New blind estimation method of evoked potentials based on minimum dispersion criterion and fractional lower order statistics

Daifeng Zha

College of Electronic Engineering, Jiujiang University, Jiujiang 332005, China. Correspondence should be addressed to Daifeng Zha (zhadaifeng@jju.edu.cn), Tel.: +86-792-8334956.

ABSTRACT

Evoked potentials (EPs) have been widely used to quantify neurological system properties. Traditional EP analysis methods are developed under the condition that the background noises in EP are Gaussian distributed. Alpha stable distribution, a generalization of Gaussian, is better for modeling impulsive noises than Gaussian distribution in biomedical signal processing. Conventional blind separation and estimation method of evoked potentials is based on second order statistics or high order Statistics. Conventional blind separation and estimation method of evoked potentials is based on second order statistics (SOS). In this paper, we propose a new algorithm based on minimum dispersion criterion and fractional lower order statistics. The simulation experiments show that the proposed new algorithm is more robust than the conventional algorithm.

Keywords: Evoked potentials (EPs), Alpha stable distribution, Blind source separation, Minimum dispersion (MD), Fractional lower order statistics (FLOS)

1. INTRODUCTION

The brain evoked potentials (EPs) are electrical responses of the central nervous system to sensory stimuli applied in a controlled manner. The EPs have a number of clinical applications including critical care, operating room monitoring and the diagnosis of a variety of neurological disorders [1, 2]. The analysis of EP characteristics is of special interest in many clinical applications, such as the diagnosis of possible brain injury and disorders in the CNS [11, 12]. Thus, the goal in the analysis of EPs is currently the estimation from the several potentials, or even from a single potential. In recent years, signal processing techniques including adaptive filtering, three-order correlation, and singular value decomposition (SVD) have been used in fast estimation of EPs. Independent component analysis (ICA) appeared as a promising technique in signal processing. Its main applications re in feature extraction, blind source separation, biomedical signal processing. ICA is based on the following principles. Assume that the original (or source) signals have been linearly mixed, and that these mixed signals are available. Conventional ICA is optimal in approximating the input data in the mean-square error sense, describing some second order characteristics of the data. Nonlinear ICA [3] method related to higher order statistical techniques is a useful extension of standard ICA. The data are represented in an orthogonal basis determined merely by the second-order statistics (covariance) of the input data [4]. Recent studies [5, 6] show that alpha stable distributions is better for modeling impulsive noise, including underwater acoustic, low-frequency atmospheric, and impulsive EEG,ECG, than Gaussian distribution in signal processing. In general, EP signals are always accompanied by ongoing electroencephalogram (EEG) signals which are considered noises in EP analysis. Often the EEG signals are assumed to be Gaussian distributed white noise for mathematical convenience. However, the EEG signals are found to be non-Gaussian in other studies (e.g., [9, 10]). Consequently, EP analysis algorithms developed under the Gaussian EEG assumption may fail or may not perform optimally. Developing EP analysis algorithms without the Gaussian distribution assumption for the background noise thus becomes a key to ensuring the reliability of the analysis results. There are two kinds of noises in the EP signals obtained. The first one is the background EEG noise found in all EP recordings. The second one is the noise introduced by the impact acceleration experiment. An analysis shows that the alpha stable model fits the noises found in the impact acceleration experiment under study better than the Gaussian model [8].

The kind of alpha stable distribution process has no its second order or higher order statistics. It has no close form probability density function so that we can only describe it by its characteristic function:

$$\varphi(t) = \exp\left\{ j\mu t - \gamma |t|^{\alpha} \left[1 + j\beta \,\mathrm{sgn}(t)\varpi(t,\alpha) \right] \right\} \quad (1)$$

