Journal of ELECTRICAL ENGINEERING, VOL. 63, NO. 6, 2012, 380–385

APPLICATION OF INDEPENDENT COMPONENT ANALYSIS FOR SPEECH–MUSIC SEPARATION USING AN EFFICIENT SCORE FUNCTION ESTIMATION

Arash Pishravian — Masoud Reza Aghabozorgi Sahaf *

In this paper speech-music separation using Blind Source Separation is discussed. The separating algorithm is based on the mutual information minimization where the natural gradient algorithm is used for minimization. In order to do that, score function estimation from observation signals (combination of speech and music) samples is needed. The accuracy and the speed of the mentioned estimation will affect on the quality of the separated signals and the processing time of the algorithm. The score function estimation in the presented algorithm is based on Gaussian mixture based kernel density estimation method. The experimental results of the presented algorithm on the speech-music separation and comparing to the separating algorithm which is based on the Minimum Mean Square Error estimator, indicate that it can cause better performance and less processing time.

K e y w o r d s: independent component analysis, speech-music separation, score function estimation, mutual information

1 INTRODUCTION

Usually existing audio signals are combinations of several audio source signals which are mixed at the same time. By progress in the speech and signal processing techniques, the vital need to audio separation is so much sensible. The most important audio sources are the speech and music, thus separating them is included in many applications, such as automatic speech recognition, speech enhancement, music information retrieval and electronic music composition. For attaining to speech-music discriminating various techniques have been used by now [1–4], but in these techniques usually many special assumptions have been considered on signals or combination. Recently after introducing blind source separation techniques, it is obvious that they can be used for this problem.

Blind source separation is a basic and challenging research problem in signal processing which has been firstly introduced by Herault and Jutten [5] and has received a great deal of attention in recent years, with a broad range of applications. BSS consists of recovering source signals from several observed mixtures of them. The observations are obtained from a set of sensors, each receiving a different combination of source signals. The problem is called "blind" because no information is available about the mixture. Thus far, the problem of the BSS has been solved using various techniques and algorithms. The lack of prior information must be compensated by considering some special assumptions. The most popular condition used in BSS techniques is the statistical independence of source signals. The goal in these techniques is to achieve a separation process that produces most independent outputs, so is called independent component analysis (ICA) [6, 7]. Various criterions of independence cause various algorithms in the ICA method [8–10]. The criteria that is used in this paper to measure the independence is the outputs' mutual information, which is demonstrated in [7] that source separation based on minimization of the mutual information is asymptotically a Maximum Likelihood (ML) estimation of the sources.

The paper is organized as follows: In Section 2 the problem formulation is presented. Section 3 expresses the score function estimation. The separation algorithm is explained in Section 4. In Sections 5 the experimental results are presented, and concluding remarks are given in Section 6.

2 PROBLEM FORMULATION

In this section a model for the problem and some notions of blind identification are presented.

2.1 The Model

Assume that d signals $\mathbf{S}(t) = \big(s_1(t), \ldots, s_N(t)\big)^\top$ are transmitted from d sources. Considering a narrowband time-invariant channel, what we receive at N sensors will be the instantaneous linear combinations of these signals that construct measured data, *ie* N sensor signals $\mathbf{X}(t) = \big(x_1(t), \ldots, x_N(t)\big)^\top$

$$\mathbf{X}(t) = \boldsymbol{a}_1 s_1(t) + \cdots + \boldsymbol{a}_d s_d(t)\,. \tag{1}$$

Thus the model is

$$\mathbf{X}(t) = \mathbf{A} \cdot \mathbf{S}(t) \tag{2}$$

* Signal Processing Research Group, Electrical and Computer Eng. Dep., Yazd University, Iran.
ar_pishravian@yahoo.com, aghabozorgi@yazduni.ac.ir

DOI: 10.2478/v10187-012-0056-0, ISSN 1335-3632

where $\boldsymbol{x}(t) \in \Re^{m\times 1}$ is the measured data vector, $\boldsymbol{s}(t) \in \Re^{d\times 1}$ is the signal vector, $\mathbf{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d] \in \Re^{m\times d}$ characterizes the unknown channel and is referred to as the "mixing matrix".

