INDEPENDENT COMPONENT ANALYSIS: COMPARISON OF ALGORITHMS ON SIMULATED FMCG DATA

S. Comani*, D. Mantini**, K.E. Hild*** and G. Alleva*

* ITAB-Institute of Advanced Biomedical Technologies, University Foundation "G. D'Annunzio", Chieti University, Italy
** Department of Informatics and Automation Engineering, Marche Polytechnic University, Ancona, Italy
*** CNEL-Computational Neuroengineering Laboratory, University of Florida, Gainesville, Florida, USA

comani@itab.unich.it

Abstract: We quantitatively studied the performances of six ICA algorithms (FastICA, JADE, Infomax, CubICA, TDSEP and MRMI-SIG) in the separation of signal components embedded in fetal magnetocardiographic recordings, with particular attention to the reconstruction of fetal cardiac signals. Synthetic datasets of increasing complexity were prepared, and real fMCG recordings were simulated with linear combinations of usual fMCG source signals: maternal and fetal cardiac activity, ambient noise, maternal respiration and sensor spikes. The signal-to-interference ratio (SIR) was calculated for all separated components and for fetal traces in particular. Computation times for fetal signals of at least 20 dB SIR were calculated for all algorithms. No significant dependency on gestational age or cluster dimension was observed. Infomax, TDSEP and MRMI-SIG were sensitive to additive noise. FastICA, CubICA and JADE showed the best overall performances, FastICA having the best values of fetal SIR and very short computation times.

Introduction

Independent Component Analysis (ICA) is a signal processing technique able to recover mutually independent but otherwise unknown original source signals from their linear instantaneous mixtures [1-5]. ICA has been lately employed also in the field of biomedical signal processing [6], in particular when biomedical signals recorded with multi-channel devices need to be separated into their components [7-10].

A recent biomedical application of ICA regards the processing of datasets recorded with fetal magnetocardiography (fMCG) [11,12], which is a noninvasive technique useful for the prenatal assessment of the fetal heart function and fetal well-being in a variety of clinical situations [13-17]. Indeed, fMCG recordings are linear mixtures of signals related to fetal cardiac activity, maternal cardiac function and environmental magnetic noise. We demonstrated in previous papers the reliability of ICA for the retrieval of high-quality fetal cardiac signals from fMCG recordings [18-21].

The present study regards the evaluation of the performances of six different ICA algorithms, commonly used for biomedical analysis, with particular attention to fetal signals. Those performances were evaluated on simulated datasets of increasing complexity, obtained from the combination of several variables able to represent different fetal maturation and types of noise.

Materials and Methods

The solution of the general ICA problem consists in the separation of source signals that are linearly mixed in the input recordings by using vectorial operations.

Let us assume that the observed *n* random variables were generated by a linear instantaneous mixture of *m* source signals, or independent components. The ICA linear expansion is:

$$x(t) = A\, s(t) \qquad (1)$$

where $s_1(t)$, $s_2(t)$,…$s_m(t)$ are the source signals, $x_1(t)$, $x_2(t)$,…$x_n(t)$ are the instantaneous mixtures, and A denotes the [$n\times m$] *mixing matrix* [1-3].

A basic assumption of ICA is that source signals are independent, and that at most one of them may have a Gaussian distribution. The independence of source signals can be defined as:

$$f(s_1, s_2, ..., s_m) = \prod_{i=1}^{m} f_i(s_i) \qquad (2)$$

where f_i is the probability density function (pdf) of s_i, and f is the joint density of s_1, s_2,…s_m .

For systems fulfilling the ICA basic assumption, equation (1) can be solved using information contained in x(t) only, and the independent components (ICs) can be retrieved determining a [$m\times n$] matrix W, named *unmixing matrix*, such as:

$$y(t) = W\, x(t) \qquad (3)$$

ISSN: 1727-1983 © 2005 IFMBE

where the m-dimensional vector y(t) is the best estimate of the source vector s(t). According to the ICA theory [1-3], y(t) can be found minimizing the average mutual information (AMI)

$$\mathrm{AMI} = \int p_y(u) \log \frac{p_y(u)}{\prod_{i=1}^{n} p_{yi}(u)} du \tag{4}$$

where p_y is the probability distribution function associated with y(t), and p_{yi} is the probability distribution function associated with $y_i(t)$.