Where $\varpi(t,\alpha) = \tan\dfrac{\alpha\pi}{2}(if\,\alpha \neq 1)$ or $\dfrac{2}{\pi}\log|t|(if\,\alpha = 1)$,
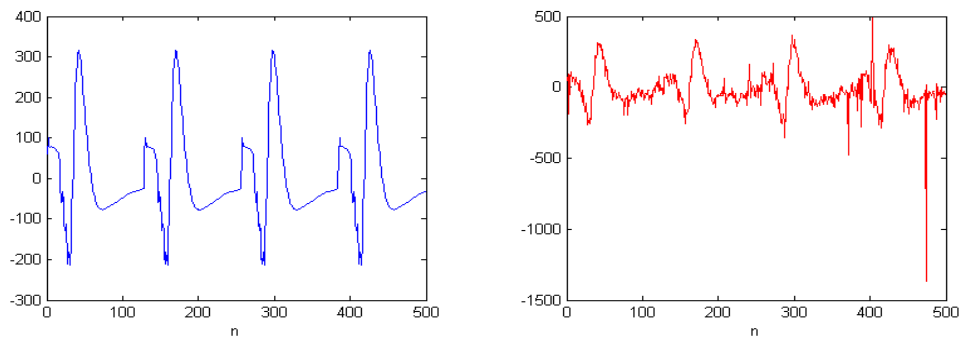
Figure 1. Evoked potentials (Left: without noises; Right: with noises).

$-\infty < \xi < \infty$, $\delta > 0, 0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$. The characteristic exponent α determines the shape of the distribution. Especially, if $\alpha = 2$, it is a Gaussian distribution. The dispersion δ plays a role analogous to the variance of the second order process. β is the symmetry parameter and μ is the location parameter. The distinct characteristics of lower order stable process are its impulsive waveform and the thick tail in its distribution function. Due to the thick tails, lower order stable processes do not have finite second or higher-order moments. This feature may lead all second order moment based algorithms to fail or to function sub-optimally. The typical Evoked potentials are shown in **Figure 1**.

2. DATA MODEL

In the following, we present the basic data model used in both PCA and the source separation problem plotted in **Figure 2**, and discuss the necessary assumptions. We assume that P signals $s_i(n)$, $i=1,2,\ldots P$ are non-coherent, statistically independent. The noiseless linear ICA model with instantaneous mixing may be described by the equation

$$x(n) = As(n) = \sum_{i=1}^{P} a_i s_i(n) \qquad (2)$$

where $x(n) = [x_1(n), x_2(n), \ldots x_M(n)]^T, n = 1, 2, \ldots N, M \geq P$ (T denoting the transpose) are observed signals, $\mathbf{S(n)} = [s_1(n), s_2(n), \ldots, s_P(n)]^T$ are the source signals containing alpha stable distribution signals or noises which are supposed to be stationary and independent, and A is an unknown mixing matrix. Our goal is to estimate S from X, with appropriate assumptions on the statistical properties of the source distributions. The solution is

$$Y(n) = WZ(n) \qquad (3)$$

where W is called the de-mixing matrix, $Z(n)$ is the whitening vector. The general ICA problem requires A to be an $N \times P$ matrix of full column rank, with $M \geq P$ (i.e., there are at least as many mixtures as the number of independent sources). In this paper, we assume an equal number of sources and sensors to make calculation simple. We can write the signal model in matrix form

as $X = AS$. Here X is observation data matrix, S is source signals data matrix, mixture matrix A is unknown.

3. WHITENING BY NORMALIZED COVARIANCE MATRIX

Generally, it is impossible to separate the possible noise in the input data from the source signals. In practice, noise smears the results in all the separation algorithms. If the amount of noise is considerable, the separation results are often fairly poor. Some of the noise can usually be filtered out using standard PCA if the number of mixtures is larger than the number of sources [13].

We introduce here a two-step separation method that achieves the BSS through minimization of a dispersion criterion. The first step is a whitening procedure that orthogonalizes the mixture matrix. Here we search for a matrix B which transforms mixing matrix A into a unitary matrix. Classically, for a finite variance signal, the whitening matrix is computed as the inverse square root of the signal covariance matrix. In our case, impulsive EEG noises have infinite variances. However, we can take advantage from the normalized covariance matrix.

Theorem 1 [7]：Let $X = [x(1), x(2), \ldots, x(N)]$ be a stable process vectors data matrix, then normalized covariance matrix of X

$$\Gamma_x = \frac{X \; X^T}{N \cdot Trace(X \; X^T / N)} \qquad (4)$$

Converges asymptotically to the finite matrix when $N \to \infty$, i.e.,

$$\lim_{N \to \infty} \Gamma_x = AD A^T, D = diag(d_1, d_2, \ldots, d_M)$$

where $d_i = \lim_{N \to \infty} \Delta_i \Big/ \sum_{j=1}^{P} \Delta_j \|a_j\|^2$, a_j is column of A, $\Delta_i = \sum_{n=1}^{N} x_i^2(n)/N$.