Above model is instantaneous BSS model. The aim of separating algorithm is to construct a separating matrix $\mathbf{B}$ such that the components of the vector $\mathbf{Y}(t) = \mathbf{B}\mathbf{X}(t)$ contain an estimate of original signals, up to a few indeterminacies (scale and permutation). If we suppose that N mixed signals are linear combination of unknown mutually statistically independent source signals, as mentioned above the problem is called independent component analysis (ICA). In this case, the goal is to achieve a separation process that produces most independent outputs, in other word the independence of the outputs insures the separation of the sources.

It is obvious that our problem can be coincided with above mentioned model. Considering speech and music as two original sources that have been combined in an unknown channel, and their combination (at least by two receivers) have been recorded, we can use a 2x2 BSS(ICA) model for problem.

2.2 Independence Identification

As mentioned in ICA, separation process is achieved by obtaining outputs that are independent as most as possible. So in ICA we need a technique that measures the independence. By now various techniques for measuring independence and corresponding various ICA algorithms have been introduced such as JH alg. [5], Cumulant based, HOS (Higher Order Statistics), FOBI (Fourth Order Blind Identification) and using IM (Information Maximization) and ML (Maximum Likelihood), MI (Mutual Information) score functions [8–13].

In this paper for measuring the independence the mutual information can be used as below

$$I(X) = \int_X P_X(X) \operatorname{Ln} \frac{P_X(X)}{\prod_i P_{x_i}(x_i)} \mathrm{d}X = \sum_i H(x_i) - H(x) \quad (3)$$

where p_x and p_{x_i} are the PDFs of X and x_i respectively and H denotes the Shannon's entropy. According to above equation, we conclude that the mutual information is always non-negative and it is zero if and only if $P_X(X) = \prod_{i=1}^{N} p_{x_i}$, i.e. when $x_1, \ldots, x_N$ are independent. In fact, for separating the sources, we must minimize the mutual information of the outputs.

3 SCORE FUNCTION ESTIMATION

Because the estimation of the score function has an important role in separating algorithm, in this section score function and two methods for it's estimation introduced.

3.1 Score Function

Score function of a random variable x is introduced as [14]

$$\psi_x(x) = -\frac{\mathrm{d}}{\mathrm{d}x} \ln p_x(x) = -\frac{p'_x(x)}{p_x(x)} \quad (4)$$

where p_x is pdf of x.

For a random vector $X = (x_1, \ldots, x_N)^\top$ two score functions are introduced:
– MSF (Marginal Score Function)

$$\psi_X(X) = \big(\psi_1(x_1), \psi_2(x_2), \ldots, \psi_N(x_N)\big)^\top \quad (5)$$

where $\psi_i(x_i)$ is score function of i-th component.
– JSF (Joint Score Function)

$$\varphi_X(X) = \big(\varphi_1(X), \varphi_2(X), \ldots, \varphi_N(X)\big)^\top \quad (6)$$

where

$$\varphi_i(x) = -\frac{\partial}{\partial x_i} \ln p_X(X) = -\frac{\frac{\partial}{\partial x_i} p_X(X)}{p_X(X)}\,. \quad (7)$$

The difference between marginal score function and joint score function is called Score Function Difference (SFD)

$$\beta_X(X) = \psi_X(X) - \varphi_X(X)\,. \quad (8)$$

This function contains some information about independence of vector components. Score functions have some properties that are used in next section such as:
(a) Components of vector $X = (x_1, \ldots, x_N)^\top$ are independent if and only if $\beta_X(X) = 0$, *ie* $\psi_X(X) = \varphi_X(X)$.
(b) If X is a bounded vector then $E\{\varphi_X(X)X^\top\} = \mathbf{I}$.

As it will be mentioned in the next section, the algorithm at each iteration needs to estimation of score function hence the speed and accuracy of the estimation will affect on the quality of the separated signals and the processing time of algorithm.

The utilized method of our algorithm for score function estimation is based on the Gaussian mixture (GM) model which is introduced in the next section. To show the proper performance of the presented estimation, the results are compared with Minimum Mean Square Error (MMSE) method which is a usual approach in score function estimation. In addition this method is explained in Section 3.3.

