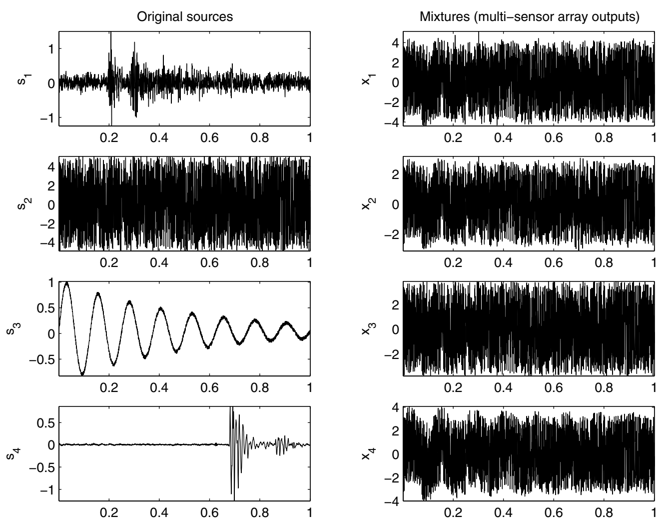

Fig. 2. Left column: One of the 20 quartets of original sources to be mixed, which in turn constitutes one of the 20 inputs to the ICA algorithms. Right column: The linear mixtures.

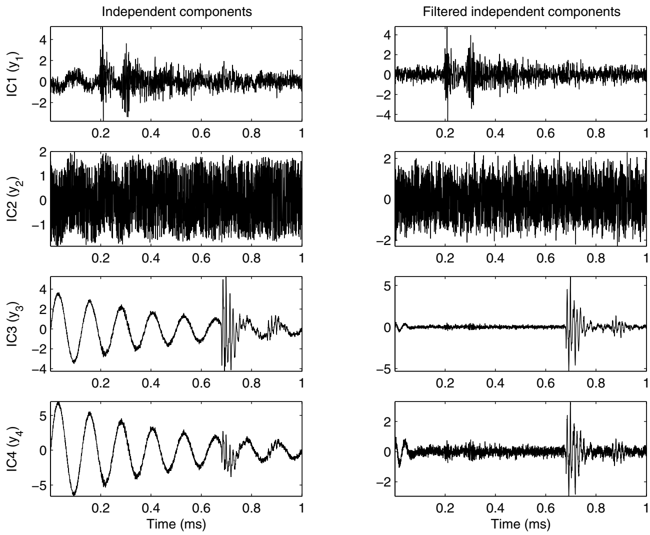

Fig. 3. Estimated and filtered sources via CumICA (ICs; Independent Components). Left column: AE event, noise, damping sine wave plus impulse, idem. Right column: Filtered signals.

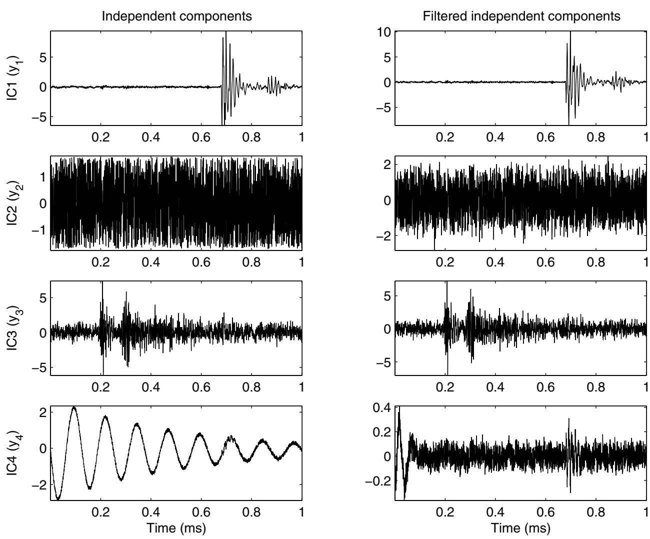

Fig. 4. Estimated and filtered sources (independent components, ICs) via FastICA. Right column (very similar to the left) top to bottom: Impulse, noise, AE event, noise. Post-filtering is not necessary to recover the AE event and the impulse.

column of both figures shows how for any IC, the values are quite random. This means that for a value (a point in the signal-to-signal graphic) of an IC, almost all the values of the another IC are allowed. On the other hand, the joint

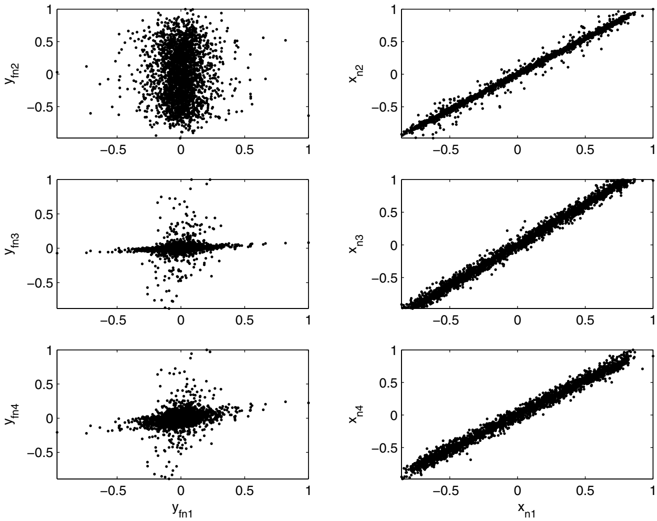

Fig. 5. Signal-to-signal diagram for the CumICA outputs. Left column: Independent components. Right column: Mixtures.