Theorem 2：We have eigen-decomposition of Γ_x as $\Gamma_x = U\Omega^2 U^T$ and we can obtain whitening matrix $B = \Omega^1 U^T$, then $Z = BX$ is orthogonal.
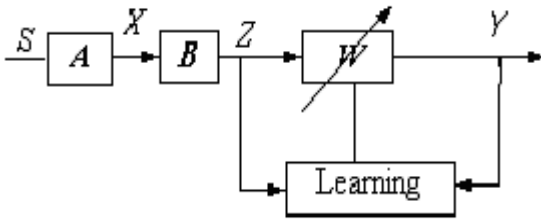
Figure 2. Data and system model.

Proof：

$$ZZ^T = BXX^T B^T = B\Gamma_x N \cdot Trace(XX^T/N)B^T$$
$$= N \cdot Trace(XX^T/N) \cdot B \cdot U\Omega^2 U^T \cdot B^T$$
$$= N \cdot Trace(XX^T/N) \cdot \Omega^{-1}U^T \cdot U\Omega^2 U^T \cdot U \cdot (\Omega^{-1})^T$$
$$= N \cdot Trace(XX^T/N) \cdot \Omega^{-1} \cdot I \cdot \Omega^2 \cdot I \cdot (\Omega^{-1})^T$$
$$= N \cdot Trace(XX^T/N) \cdot I$$

So we can write

$B\Gamma_x B^T = BAD A^T B^T = (BAD^{1/2})(BAD^{1/2})^T = I$ and thus the whitening of $x(n)$ is $z(n) = Bx(n)$.

4. DEMIXING ALGORITHM

The core part and most difficult task in BSS is learning of the separating matrix **W**. During the last years, many neural blind separation algorithms have been proposed. In the following, we discuss and propose separation algorithms which are suitable for alpha stable noise environments in PCA-type networks.

Let us consider i th output weight vector $W_i, i = 1,2...P$, standard PCA is based on second order statistics and maximize the output variances $E\{|y_i(n)|^2\} = E\{|\mathbf{W}_i^T\mathbf{Z}(n)|^2\}$ subject to orthogonal constraints $\mathbf{W}_i\mathbf{W}_i^T = \mathbf{I}_P$. As lower order alpha stable distribution noise has no second order moment, we must select appropriate optimal criterion. FLOS and related other statistics are clearly defined in [5, 6]. So we must use fractional lower order statistics (FLOS) [5, 6], that is to say, the PCA problem corresponding to p-order moment maximization is solution to optimization problem. For each $W_i, i = 1,2...P$

$$W_i^{opt} = \arg\max_{W_i} E\{\frac{1}{p}|Z^T(n)W_i|^p\} \quad (W_iW_i^T = I_P) \quad (5)$$

Let objective function be

$$J(W_i) = E\{\frac{1}{p}|Z^T(n)W_i|^p\} + \frac{1}{2}\lambda_{ii}(W_i^TW_i - 1) + \frac{1}{2}\sum_{j=1,j\neq i}^{P}\lambda_{ij}W_i^TW_j$$

Here the Lagrange multiplier is λ_{ij}, imposed on the orthogonal constraints. For each neuron, $\mathbf{W}_i$ is orthogonal to the weight vector $\mathbf{W}_j, j \neq i$.

The estimated gradient of $J(\mathbf{W}_i)$ with respect to $\mathbf{W}_i$ is

$$\hat{\nabla}J(W_i) = E\{Z(n)|Z^T(n)W_i|^{p-2} conj(Z^T(n)W_i)\} + \sum_{j=1}^{P}\lambda_{ij}W_j \quad (6)$$