3.2 Gaussian Mixture Estimator

A general class of density models is the Gaussian mixture model. This method models the unknown density with a sum of Gaussian kernels as in the following form [15]

$$f(y) = \sum_{k=1}^{m} \pi_k g\big(y, \mu_k, \sigma_k^2\big) \quad (9)$$

where $g(\cdot)$ is the Gaussian kernel with center mean μ_k and variance σ_k^2.

$$g\big(y,\mu_k,\sigma_k^2\big)=\frac{1}{\sqrt{2\pi\sigma_k^2}}\exp\frac{(y-\mu_k)^2}{2\sigma_k^2}\,. \tag{10}$$

Under the Gaussian mixture model, the score function of y reads

$$\psi(y)=\frac{\sum_{k=1}^m \psi_k g(y,\mu_k,\sigma_k^2)\frac{y-\mu_k}{\sigma_k^2}}{f(y)} =\sum_{k=1}^m P(k|y)\frac{y-\mu_k}{\sigma_k^2} \tag{11}$$

for any y it holds $\sum_{k=1}^m P(k|y)=1$.

For our estimation model we use $m=400$, with their centers equidistantly positioned over the range of y and with constant variance σ^2. If $\mu_1\le\min(y)$ is the center of the first kernel then the center of the k- th kernel is $\mu_k=\mu_1+(k-1)\delta$ and $\delta=(\mu_m-\mu_1)/(m-1)$ is the distance between centers. In this case the score function becomes

$$\psi(y)=\frac{1}{\sigma^2}\Big[y-\sum_{k=1}^m P(k|y)\mu_k\Big]. \tag{12}$$

Our choice for variance is $\sigma=n^{-2/7}std(y)$ where $std(y)$ is the standard deviation of each random variable y. To estimate the score function in (12) on all points y_i, we need to compute for all points the quantity $\sum_{k=1}^m P(k|y)\mu_k=h(y)/f(y)$, where

$$\begin{aligned} h(y)&=\sum_{k=1}^m \pi_k\mu_k\exp\big[-0.5(y-\mu_k)^2/\sigma^2\big],\\ f(y)&=\sum_{k=1}^m \pi_k\exp\big[-0.5(y-\mu_k)^2/\sigma^2\big]. \end{aligned} \tag{13}$$

However, instead of computing directly by substituting the y_i in the above formulas, we can evaluate both $h(y)$ and $f(y)$ only at the points μ_k and then interpolate for computing the values of the functions at the points y_i. Using the μ_k from above, the values of $h(y)$ and $f(y)$ become

$$\begin{aligned} h(\mu_l)&=\sum_{k=1}^m \pi_k\mu_k\exp\big[-0.5(l-k)^2\delta^2/\sigma^2\big],\\ f(\mu_l)&=\sum_{k=1}^m \pi_k\exp\big[-0.5(l-k)^2\delta^2/\sigma^2\big]. \end{aligned} \tag{14}$$

Both quantities are in the form $b_l=\sum_{k=1}^m a_k g_{l-k}$ which is a discrete convolution and can be efficiently carried out with the FFT algorithm very quickly. Moreover, the mixing weights π_k can be approximated by the histogram of y [5].

3.3 MMSE Estimator

Let x be a random variable with the PDF $P_x(x)$ and $f(x)$ be a continuously differentiable function and $\lim_{x\to\pm\infty} f(x)p_x(x)=0$, then [16]

$$\mathrm{E}\{f(x)\psi(x)\}=\mathrm{E}\Big\{\frac{\partial f}{\partial x}(x)\Big\}. \tag{15}$$

The above equation shows that one can easily design a MMSE estimator for the $\psi(x)$ using the parametric function $f(x;W)$ where $W=(w_1,\dots,w_k)^\top$ denotes the parameter vector

$$\arg\min_W \mathrm{E}\big\{\big(\psi_k(x)-f(x;W)\big)^2\big\} = \arg\min_W\Big\{\mathrm{E}\{f^2(x;w)\}-2\,\mathrm{E}\Big\{\frac{\partial f}{\partial x}(x;W)\Big\}\Big\}. \tag{16}$$

For instance, we would like to estimate the $\psi(x)$ as a linear combination of $k_i(x)$, which is

$$\hat\psi(x)=\sum_{i=1}^L w_i k_i(x)=K^\top(x)W \tag{17}$$

where $W=(w_1,\dots,w_L)^\top$, $K(x)=\big(k_1(x),\dots,k_L(x)\big)^\top$.

