Processing Human Colonic Pressure Signals by Using Overdetermined ICA

She-ping TIAN(田社平), Cheng PAN(潘　城), Guo-zheng YAN(颜国正)
(*Dept. of Information Measurement Technology and Instruments, Shanghai Jiaotong University, Shanghai* 200240, *China*)

Abstract – Independent component analysis (ICA) is a widely used method for blind source separation (BSS). The mature ICA model has a restriction that the number of the sources must equal to that of the sensors used to collect data, which is hard to meet in most practical cases. In this paper, an overdetermined ICA method is proposed and successfully used in the analysis of human colonic pressure signals. Using principal component analysis (PCA), the method estimates the number of the sources firstly and reduces the dimensions of the observed signals to the same with that of the sources; and then, Fast-ICA is used to estimate all the sources. From 26 groups of colonic pressure recordings, several colonic motor patterns are extracted, which not only prove the effectiveness of this method, but also greatly facilitate further medical researches.

Key words – *medical signal processing*; *overdetermined ICA*; *PCA*; *colonic motor pattern*

Manuscript Number: 1674-8042(2010)04-0401-05

dio: 10.3969/j.issn.1674-8042.2010.04.22

1 Introduction

Blind source separation (BSS) is a technique for estimating original source signals using only sensor observations that are mixtures of the original signals. Recently BSS has become an increasingly important research area due to its rapidly growing applications in various fields such as biomedical data processing (e.g. separation and analysis of EEG, MEG and ECG signals)[1-2], audio and image processing, wireless communications, and others. Independent component analysis (ICA) is a well-known technique for solving such a task. So far, most studies have focused on the complete ICA where there are just as many sensors as the source signals. Furthermore, ideal sensors are usually assumed, which have no additional sensor noise[3]. However, in most practical cases, the number of source signals is unknown and changes over time. Especially for the medical signals, a variety of human physiological activities mix together[4], making it difficult to ensure how many sources there are. This paper proposes a new method to solve the problem with the overdetermined ICA, where there are more sensors than sources and a low SNR is present at the sensors. Using principal component analysis (PCA), the algorithm estimates the number of sources firstly and reduces the dimensions of the observed signals to be equal to that of the sources; and then, Fast-ICA is used to estimate all the sources. By applying the algorithm to the human colonic pressure data, desired results are obtained.

2 Overdetermined ICA

2.1 Model and Assumptions

Consider the overdetermined ICA model with n sources $s_1(t), s_2(t), \cdots, s_n(t)$ and m sensors that give mixed signals $x_1(t), x_2(t), \cdots, x_m(t)$ with additional noises $v_j(t)(j = 1,2,\cdots,m)$. The mixing process can be described by

$$x_j(t) = a_{j1}s_1(t) + a_{j2}s_2(t) + \cdots + a_{jn}s_n(t) + v_j(t), j = 1,2,\cdots,m, \quad (1)$$

Transformed into the vector form, equation (1) becomes

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{v}(t). \quad (2)$$

where $\boldsymbol{x}(t) \in \mathbf{R}^m$, $\boldsymbol{s}(t) \in \mathbf{R}^n$ and $\boldsymbol{v}(t) \in \mathbf{R}^m$ are vectors of the observed variables, independent sources and additional noise, respectively. $\boldsymbol{A} = (a_{ij})_{m\times n}$ is the unknown mixing matrix that has full rank. In the overdetermined case, we have $m < n$. Thus rank $(A) = n$.

The overdetermined ICA model is based on a few assumptions as follows, most of which are easily to meet in practice[3].

a) The source signals are mutually statistic independent and have zero mean.

b) All the source signals but possibly one are non-Gaussian. All the noises are additive white Gaussian noises and are mutually independent.

c) The source signals and the noises are mutually in-

※ **Received**: 2010-09-15
Project supported: This work was supported by National Natural Science Foundation(No. 60875061)
Corresponding author: She-ping TIAN(sptian@sjtu.edu.cn)

dependent.

2.2 Estimation of the number of the source signals

To solve the overdetermined ICA problem, the number of source signals $n = \text{rank}(\boldsymbol{s})$ first is estimated, and then the m observed signals are projected onto an n-dimension signal-plus-noise subspace and an m-n-dimension noise subspace. Discarding the noise subspace, the overdetermined ICA is converted to be a complete ICA problem, in which the number of the sensors is equal to that of the sources and the source signals can be estimated by Fast-ICA algorithm.