A set of multichannel fMCG recordings satisfies the requirements of the ICA problem, because a large number of simultaneous observations are available for the solution of the ICA problem, and also because the source signals, produced by the maternal and fetal hearts, are independent signals and linearly mixed in the recordings [18].

In the present study, six ICA algorithms were investigated; they were three classical ICA algorithms (FastICA [22], JADE [23], Infomax [24]) and three more recent algorithms (CubICA [25], TDSEP [26], and MRMI-SIG [27]). Each algorithm used a different approach to solve equation 3, hence to minimize AMI.

FastICA: the fixed-point ICA algorithm (FastICA) minimizes AMI by maximizing the normalized differential entropy J of the estimated source signals, which is called *negentropy* [1, 28]; it is defined as the difference between the entropy of a Gaussian random variable y_{gauss} with the same variance of the observed random variable y and the entropy of y:

$$J(y) = H(y_{gauss}) - H(y) \tag{5}$$

being the entropy H of a random vector y(t) with density $p_y(.)$ defined as:

$$H(y) = -\int p_y \log(p_y) dy \tag{6}$$

JADE: the Joint Approximate Diagonalization of Eigen-matrices (JADE) estimates the uncorrelation and statistical independence of sources respectively with the reduction of the second-order and fourth-order cumulants ($C^{(y)}_{\alpha\beta}$ and $C^{(y)}_{\alpha\beta\gamma\delta}$) to zero.

Infomax: The Bell-Sejnowski algorithm is a gradient-based neural network algorithm in which the signal information is maximized (hence the name Infomax). Information can be expressed through the entropy H(y) of a transformed signal y=*g*(x), where *g* is a non-linear function and H(y) can be written as:

$$H(y) = E[\ln|J|] + H(x) \tag{7}$$

The maximization of H(y) can be achieved by maximizing only the first term $E[\ln|J|]$, i.e. by changing the un-mixing matrix W. A neural gradient training algorithm is used to this purpose.

CubICA: based on the Comon's theory, the Cumulant Based Independent Component Analysis (CubICA) uses the diagonalization of cumulant tensors, taking the third-order and fourth-order cumulant tensors ($C^{(y)}_{\alpha\beta\gamma}$ and $C^{(y)}_{\alpha\beta\gamma\delta}$) into account simultaneously to estimate the independence of source signals.

TDSEP: the Temporal Decorrelation source SEParation (TDSEP) algorithm performs a simultaneous diagonalization of several time-delayed correlation matrices to estimate the statistical independence of the sources. Using time delays τ = 1, 2, 3,..., the cross-covariance function C^{τ}_{x} of the signal is obtained:

$$E\{x(t,\tau)x^T(t,\tau)\} = C^{\tau}_{x} = AC^{\tau}_{s}A^T \tag{8}$$

MRMI-SIG: The MRMI-SIG algorithm [29] is a spatio-temporal ICA method that minimizes the mutual information of source signals written as the sum of (Shannon) marginal entropies minus the (Shannon) joint entropy:

$$I_1(Y) = \sum_{m=1}^{M} H_1(Y_m) - H_1(Y) \tag{9}$$

where Y_m is the random variable that represents the m^{th} estimated source and Y is the M-dimensional random vector representing all estimated sources. To reduce estimator variance and to significantly decrease the computational complexity [30], MRMI-SIG replaces Shannon's definition of entropy, $H_1(Y)$, with Renyi's quadratic entropy, $H_2(Y)$. Renyi's joint entropy is invariant to rotations, so the criterion reduces to a sum of marginal entropies,

$$I_2(Y) = \sum_{m=1}^{M} H_2(Y_m) \tag{10}$$

where each of the marginal entropies is estimated using non-parametric probability distribution function (pdf) estimation.

Whenever possible, we used the original MATLAB codes provided by the authors; the algorithms' performances were tested on synthetic fMCG datasets.