The coefficient W must be determined such that the error term $\mathrm{E}\big\{[\psi(x)-\hat\psi(x)]^2\big\}$ [is minimized. From the orthogonality principle and equation (15) we can write

$$\mathrm{E}\big\{K(x)K^\top(x)\big\}W=\mathrm{E}\{K'(x)\}\,. \tag{18}$$

In our separating algorithm, for score function estimation we have applied the following kernels

$$K_1(x)=1\,,\ k_2(x)=x\,,\ k_3(x)=x^2,\ k_4(x)=x^3. \tag{19}$$

4 SEPARATING ALGORITHM

In the separating algorithm, the MI of the outputs $I(Y)$ has been chosen as the independence criterion. Furthermore the natural gradient approach is used to minimize the MI, *ie*, in order to mutual information minimization, we need to estimate the gradient of mutual information with respect to the parameters of the separating system, in other word derivative of MI with respect to the separating matrix is needed, which is calculated as [13]

$$\frac{\partial}{\partial B}I(Y)=\mathrm{E}\big\{\beta_Y(Y)X^\top\big\} \tag{20}$$

where $\beta_Y(Y)$ is the SFD of the vector Y. This relation is converted to the following relation in instantaneous mixtures

$$\Delta_B(I)=\frac{\partial I}{\partial B}B^\top=\mathrm{E}\big\{\beta_Y(Y)Y^\top\big\}\big\}. \tag{21}$$

> Initialization: $B = I$ and $Y = X$
> Loop:
a) Estimate $\Psi_Y(Y)$ (MSF)
b) $\Delta_B I = E\{\Psi_Y(Y)\, Y^T\} - I$
c) $B \longleftarrow (I - \mu\Delta_B I)\, B$
d) $Y = BX$
e) Normatialization:
* $y_i = y_i / \sigma_i$, where σ^2_i is the energy y_i
* Divide the i - th row of the matrix B by σ_i

Fig. 1. The iteration procedure of the separating algorithm

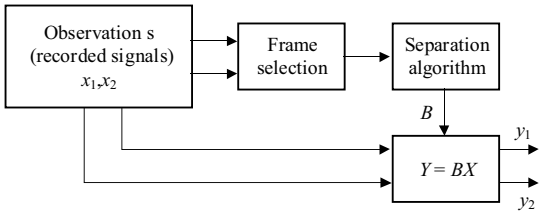

Fig. 2. The iteration procedure of the separating algorithm

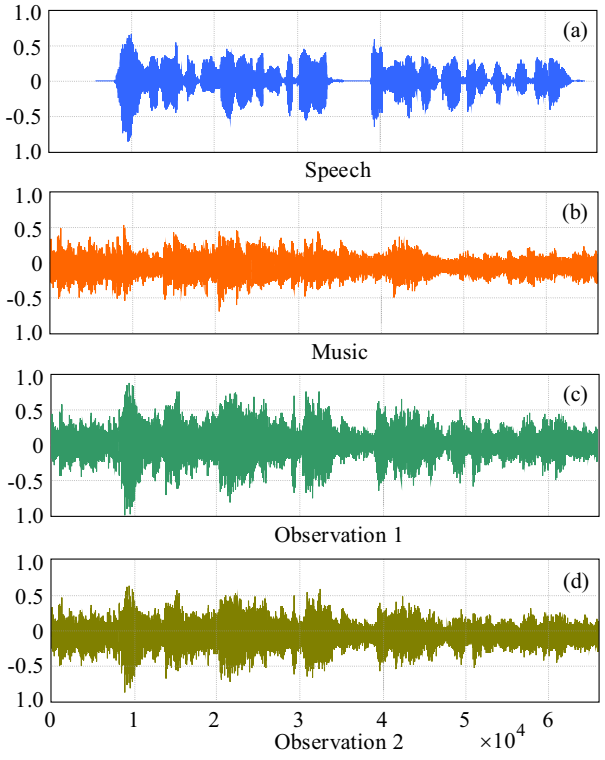

Fig. 3. (a),(b) – original speech and music signals, (c),(d) – mixed signals

Considering the definition of SFD and it's property mentioned in Section 3.1 we can rewrite (21) as

$$\Delta_B I = \mathrm{E}\{\psi_Y(Y)Y^\top\} - I \tag{22}$$

where I denotes the Identity matrix and $\psi_Y(Y)$ is the marginal score function of the vector Y, *ie*

$$\psi_y(y) = \big(\psi_1(y_1), \ldots, \psi_N(y_N)\big)^\top \tag{23}$$

where $\psi_i(y_i) = -\frac{\mathrm{d}}{\mathrm{d}y_i} \mathrm{Ln}\, p_{y_i}(y_i) = -\frac{p'_{y_i}(y_i)}{p_{y_i}(y_i)}$.