Denote the covariance matrix of the observed signals vector by

$$\boldsymbol{C} = \mathrm{E}[xx^{\mathrm{H}}], \tag{3}$$

where $\boldsymbol{C}$ is a nonnegative definite Hermitian matrix. By applying the SVD (singular-value decomposition) on $\boldsymbol{C}$, we obtain

$$\boldsymbol{C}_{\mathrm{m\times n}} = \boldsymbol{U}_{m\times n}\Lambda_{m\times m}\boldsymbol{U}^{\mathrm{H}}_{m\times m}, \tag{4}$$

where $\Lambda = \text{diag}[\lambda_1\lambda_2\cdots\lambda_m]$ is a diagonal matrix with eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m \geqslant 0$. $\boldsymbol{U}$ is a unitary matrix that $\boldsymbol{U}^{-1} = \boldsymbol{U}^{\mathrm{H}}$. And the columns of $\boldsymbol{U}$ are denoted by $u_1, u_2, \cdots, u_m$, each of which is the eigenvector corresponding to an eigenvalue. With Eq.(2) and Eq.(3), we obtain

$$\boldsymbol{C} = \mathrm{E}[(As)(As)^{\mathrm{H}}] + \mathrm{E}[vv^{\mathrm{H}}] + \mathrm{E}[(As)v^{\mathrm{H}}] + E[v(As)^{\mathrm{H}}]. \tag{5}$$

Based on the assumption c), we have $\mathrm{E}[(As)v^{\mathrm{H}}] = \mathrm{E}[v(As)^{\mathrm{H}}] = 0$, then Eq.(5) becomes

$$\boldsymbol{C} = \mathrm{E}[(As)(As)^{\mathrm{H}}] + \mathrm{E}[vv^{\mathrm{H}}]. \tag{6}$$

Let $\boldsymbol{x}' = \boldsymbol{As}$ and $\boldsymbol{C}' = \mathrm{E}[x'x'^{\mathrm{H}}]$, where x represents the mixtures of source signals without noises, and $\boldsymbol{C}'$ is the covariance matrix of x'. Then Eq.(6) could be rewritten as

$$\boldsymbol{C} = \boldsymbol{C}' + \mathrm{E}[vv^{\mathrm{H}}], \tag{7}$$

The literature [5] proves that the number of the source signals $n = \text{rank}(\boldsymbol{s}) = \text{rank}(\boldsymbol{C}')$, and the eigenvalues of $\boldsymbol{C}$ and $\boldsymbol{C}'$ have relations as follows

$$\lambda_i = \begin{cases} \lambda_i' + \delta_i^2, & i = 0,1,\cdots,n, \\ \delta_i^2, & i = n, n+1, \cdots, m. \end{cases} \tag{8}$$

where λ_i' are the eigenvalues of $\boldsymbol{C}'$, and δ_i^2 is the variance of the noise v_i.

When the SNR is high enough, δ_i^2 is relatively much smaller than λ_i', i.e., $\lambda_i' + \delta_i^2 \gg \delta_i^2$. Since λ_i is ordered descendingly and the first n eigenvalues $\lambda_i = \lambda_i' + \delta_i^2 (i = 0,1,\cdots,n)$ are much bigger than the last m-n eigenvalues $\lambda_i = \delta_i^2 (i = n, n+1, \cdots, m)$, the first n eigenvalues are called principal eigenvalues and the last $m-n$ eigenvalues minor eigenvalues. So we can obtain the numbers of source signals n by calculating the number of principal eigenvalues.

Then we could reduce the dimension of observed signals to make it as many as the source signals.

The eigenvectors corresponding to the principal eigenvalues called principal eigenvectors, which span an n dimension signal-plus-noise subspace. And the rest $m-n$ eigenvectors span an $m-n$ dimension noise subspace. Discarding the $m-n$ smaller eigenvalues and their corresponding eigenvector, the unitary matrix $\boldsymbol{U}$ becomes

$$\boldsymbol{U}_{\mathrm{m\times m}} \rightarrow \boldsymbol{U}'_{m\times n}[u_1, u_2, \cdots, u_n], \tag{9}$$

So we can project the observed signals on the n dimension signal-plus-noise subspace by

$$\boldsymbol{p}_n = \boldsymbol{U}'^{\mathrm{H}}_{\mathrm{n\times m}}\boldsymbol{x}_m, \tag{10}$$

The elements of $\boldsymbol{p}$ are principal components of $\boldsymbol{x}$. With Eq.(4) and (10) we obtain the covariance matrix of $\boldsymbol{p}$ as

$$\boldsymbol{C}_{pp} = E[\boldsymbol{pp}^{\mathrm{H}}] = \boldsymbol{U}'^{\mathrm{H}}\mathrm{E}[xx^{\mathrm{H}}]\boldsymbol{U}' = \boldsymbol{U}'^{\mathrm{H}}\boldsymbol{U}\Lambda U^{\mathrm{H}}\boldsymbol{U}' = \Lambda' = \text{diag}[\lambda_1\lambda_2\cdots\lambda_n]. \tag{11}$$

It can be seen that, the principal components are mutually uncorrelated, and ordered descendingly by the power. Furthermore, they are as many as the source signals, thus we only use principal components for the further processing.