Simulated datasets: For synthetic data, the source signals are known, as well as the mixing matrix A. In this case, the un-mixing system W provided by the ICA algorithm can be assessed using the known mixing system A, and the unmixed signals y_i can be evaluated using the known source signals s_i. This setting allows estimating separation performances of different algorithms.

Five independent components were taken into account to represent real fMCG recordings: maternal and fetal cardiac signals, which were directly reconstructed from real datasets, environmental noise, maternal respiration and sensor spikes, the last three components being synthesized respectively as a gaussian signal, as a sine wave, and as a triangular function. Signal duration was 1 minute, and sampling frequency was set at 1 kHz. Different gestational ages were represented with fetal signals with amplitudes of 0.3 pT, 0.8 pT, 1.4 pT and 2.0 pT, corresponding to 24, 28, 32, 36 weeks of gestation on average. For each fetal signal amplitude, 3 synthetic fMCG datasets were prepared using appropriate mixing matrices, based on realistic source intensities, which permitted to reproduce clusters of 55 traces; the different combinations of components are shown in Table 1. Figure 1 shows examples of four synthetic fMCG recordings belonging to dataset 3, fetus at 36 weeks.

Table 1: Summary of the signals used to produce the simulated fMCG datasets.

	dataset 1	**dataset 2**	**dataset 3**
maternal cardiac signal	X	X	X
fetal cardiac signal	X	X	X
environmental noise	X	X	X
maternal respiration		X	X
sensor spikes			X

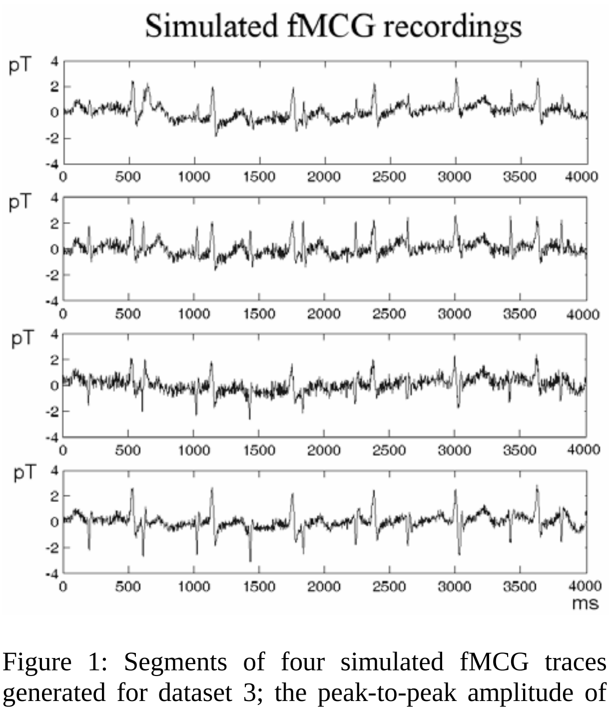

Figure 1: Segments of four simulated fMCG traces generated for dataset 3; the peak-to-peak amplitude of the fetal signal was 2.0 pT on average. Time is given in milliseconds.

Simulated datasets were centered, and then PCA was used to whiten the datasets and to reduce their dimensions. Dimension reduction was required because some of the analyzed ICA algorithms are only able to estimate a square [$n \times n$] un-mixing matrix W. Through dimension reduction we obtained a new set of n recordings and a new [$n \times n$] mixing matrix A' for every ICA algorithm, without any lack of information as to the independent components.

Algorithms' performances were evaluated through the accuracy with which each algorithm was able to separate components, with particular attention to the fetal ones.

For each ICA algorithm, a vector containing the estimated source signals $y_i(t)$ was available. The distortion of a source signal $s_i(t)$ in its estimate $y_i(t)$ was measured, and the accuracy of each ICA algorithm was then quantified in terms of signal-to-interference ratio (SIR_i), which was defined as:

$$SIR_i = -10 \log_{10} \frac{\left\| e_{interf} \right\|^2}{\left| \left\langle y_i, s_i \right\rangle \right|^2} \tag{11}$$

where e_{interf} was the interference term according to the method proposed by Gribonval et al. [31].