At the optimum, the gradients must vanish for $i = 1,2,...P$, and $W_i^TW_j = \delta_{ij}$. These can be taken into ac-

count by multiplying (6) by $\mathbf{W}_j^T$ from left. We can obtain $\lambda_{ij} = -\mathbf{W}_j^T E\{Z(n)|Z^T(n)W_i|^{p-2} conj(Z^T(n)W_i)\}$. Inserting these into (6), we can get

$$\hat{\nabla}J(W_i) = [I - \sum_{j=1}^{P}W_jW_j^T]E\{Z(n)|Z^T(n)W_i|^{p-2} conj(Z^T(n)W_i)\}$$

$$(7)$$

A practical gradient algorithm for optimization problem (5) is now obtained by inserting (7) into

$$\mathbf{W}_i(n+1) = \mathbf{W}_i(n) - \mu(n)\hat{\nabla}J(\mathbf{W}_i(n)),$$

where $\mu(n)$ is the gain parameter. The final algorithm is thus

$$W_i(n+1) = W_i(n) - \mu(n)[I - \sum_{j=1}^{P}W_j(n)W_j^T(n)]Z(n)|$$
$$Z^T(n)W_i(n)|^{p-2} conj(Z^T(n)W_i(n)) \quad (8)$$

As $y_i(n) = \mathbf{Z}^T(n)\mathbf{W}_i(n)$, (8) can be written as follow

$$W_i(n+1) = W_i(n) - \mu(n)|y_i(n)|^{p-2} conj(y_i(n))$$
$$[Z(n) - \sum_{j=1}^{P}y_j(n)W_j(n)] \quad (9)$$

Let $g(t) = |t|^{p-2} conj(t)$, then $g(t)$ is appropriate PCA network nonlinear transform function for lower order alpha stable distribution impulse noises.

Considering that during the iteration error item of gradient $I - \sum_{j=1}^{P}\mathbf{W}_j(n)\mathbf{W}_j^T(n)$ might be zero instantaneously, we modify (9) in order to improve robustness of algorithm as

$$\mathbf{W}_i(n+1) = \mathbf{W}_i(n) - \mu(n)g(y_i(n))[\mathbf{Z}(n) - \sum_{j=1}^{P}g(y_j(n))\mathbf{W}_j(n)] \quad (10)$$

Thus $\mathbf{W}_1, \mathbf{W}_2,...,\mathbf{W}_P$ can be obtained.

Let $Y(n) = [y_1(n), y_2(n), ..., y_P(n)]^T$, $W = [W_1, W_2, ..., W_P]$. For whole network, solution to W and optimization problem is

$$W_{opt} = \arg\max_{W} \sum_{i=1}^{P} E\{\frac{1}{p}|Z^T(n)W_i|^p\} \quad (11)$$

According above derivation, by using $g(t) = |t|^{p-2} conj(t)$, the algorithm for learning W is thus

$$W(n+1) = W(n) - \mu(n)[Z(n) - W(n)g(Y(n))]g(Y^T(n)) \quad (12)$$

5. Performances Analysis

Different nonlinear function can be applied to different blind signal separation problem. Many popular functions are $g(t) = sign(t)$ and $g(t) = \tanh(t)$ corresponding to the double exponential distribution $\frac{1}{2}\exp(-|x|)$ and the inverse-cosine-hyperbolic distribution $\frac{1}{\pi}\frac{1}{\cosh(x)}$, respectively. For the class of symmetric normal inverse
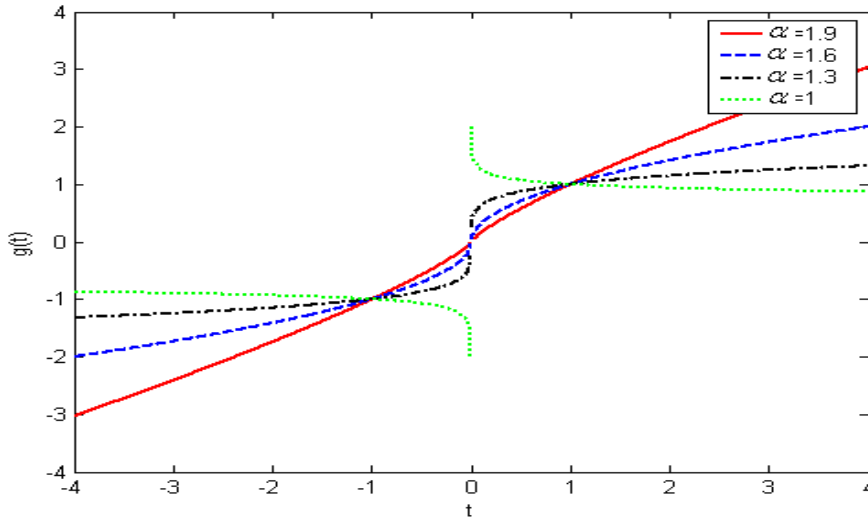
Figure 3. The nonlinear function of alpha stable distribution.