2.3 The proposed algorithm

After the overdetermined ICA problem has been converted into complete ICA, we can use Fast-ICA algorithm to solve it easily. The Fast-ICA algorithm is a computationally highly efficient method for performing the estimation of ICA. It is introduced by Hywarinen[6]. Fast-ICA algorithm is based on the fixed-point iteration scheme, and has a very fast convergence (cubic).

ICA separates the source signals by the principle of independence, which could be measured by nongaussianity. A widely used approximation for the nongaussianity is

$$J(\hat{s}_i) \propto [E\{G(\hat{s}_i)\} - E\{G(v)\}]^2. \tag{12}$$

where $J(\hat{s}_i)$ is the nongaussianity of the random variable $\hat{s}_i$, v is a zero-mean, unite-variance Gaussian variable, and G is an nonquadratic function. Selection of a proper G is given by Hywarinen in Ref.[7].

The overall algorithm is given as follow:

i) Apply PCA to the observed vector $\boldsymbol{x}$. Estimate the number of source signals and obtain the principal component vector $\boldsymbol{p}$, which have already reduced the dimension.

ii) Apply whiten transformation to $\boldsymbol{p}$. Get a new vector $\boldsymbol{z}$, the elements of which are unite-variance and mutually uncorrelated.

iii) Choose an initial (e.g. random) weight vector $\boldsymbol{w}_i$.

iv) Let $\boldsymbol{w}_i^+ = \mathrm{E}\{zg(\boldsymbol{w}_i^{\mathrm{H}}z)\} - \mathrm{E}\{g'(\boldsymbol{w}_i^{\mathrm{H}}z)\}\boldsymbol{w}_{\mathrm{i}}$, where is the derivative of function G in Eq.(12)

v) Let $\boldsymbol{w}_i = \dfrac{\boldsymbol{w}_i^+}{\|\boldsymbol{w}_i^+\|}$.

vi) If not converged, go back to iv).

vii) Put the converged $\boldsymbol{w}_i$ into $y_i = \boldsymbol{w}_i^{\mathrm{H}}\boldsymbol{x}$ to obtain one of the source signals.

To estimate all the source signals, we need to run the iii)~vii) n times, and each time choose a different initial $\boldsymbol{w}_i$. To prevent different vectors from converging to the same maxima we have to decorrelate w_i in each iteration. For example, when $j-1$ source signals have been estimated, we could apply Gram-Schmidt-like decorrelation to w_j between iv) and v) in each iteration to estimate the jth source signal.

3 Processing of Colonic Pressure Signal

The Colonic motor function is a high-incidence disease, with the representative clinical manifestation of slow transit constipation (STC) and colonic inertia (CI)[8]. The Colonic pressure signal is the important physiological information to evaluate the colonic motor function[9]. The subjects in the experiment consisted of 26 volunteers, including a group of 14 STC patients and a control group of 12 healthy ones. All subjects gave written informed consent, and the study was approved by the Shanghai Human Ethics Committee, China. The instrument used to collect pressure data is the gastrointestinal function monitor made by CTD-SYNETICS Company from Sweden. The measuring system consisted of the monitor, a computer, a data acquisition card and a manometry catheter. The manometry catheter has 4-channels with a 5 cm intersidehole distance. A water balloon was attached to the end of the catheter in order to help position the sensors by a B-type ultrasonic inspection system. There was a barostat balloon aiming to simulate the filled situation of colon 10 cm away from the water balloon. Each measurement took about 20 minutes, and a group of pressure data was outputted in the computer. Then, the catheter was pulled out for 20 cm and the same process was repeated until the whole colon was tested.