For each algorithm and each dataset, we calculated both the average SIR value among all separated components and the fetal SIR.

Results

The results of separation performance are given as a function of dataset identifier and weeks of gestation. For each item the SIR given was calculated with respect to the other item. Both the average SIR (averaged over all source estimates) and the SIR pertaining to only the fetal source are shown.

For each algorithm, the average SIR (upper row) and the fetal SIR (lower row) are given as a function of the dataset identifier in Table 2. The values obtained for the average SIR are comparable with the exception of the Infomax algorithm, which performs poorly as soon as the maternal respiration source is included. The results for the fetal SIR indicate that FastICA performs noticeably better than the others.

Table 3 shows the separation performances as a function of the number of gestational weeks. The average SIR of FastICA, CubICA, JADE, and MRMI-SIG are comparable, whereas the results for TDSEP are slightly lower and Infomax is much lower. The fetal SIR results show a clear preference for FastICA.

All algorithm reached convergence in an interval ranging from 0.04 seconds to 5.30 seconds.

Discussion

The outcomes of our study indicate TDSEP, MRMI-SIG and Infomax as the less performing algorithms. Infomax presents the worst figures as soon as the maternal respiration source is included in the datasets, which is not surprising since maternal respiration is modeled with a sinusoid (bimodal distribution) while the Infomax implementation is usually tuned for unimodal super-Gaussian sources.

ISSN: 1727-1983 © 2005 IFMBE

Table 2: Average SIR (upper row) and fetal SIR (lower row) values for each analyzed ICA algorithm; figures, in dB, are given in function of the dataset complexity, expressed by the dataset identifier.

	dataset 1	**dataset 2**	**dataset 3**
FastICA	37,69 44,57	36,70 42,82	36,74 39,32
JADE	37,13 32,45	33,33 31,79	33,72 31,53
Infomax	34,08 34,49	3,19 34,02	4,16 32,96
CubICA	37,38 33,03	33,37 31,93	34,04 31,96
TDSEP	33,60 29,18	34,63 28,94	34,85 28,87
MRMI-SIG	37,97 33,49	37,69 32,66	37,58 32,54

The drop in Infomax performance is clearly visible in Tables 2 and 3, which show average SIR performances as a function of dataset identifier and weeks of gestation. On the other hand, Infomax performs quite well in terms of fetal SIR: this is possible because the fetal signal is super-Gaussian.

Table 3: Average SIR (upper row) and fetal SIR (lower row) values for each analyzed ICA algorithm; figures, in dB, are given in function of fetal gestational age, expressed in weeks.

Gestational weeks	**24**	**28**	**32**	**36**
FastICA	37,04 41,21	37,04 41,89	37,05 42,11	37,05 42,17
JADE	34,53 31,83	34,74 32,16	34,80 31,97	34,80 31,87
Infomax	11,79 33,83	11,82 33,82	11,83 33,81	11,83 33,82
CubICA	34,77 31,88	34,91 33,27	35,39 33,47	34,83 31,98
TDSEP	31,43 27,06	32,33 28,88	32,34 28,97	32,32 28,98
MRMI-SIG	35,31 29,89	37,41 32,79	37,86 32,84	37,61 32,93

TDSEP and MRMI-SIG had difficulty extracting the fetal signal buried in noise when the energy of the fetal signal is low, which explains the results obtained for datasets representing acquisitions at 24 weeks of gestation. It is worth noting that TDSEP and MRMI-SIG are the only algorithms considered herein that use temporal information for separation.

Therefore, all the above-mentioned algorithms are sensitive to noise.

On the other hand, FastICA, JADE and CubICA have performances that are fairly independent on noise (represented by the dataset identifier) and on gestational age (Tables 2 and 3); in fact, these algorithms always present the best overall performances. Moreover, they provide similar results in terms of average SIR and fetal SIR, although the FastICA algorithm always shows the best separation performance pertaining to the fetal SIR.