Gaussian (NIG), it is straightforward to obtain according to [14] $g(t) = \dfrac{-\alpha t}{\sqrt{\delta^2 + t^2}} \dfrac{K_2(\alpha\sqrt{\delta^2 + t^2})}{K_1(\alpha\sqrt{\delta^2 + t^2})}$, where $K_1(.)$ and $K_2(.)$ are the modified Bessel function of the second kind with index 1 and 2.

For lower order alpha stable distribution noise has no second order or higher order moment, we must select appropriate nonlinear function $g(t) = |t|^{p-2} conj(t)$ ($p < \alpha$). If t is real data, then $g(t) = |t|^{p-1} sign(t)$. If $p=1$, then $g(t) = sign(t)$. **Figure 3** shows the nonlinear function of alpha stable distribution for different α.

We start from the learning rule (12), and we assume that there exists a square separating matrix $\mathbf{H}^T$ such that $\mathbf{U}(n) = \mathbf{H}^T \mathbf{Z}(n)$. The separating matrix $\mathbf{H}^T$ must be orthogonal. To make the analysis easier, we multiply both sides of the learning rule (12) by $\mathbf{H}^T$. We obtain

$$H^T W(n+1) = H^T W(n) + \mu(n)[H^T Z(n) \\ - H^T W(n) g(W(n)^T Z(n))] \lg(Z(n)^T W(n)) \tag{13}$$

For the sake of $\mathbf{H}\mathbf{H}^T = \mathbf{I}_P$, we can get

$$H^T W(n+1) = H^T W(n) + \mu(n)[H^T Z(n) - \\ H^T W(n) g(W(n)^T H H^T Z(n))] g(Z(n)^T H H^T W(n)) \tag{14}$$

Define $\mathbf{Q}(n) = \mathbf{H}^T\mathbf{W}(n)$, $\mathbf{W}(n) = (\mathbf{H}^T)^{-1}\mathbf{Q}(n)$, (14) is written as

$$Q(n+1) = Q(n) + \mu(n)[U(n) - \\ Q(n) g(Q(n)^T U(n))] g(Q(n) U(n)^T) \tag{15}$$

Geometrically the transformation multiplying by the orthogonal matrix $\mathbf{H}^T$ simply means a rotation to a new set of coordinates such that the elements of the input vector expressed in these coordinates are statistically independent.

Analogous differential equation of (15) is obtained as matrix form:

$$dQ/dt = E\{Ug(U^T Q)\} - E\{g(Q^T U)g(U^T Q)\} \tag{16}$$

According to [15], we can easily prove that (16) has stable solution. For the sake of $\mathbf{Q} = \mathbf{H}^T\mathbf{W}$, thus $\mathbf{W} = (\mathbf{H}^T)^{-1}\mathbf{Q}$ is asymptotic stable solution of (12). **Figure 4** shows the stability and convergence of algorithm based on SOS and FLOS. From **Figure 4**, we know the algorithm based on FLOS has better stability and convergence than the algorithm based on SOS.

6. EXPERIMENTAL RESULTS

From Section I we know that the noise for EP could be a lower order stable process. Through computer simulations, we will demonstrate the effectiveness of the proposed algorithm under alpha stable noise conditions. We use correlation coefficient as follows to evaluate the performances of the proposed algorithms:

$$\tau(s_i, y_j) = \left| \sum_{n=1}^{N} s_i(n) y_j(n) \right| \Big/ \sqrt{\sum_{n=1}^{N} s_i^2(n) \sum_{n=1}^{N} y_j^2(n)} \tag{17}$$

Experiment 1

Two independent sources are linearly mixed. One is the periodical noise free EP signal, and the period is 128 points, the sampling frequency is 1000Hz. The other is an alpha stable non-Gaussian noise with $\alpha = 1.7$. Two algorithms are used in the experiment, including: (1) SOS with nonlinear function $g(t) = \tanh(t)$;(2) FLOS with $g(t) = |t|^{p-2} conj(t)$, respectively. **Figure 4** shows the stability and convergence of algorithm based on SOS and FLOS. We know the algorithm based on FLOS has better stability and convergence than the algorithm based on SOS.