The procedure of the separating algorithm would be the way that in each iteration, the score functions (MSF) from observation signal samples is estimated at first, and then the B matrix will be updated as

$$B \longleftarrow (I - \mu\Delta_B I)B \tag{24}$$

where μ is a learning rate. Finally a normalization step is executed for the convergence of the algorithm. The mentioned procedure would be repeated as much as the algorithm is converged and stopped.

The final separating algorithm is summarized in Fig. 1.

5 EXPERIMENTAL RESULTS

To evaluate the algorithm performance, a male speech signal from FARSDAT data and a music signal have been used. In all the experiments, we have $\mu = 0.1$.

Also for comparing the two estimators, the separating algorithm has been exercised as the following two cases:Case 1) separating algorithm by MMSE estimator, Case 2) separating algorithm by GM estimator.

In both cases, at first the 3000 samples of the mixing signals (selected frame) have been used as the input of the algorithm, then the obtained separating matrix (B) from separation algorithm has been applied to whole of mixed signals (observations) to achieve the original signals, see Fig. 2.

For measuring the separation quality, we use the output SNR defined by:

$$\mathrm{SNR}_i = 10 \log \frac{\mathrm{E}\{y_i^2\}}{\mathrm{E}\{y_i^2|_{s_i=0}\}} \tag{25}$$

where $y_i|_{s_i=0}$ stands for what is at the ith output when the ith input is zero (assuming there is no permutation). By using this definition, SNR will be a measure of separation, and a high SNR means that there is not a large leakage from the other sources to the output corresponding to ith input. To be noted, the total SNR can be calculated as

$$\mathrm{SNR} = \frac{\sum_{i=1}^{N}(\mathrm{SNR}_i)}{N}\,. \tag{26}$$

In Fig. 3(a–d), 6 seconds from the original and mixed (observations) have been shown. At first, the separating algorithm using MMSE estimator (case 1) has been examined. In Fig. 4 output (separated) signals and in Fig. 6a the output SNR versus the number of the algorithm iterations have been displayed.