With regard to the computation time, all algorithms were very fast. However, Infomax and MRMI-SIG were the slowest algorithms, while TDESP and FastICA had the best computation times for all datasets.

Conclusions

The results of this study provide a tool for the selection of the most appropriate ICA algorithm to process fMCG datasets in order to extract fetal cardiac signals with best signal quality. Among the analyzed algorithms, FastICA seems to have the best overall performances, as far as we could verify on the simulated datasets.

References

[1] COMON P. (1994): "Independent component analysis - a new concept?", *Signal Process.*, **36**, pp. 287-314
[2] LEE T.W. (1998): "Independent component analysis: theory and applications", (Kluwer Academic Publishers, Boston)
[3] HYVÄRINEN A., KARHUNEN J. and OJA E. (2001): "Independent component analysis", (John Wiley and Sons, New York)
[4] AMARI S., CICHOCKI A. and YANG H.H. (1996): "A new learning algorithm for blind signal separation", in TESAURO G., TOURETZKY D.S., LEEN T.K. (Eds): "Advances in Neural Information Processing Systems", (The MIT Press, Cambridge MA), pp. 757-763
[5] BIFULCO P., CESARELLI M., ALLEN R., SANSONE M., and BRACALE M. (2001): "Automatic Recognition of Vertebral Landmarks in Fluoroscopic Sequences for Analysis of Intervertebral Kinematics", *Med. Biol. Eng. Comput.*, **39**, pp. 65-75
[6] JAMES C.J. and HESSE C.W. (2005): "Independent component analysis for biomedical signals", *Physiol Meas*; **26**, pp. R15-R39
[7] MAKEIG S., BELL A.J., JUNG T.P. and SEJNOWSKI T.J. (1996) "Independent component analysis of electroencephalographic data" in TESAURO G., TOURETZKY D.S., LEEN T.K. (Eds): "Advances in Neural Information Processing Systems", (The MIT Press, Cambridge MA), pp. 145-151
[8] VIGARIO R., JOUSMÄKI V., HÄMÄLÄINEN M., HARI R. and OJA E. (1998) "Independent component analysis for identification of artifacts in magnetoencephalographic recordings", in "Advances in Neural Information Processing", (The MIT Press, Cambridge MA), pp. 229-235
[9] MCKEOWN M.J., MAKEIG S., BROWN G.G., JUNG T.P., KINDERMANN S.S. and SEJNOWSKI T.J. (1998): "Analysis of fMRI by Blind Separation

into Independent Spatial Components", *Human Brain Mapping*, **6**, pp.160–88

[10] ESPOSITO F., SCARABINO T., HYVARINEN A., HIMBERG J., FORMISANO E., COMANI S., TEDESCHI G., GOEBEL R., SEIFRITZ E. and DI SALLE F. (2005): "Independent component analysis of fMRI group studies by self-organizing clustering", *Neuroimage*; **25**, pp.193-205

[11] PETERS M., CROWE J., PIERI J.F., QUARTERO H., HAYES-GILL B., JAMES D., STINSTRA J. and SHAKESPEARE S. (2001): "Monitoring the fetal heart non-invasively: a review of methods", *J Perinat Med*, **29**, pp. 408-416

[12] LEWIS M.J. (2003): "Review of electromagnetic source investigations of the fetal heart", *Med Eng Phys*, **25**, pp. 801-810

[13] WAKAI R.T., LEUTHOLD A.C., CRIPE L. and MARTIN C.B. (2000): "Assessment of fetal rhythm in complete congenital heart block by magnetocardiography", *Pacing Clin Electrophysiol*, **23**, pp.1047-1050

[14] HOSONO T., KAWAMATA K., CHIBA Y., KANDORI A. and TSUKADA K. (2002): "Prenatal diagnosis of long QT syndrome using magnetocardiography: a case report and review of the literature", *Prenat Diagn*, **22**, pp.198-200

[15] KÄHLER C., SCHLEUSSNER E., GRIMM B., SCHNEIDER U., HAUEISEN J., VOGT L. and SEEWALD H.J. (2002): "Fetal magnetocardiography in the investigation of congenital heart defects", *Early Hum Dev*, **69**, pp. 65-75