We can get signals waveforms in time domain shown in **Figure 5**, where (a) and (b) are source signals, (c) and (d) are separated signals based on SOS, (e) and (f) are separated signals based on FLOS. For FLOS algorithm,
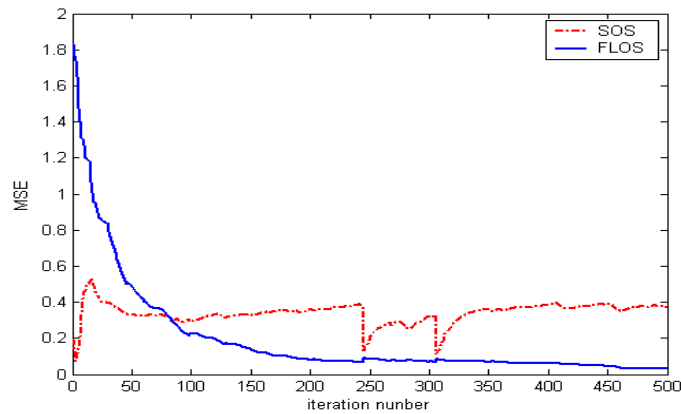
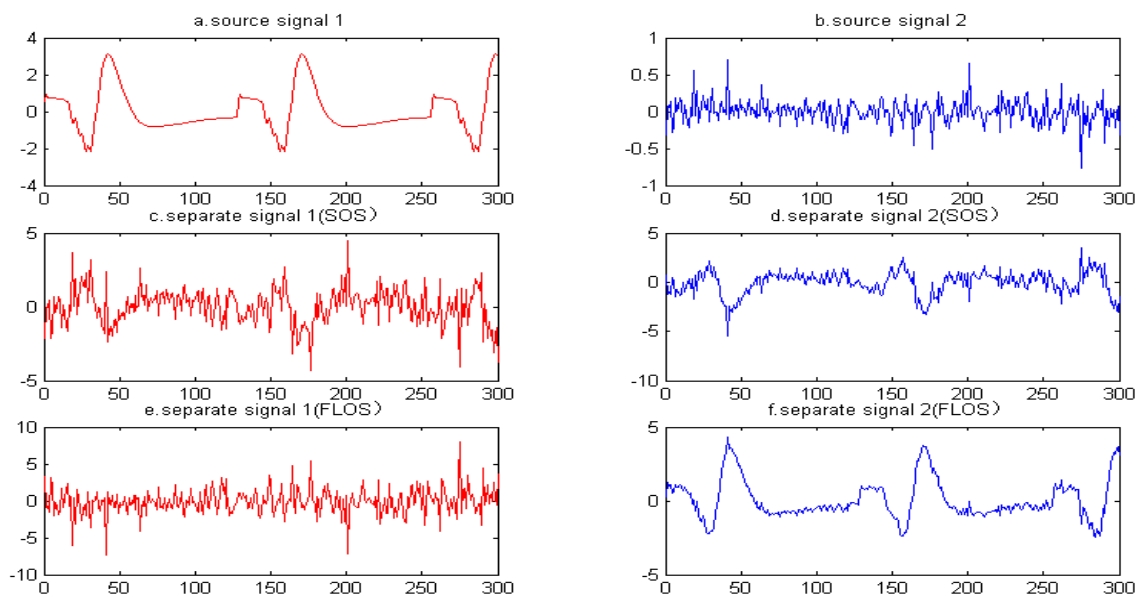**Figure 4.** The stability and convergence.