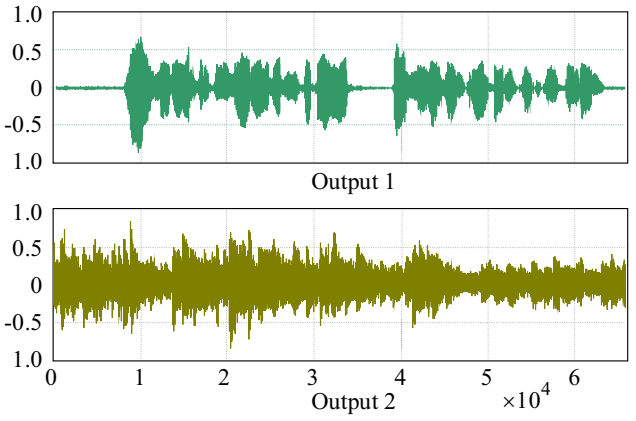

Fig. 4. Separated signals for case 1

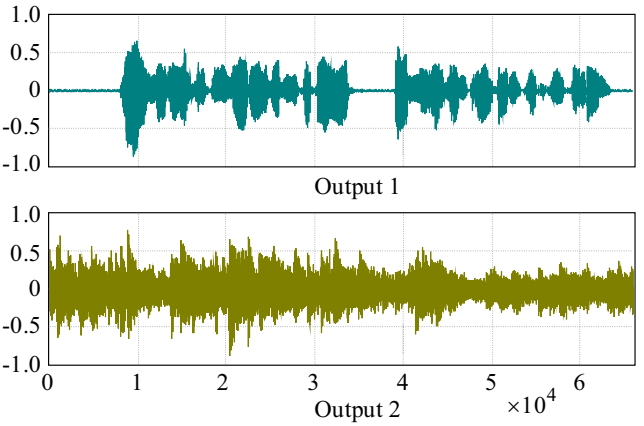

Fig. 5. Separated signals for case 2

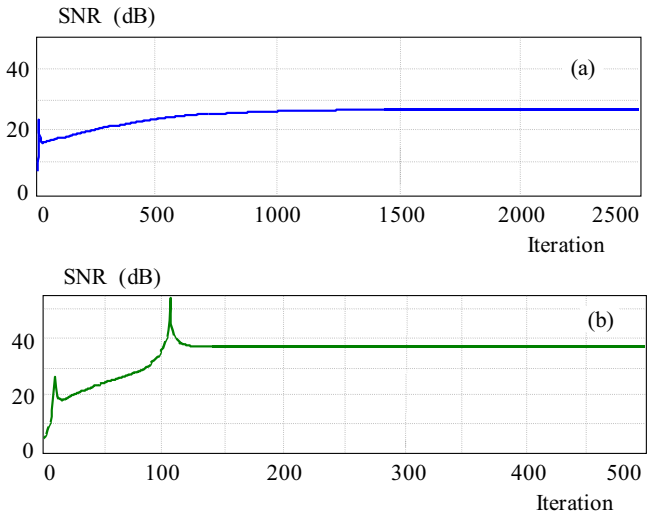

Fig. 6. Output SNR (in Db) versus iteration: (a) – case 1, (b) – case 2

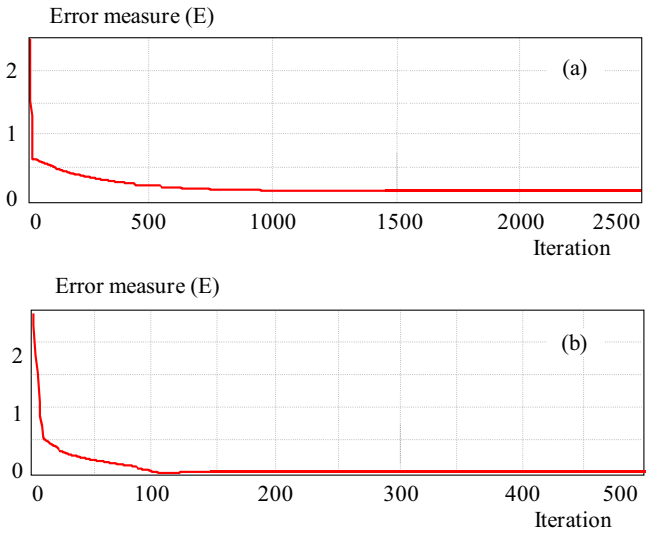

Fig. 7. The error measure versus the number of iterations: (a) – MMSE estimator, (b) – GM estimator

Next, we applied the separating algorithm using GM estimator (case 2) to the above-mentioned signals which results have been displayed in Figs. 5 and 6b.

Also the algorithm performance is measured by the following error measure [17]

$$E = \sum_{i=1}^{d}\Big(\sum_{j=1}^{d}\frac{|p_{ij}|}{\max_k |p_{ik}|} - 1\Big) + \sum_{j=1}^{d}\Big(\sum_{i=1}^{d}\frac{|p_{ij}|}{\max_k |p_{kj}|} - 1\Big) \tag{27}$$

where p_{ij} are the elements of the matrix $P = BA$. The results are presented in Fig. 7.

From the figures it is obvious that the separating algorithm by GM estimator shows a better performance in speech-music separation which causes more SNR's, in addition the separating algorithm by GM estimator is converged faster and eventually the processing time of the algorithm would be fewer.

The comparing of the processing time of the separating algorithm in MATLAB shows, the run time for 100 iterations in GM and MMSE method are approximately 1.73 and 1.06 seconds respectively.

6 CONCLUSION

In this paper, the method of GM estimator has been used to score function estimation in blind source separation. The experimental results indicate the separating algorithm by GM estimator has a better performance in speech-music separation compare to MMSE estimator in addition the separating algorithm can be converged faster and eventually the processing time of algorithm would be less.