[16] GRIMM B., KÄHLER C., SCHLEUSSNER E., SCHNEIDER U., HAUEISEN J. and SEEWALD H.J. (2003): "Influence of intrauterine growth restriction on cardiac time intervals evaluated by fetal magnetocardiography", *Early Hum Dev*, **74**, pp. 1-11

[17] WAKAI R.T., STRASBURGER J.F., LI Z., DEAL B.J. and GOTTEINER N.L. (2003): "Magnetocardiographic rhythm patterns at initiation and termination of fetal supraventricular tachycardia", *Circulation*, **107**, pp. 307-312

[18] COMANI S., MANTINI D., PENNESI P., LAGATTA A. and CANCELLIERI G. (2004): "Independent component analysis: fetal signal reconstruction from magnetocardiographic recordings", *Comput Meth Prog Bio*, **75**, pp. 163-177

[19] MANTINI D., COMANI S., PENNESI P. and CANCELLIERI G. (2004): "Tailoring of the Independent Component Analysis to multi-channel fMCG recordings for an optimal reconstruction of the fetal cardiac signal", *Biomed Tech*, **48**, pp. 186-188

[20] COMANI S., MANTINI D., LAGATTA A., ESPOSITO F., DI LUZIO S. and ROMANI G.L. (2004): "Time course reconstruction of fetal cardiac signals from fMCG: Independent Component Analysis vs. Adaptive Maternal Beat Subtraction", *Physiol Meas*, **25**, pp. 1305-1321

[21] COMANI S., MANTINI D., ALLEVA G., DI LUZIO S. and ROMANI G.L. (2004): "Fetal Magnetocardiographic Mapping using Independent Component Analysis", *Physiol Meas*, **25**, pp. 1459-1472

[22] HYVÄRINEN A. (1999): "Fast and robust fixed-point algorithms for independent component analysis", *IEEE Trans. on Neural Networks*, **10**, pp. 626-634

[23] CARDOSO J.F. and SOULOUMIAC A. (1996): "Jacobi angles for simultaneous diagonalization", *J. Mat Anal Appl.*, **17**, pp: 161-164

[24] BELL A.J. and SEJNOWSKI T.J. (1995): "An information-maximization approach to blind separation and blind deconvolution", *Neural Computation*, **7**, pp. 1129-1159

[25] BLASCHKE T. and WISKOTT L. (2004): "CuBICA: Independent Component Analysis by Simultaneous Third- and Fourth-Order Cumulant Diagonalization", *IEEE Transactions on Signal Processing*, **52**, pp. 1250-1256

[26] ZIEHE A., MULLER K.R., NOLTE G., MACKERT B.M. and CURIO G. (2000): "Artifact reduction in magnetoneurography based on time-delayed second order correlations", *IEEE Trans. on Biomedical Engineering*, **41**, pp. 75-87

[27] HILD K.E., ERDOGMUS D. and PRINCIPE J.C. (2001): "Blind Source Separation Using Renyi's Mutual Information", *IEEE Signal Processing Letters*, **8**, pp. 174-176

[28] HYVÄRINEN A. (1998): "New approximations of differential entropy for independent component analysis and projection pursuit" in: "Advances in Neural Information Processing", (The MIT Press, Cambridge MA), pp. 273–279

[29] HILD II K.E., ERDOGMUS D. and PRINCIPE J.C. (2002): "Blind Source Separation of Time-Varying, Instantaneous Mixtures Using an On-Line Algorithm," *Intl. Conf. on Acoustics, Speech, and Signal Processing* (ICASSP '02), Orlando, FL, pp. 993-996

[30] HILD II K.E., ERDOGMUS D. and PRINCIPE J.C. (2005): "An Analysis of Entropy Estimators for Blind Source Separation," accepted to *Signal Processing*

[31] GRIBONVAL R., VINCENT E., FÉVOTTE C. and BENAROYA L. (2003): "Proposals for performance measurement in source separation" in: *Proceedings of ICA*, Nara, Japan.