Figure 5. Separating results: (a)-(b) are the source signals. (c)-(d) are the separated signals with SOS. (e)-(f) are the separated signals with FLOS.
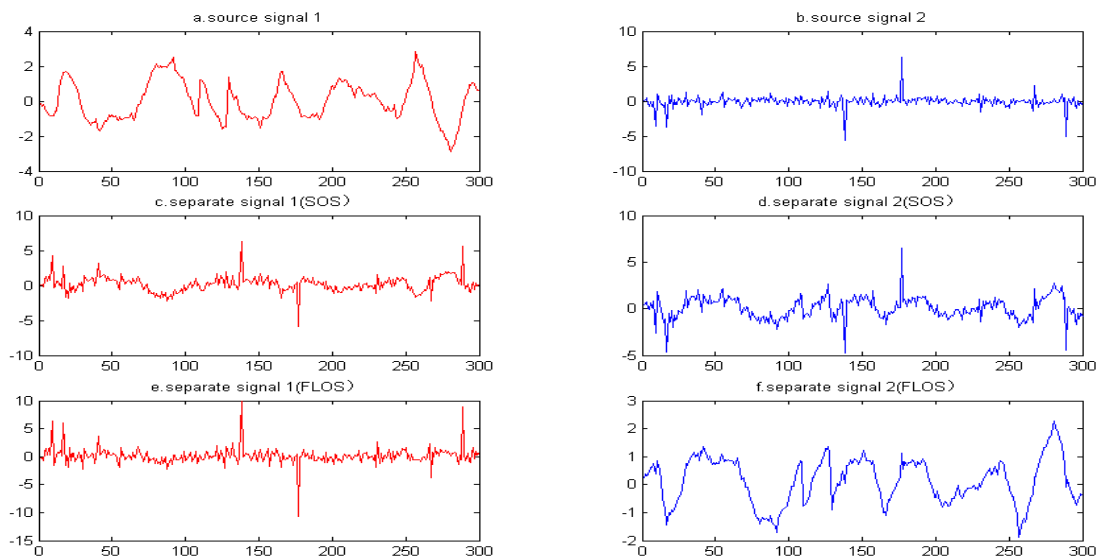

Figure 6. Separating results: (a)-(b) are the source signals. (c)-(d) are the separated signals with SOS. (e)-(f) are the separated signals with FLOS.

Table 1. Comparison between the two algorithms.

Iteration times	Correlation coefficient (FLOS)		Correlation coefficient (SOS)	
	EP	noise	EP	noise
50	0.1244	0.1044	0.0044	0.0004
100	-0.3450	-0.3050	-0.0050	-0.0063
150	0.4378	0.4378	0.1378	0.1072
200	0.6766	0.7706	0.1716	0.1212
250	-0.9291	-0.9091	-0.1711	-0.1451
300	-0.9287	-0.9107	-0.3937	-0.2231
350	-0.9293	-0.9113	-0.4993	-0.2923
400	0.9295	0.9195	0.3945	0.3045
450	0.9299	0.9292	0.2935	0.1935
500	-0.9501	-0.9593	-0.2804	-0.1904
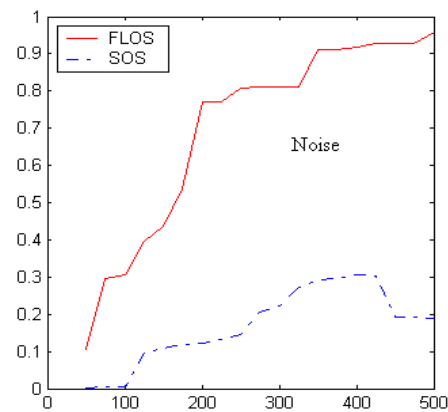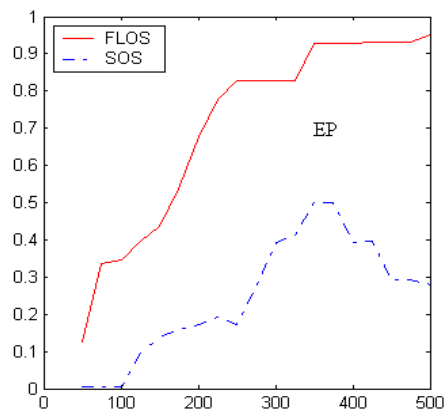
Figure 7. The correlation coefficients of EP and noise.

the correlation coefficient between the separated and source EP signals is 0.9213, and the correlation coefficient between the separated and source alpha stable non-Gaussian noises is –0.9098.

Experiment 2

We repeat simulations when GSNR is 20dB. Two independent sources are linearly mixed. One is the periodical noise free the brain evoked potential (EP) signal, and the period is 128 points, the sampling frequency is 1000Hz. The other is an alpha stable non-Gaussian noise with $\alpha = 1.7$. Two algorithms are used in the experiment, including: (1) SOS with nonlinear function $g(t) = \tanh(t)$;

(2) FLOS with $g(t) = | t |^{p-2} conj(t)$, respectively. We can get signals in time domain shown in **Figure 6**, where (a) and (b) are source signals, (c) and (d) are separated signals based on SOS,(e) and (f) are separated signals based on FLOS. For FLOS algorithm, the correlation coefficient between the separated and source EP signals is–0.9213, and the correlation coefficient between the separated and source alpha stable non-Gaussian noises is –0.9098.