Acknowledgment

This paper is the results of a research project that was supported by Telecommunication Research Center with grant number 500/8478. Hereby, we appreciate the financial support of the Telecommunication Research Center.

References

[1] PANAGIOTAKIS, C.—TZIRITAS, G.: Speech/Music Discriminator based on RMS and Zero-Crossing, IEEE Transactions on Multimedia **7** No. 2 (2005), 155–166.

[2] PINQUIER, J.—SÉNAC, C.—ORBRECHT, R.: Speech and Music Classification in Audio Documents, In Proceedings of ICASSP'2002,, vol. 4, USA, May 2002.

[3] SCHEIRER, E.—SLANEY, M.: Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator, In Proceedings of ICASSP'97, Munich, Germany,, Apr 1997, pp. 1331–1334.

[4] CAREY, M. J.—PARRIS, E. S.—LLOYD-THOMAS, H.: A Comparison of Features for Speech- Music Discrimination, In Proceedings of ICASSP'99, vol. 1, USA, March 1999, pp.149-152.

[5] JUTTEN, C.—HERAULT, J.: Blind Separation of Sources I. an Adaptive Algorithm based on Neuromimetric Architecture, Signal Processing **24** No. 1 (1991), 1–10.

[6] COMON, P.: Independent Component Analysis, a New Concept?, Signal Processing **36** No. 3 (1994), 287–314.

[7] CARDOSO, J. F.: Blind Signal Separation: Statistical Principles, Proceedings of the IEEE **9** No. 10 (1998), 2009–2025.

[8] HYVÄRINEN, A.—OJA, E.: Independent Component Analysis: Algorithms and Applications, Neural Networks **13** No. 4 (2000), 411–430.

[9] CHOI, S.—CICHOCKI, A.—PARK, H.-M.—LEE, S.-Y.: Blind Source Separation and Independent Component Analysis: a Review, Neural Information Processing – Letters and Reviews **6** No. 1 (2005), 1–57.

[10] PEDERSEN, M. S.—LARSEN, J.—KJEMS, U.—PARRA, L. C.: A Survey of Convolutive Blind Source Separation Methods, Springer Handbook on Speech Processing and Speech Communication, Springer, 2007, pp. 1–34.

[11] CARDOSO, J.-F.: Source Separation using Higher Order Moments, Proceeding ICASSP, 1989, pp. 2109–2112.

[12] MANSOUR, A.—JUTTEN, C.: Fourth Order Criteria for Blind Separation of Sources, IEEE Trans. on Signal Processing **43** No. 8 (1995), 2022–2025.

[13] BABAIE-ZADE, M.—JUTTEN, C.—NAYEBI, K.: Differential of Mutual Information Function, IEEE Signal Processing Letters **11** No. 1 (2004), 48–51.

[14] SAMADI, S.—BABAIE-ZADE, M.—JUTTEN, C.—NAYEBI, K.: Blind Source Separation by Adaptive Estimation of Score Function Difference, Proceedings of ICA 2004, Granada, Spain, 2004, pp. 9–17.

[15] VLASSIS, N.—MOTOMURA, Y.: Efficient Source Adaptivity in Independent Component Analysis, IEEE Transaction on Neural Networks **12** No. 3 (2001), 559–566.

[16] BABAIE-ZADE, M.—JUTTEN, C.: A General Approach for Mutual Information Minimization and its Application to Blind Source Separation, Signal Processing **85** No. 5 (2005), 975–995.

[17] AMARI, S.—CICHOCKI, A.—YANG, H. H.: A New Learning Algorithm for Blind Signal Separation, Advances in Neural Information Processing Systems (1996,), 757–763, The MIT Press.

Received 28 April 2010

Masoud R. Aghabozorgi Sahaf was born in Yazd, Iran. He received his BSc, MSc and PhD in Electrical Engineering from Isfahan University of Technology in 1993, 1996 and 2002, respectively. He is currently an assistant professor at the Electrical and Computer Engineering Department, Yazd University (Iran), where he is dean of department from 2006. His research interests are in the areas of signal processing especially blind source separation, image processing, and time-frequency analysis and information theory. Dr. Aghabozorgi is a member of the IEEE, the IEEE Signal Processing Society, and a Senior Member of International Association of Computer Science and Information Technology (IACSIT).