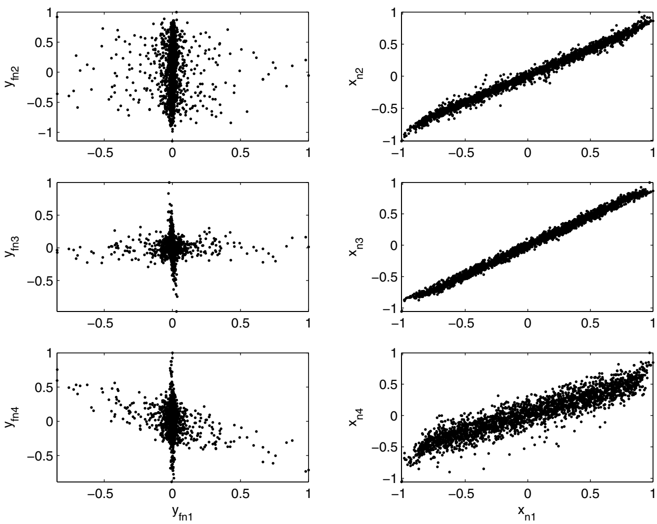

Fig. 6. Signal-to-signal diagram for the FastICA outputs. Left column: Independent components. Right column: Mixtures.

distributions of the mixtures are linearly shaped, which leads us to infer a dependency before separating sources by ICA.

These results lead us to conclude about the use of the algorithms.

5 Conclusions and Future Work

ICA is far different from traditional methods used to separate sources or to de-noise signals, as power spectrum or wavelet transforms, which obtain an energy diagram of the different frequency components, with the risk that low-level sounds or events could be masked. This experiment shows that both algorithms are able to separate the sources with small energy levels in comparison to the background noise. This is explained away by statistical independence basis of ICA, regardless of the energy associated to each frequency component. The post filtering action applied to Cum-ICA lets us work with very low SNR signals. FastICA kernel maximizes the non-Gaussianity, so it is not necessary a filter stage.

The next step regarding this research is oriented in a double direction. First, a stage involving four real mixtures will be developed. Secondly, and simultaneously, the computational complexity of the algorithms have to be reduced to perform a real implementation in a digital signal processor.

Acknowledgement

The authors would like to thank the *Spanish Ministry of Education and Science* for funding the projects DPI2003-00878, TEC2004-06096 and PTR1995-0824-OP.

References

1. De la Rosa, J.J.G., Lloret, I., Ruzzante, J., Piotrkowski, R., Armeite, M., Pumarega, M.L.: Higher-order characterization of acoustic emission signals. In: CIMSA 2005, Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applicacions, Giardini Naxos, Italy, 20-22 July 2005, ISBN 0-7803-9026-1; IEEE Catalog Number 05EX1037 (2005) 296–300 Paper CM5027. Oral Presentation in the Session 16 Advanced Signal Processing 2.
2. De la Rosa, J.J.G., Puntonet, C.G., Lloret, I.: An application of the independent component analysis to monitor acoustic emission signals generated by termite activity in wood. Measurement (Ed. Elsevier) **37** (2005) 63–76 Available online 12 October 2004.
3. Piotrkowski, R., Gallego, A., Castro, E., García-Hernández, M., Ruzzante, J.: Ti and Cr nitride coating/steel adherence assessed by acoustic emission wavelet analysis. Non Destructive Testing and Evaluation (NDT and E) International (Ed. Elsevier) **8** (2005) 260–267
4. De la Rosa, J.J.G., Puntonet, C.G., Lloret, I., Górriz, J.M.: Wavelets and wavelet packets applied to termite detection. Lecture Notes in Computer Science (LNCS) **3514** (2005) 900–907 Computational Science - ICCS 2005: 5th International Conference, GA Atlanta, USA, May 22-25, 2005, Proceedings, Part I.
5. Puntonet, C.G., De la Rosa, J.J.G., Lloret, I., Górriz, J.M.: Recognition of insect emissions applying the discrete wavelet transform. Lecture Notes in Computer Science (LNCS) **3686** (2005) 505–513 Third International Conference on Advances in Pattern Recognition, ICAPR 2005 Bath, UK, August 22-25, 2005, Proceedings, Part I.

6. Mansour, A., Barros, A.K., Onishi, N.: Comparison among three estimators for higher-order statistics. In: The Fifth International Conference on Neural Information Processing, Kitakyushu, Japan (1998)
7. Hyvärinen, A., Oja, E.: Independent Components Analysis: A Tutorial. Helsinki University of Technology, Laboratory of Computer and Information Science (1999)
8. De la Rosa, J.J.G., Puntonet, C.G., Górriz, J.M., Lloret, I.: An application of ICA to identify vibratory low-level signals generated by termites. Lecture Notes in Computer Science (LNCS) **3195** (2004) 1126–1133 Proceedings of the Fifth International Conference, ICA 2004, Granada, Spain.
9. Swami, A., Mendel, J.M., Nikias, C.L.: Higher-Order Spectral Analysis Toolbox User's Guide. (2001)