Experiment 3

Separate the mixed signals again with the new FLOS algorithm and conventional SOS algorithm, respectively. And the results of 10 independent experiments are shown

in **Table 1**. The correlation coefficients of EP and of the noise are calculated at some iteration times and plotted in **Figure 7**. From **Table 1**, we get that the performance of the new algorithm is better than the Conventional algorithm.

7. CONCLUSION

Alpha stable distributions, is better for modeling impulsive noise than Gaussian distribution in biomedical signal processing. Conventional blind separation and estimation method of evoked potentials is based on second order statistics. In this paper, we modify conventional algorithms and analyze the stability and convergence performance s of the new algorithm. From above simulation, we can easily obtain the following conclusions: the proposed class of algorithm of estimation of evoked potentials based on FLOS is more robust than conventional algorithms based on SOS so that its separation capability is greatly improved under both Gaussian and fractional lower order stable distribution noise environments.

ACKNOLEDGEMENT

REFERENCES

[1] R. R. Gharieb, A. Cichocki. (2001) Noise reduction in brain evoked potentials based on third-order correlations. *IEEE Transactions on Biomedical Engineering*, **48**(5): 501-512.

[2] C. E. Davila, R. Srebro, and I. A. Ghaleb. (1998) Optimal detection of visual evoked potential. *IEEE Transaction on Biomedical Engineering*, **45**(6): 800–803.

[3] Yanwu Zhang,Yuanliang Ma (1997) CGHA for principal component extraction in the complex domain. *IEEE. Trans. on Neural Network*,vol **8**,No.5.

[4] Mutihac, R. Van Hulle, M.M. (2003) PCA and ICA neural implementations for source separation - a comparative study. *Proceedings of the International Joint Conference on Neural Networks*, Volume: **1** , 20-24.

[5] C. L. Nikias and M. shao. (1995) *Signal Processing with Alpha-Stable Distributions and Applications.* New York: John Wiley & Sons Inc.

[6] M. shao and C. L. Nikias. (1993) Signal Processing with fractional lower order moments: stable processes and their applications. *Proceedings of IEEE*, Vol.**81**, No.7, 986-1010.

[7] G. Samorodnitsky, M. S. Taqqu. (1994) *Stable Non-Gaussian Random Process: Stochastic Models with Infinite Variance*. New York: Chapman and Hall.

[8]Xuan Kong,Tianshuang Qiu. (1999) Adaptive Estimation of Latency Change in Evoked Potentials by Direct Least Mean p-Norm Time-Delay Estimation. *IEEE Transactions on Biomedical Engineering*, vol. **46**, No. 8, August.

[9] N. Hazarika, A. C. Tsoi, and A. A. Sergejew. (1997) Nonlinear considerations in EEG signal classification. *IEEE Trans. Signal Processing*, vol. **45**, pp. 829–936.

[10] X. Ma and C. L. Nikias. (1996) Joint estimation of time delay and frequency delay in impulsive noise using fractional lower-order statistics. *IEEE Trans. Signal Processing*, vol. **44**, pp. 2669–2687, Nov.

[11] X. Kong and N. V. Thakor. (1996) Adaptive estimation of latency changes in evoked potentials. *IEEE Trans. Biomed. Eng.*, vol. **43**, pp. 189–197, Feb.

[12] C. A. Vaz and N. V. Thakor. (1989) Adaptive Fourier estimation of time varying evoked potentials. *IEEE Trans. Biomed. Eng.*, vol. **36**, pp. 448–455, Apr.

[13] Winter, S.; Sawada, H.; Makino, S. (2003) *Geometrical understanding of the PCA subspace method for overdetermined blind source separation.Acoustics, Speech, and Signal Processing*.

[14] Preben Kidmose. (2001) Blind separation of heavy tail signals. *Technical university of denmark ,IMM-Phd, LYNGBY*

[15] Juha Karhumen;Erkki Oja;Liuyue Wang;Ricardo Vigario;Jyrki Joutsensalo. (1997) A class of neural networks for independent component analysis. *IEEE. Trans. on Neural Network*,vol **8**,No.3.