Human colonic motor activiuy is complex and variable[10], showing disorder in most of times. Recentmy, 4~7 colonic motor patterns have been discovered from the ambulatory 24 h colonic manometry in healthy humans, including simultaneous pressure waves, periodic colonic motor activity (PCMA), high amplitude propagated contractions (HAPC), low amplitude propagated contractions (LAPC), etc[4]. Studies on the physiological significances of these patterns have become a hot issue. For example, HAPC has been found playing an important role in the human defecation process[11]. However, our knowledge about these patterns is still limited. Mixing of these patterns and noise caused by some other physiological activities makes it difficult to discover the physiological significances of each patterns[4]. Analysis of the data from all the 26 subjects shows that in most cases, 4 sensors used in this experiment could meet the qualification of overdetermined ICA that the number of sensors is more than source signals. Therefore, we apply the proposed algorithm to the colonic pressure signals and try to extract the waveform of colonic motor patterns independently.

Fig. 1 shows the 10 minutes, 4-channel colonic pressure signals of a subject. From the waveform, we could see that the four channels' signals are almost the same. It means that the pressure is almost the same at the 20 cm long section of colon. This is a typical case of simultaneous pressure waves, which comprised the predominant pattern of motor activity in the colon[10]. However, the simultaneous pressure signal in each channel may still mix with other motor patterns and noises. So we use overdetermined ICA model to analyze the signals.

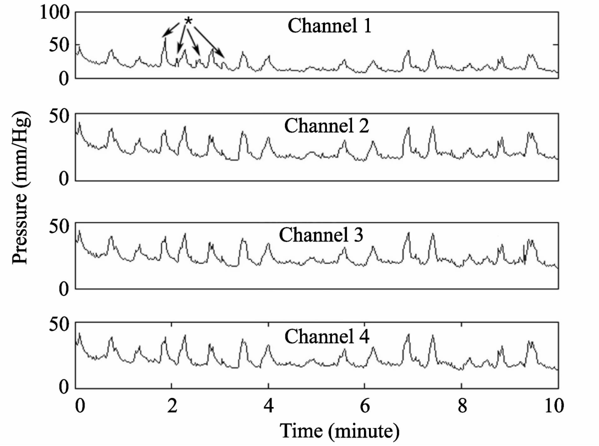

Fig. 1 Four-channel observed signals

First, apply the proposed algorithm to the 4-channel signals. Eigenvalues of the covariance matrix is shown in Fig. 2. When the SNR is high enough, we assume that the eigenvalues which are smaller than 0.5% of the sum of all eigenvalues is minor eigenvalues, and discard them. The remaining eigenvalues are principal eigenvalues, the number of which is equal to that of source signals. In this sample, the first 2 eigenvalues account for 99.5799% of the sum of all the eigenvalues. Thus other eigenvalues are minor eigenvalues and the number of source signals is 2. Reducing the dimension of observed-signal vector to 2 and applying Fast-ICA to it, we obtain 2 independent components (ICs) shown in Fig. 3, which are estimations of the 2 source signals. From Fig. 3 we could see that the 4-channel' simultaneous pressure waves are collected into IC2, which means the 20 cm-long simultaneous pressure waves are originated from the same physical source, and consequently they have the same physiological significance. IC1 shows a group of contractions occurred in the 2nd and the 4th minute. This is the bursts of pressure waves[8,12], another common pattern of human colonic activity. Indicated by the "*" in Fig. 1, it can be seen that IC1 occurs only in channel 1, with lower power compared to the simultaneous pressure waves. This makes it difficult to observe by visual analysis. Using the proposed algorithm, we successfully estimated the number of sources from only the 4-channel observed signals and extracted waveforms of two patterns of colonic activity, which will greatly facilitate further medical researches.

Fig. 4 shows another example of a subject's 4-channel colonic pressure signals, from which we can see that the 4-channels' signal are not the same, but a big step occurs in all the four channels at arround the 1st minute. Apply the

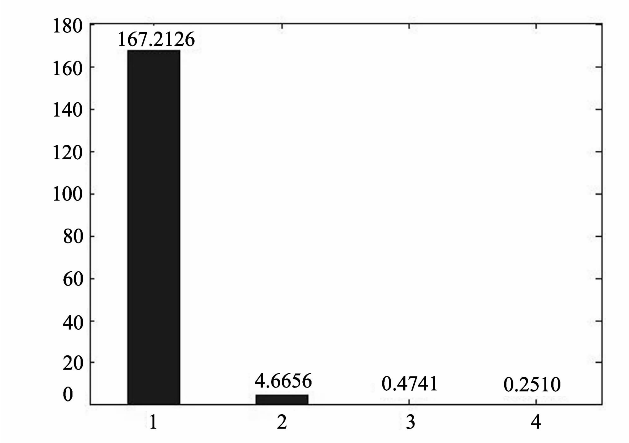

Fig. 2 Eigenvalues of observed signals' covariance matrix

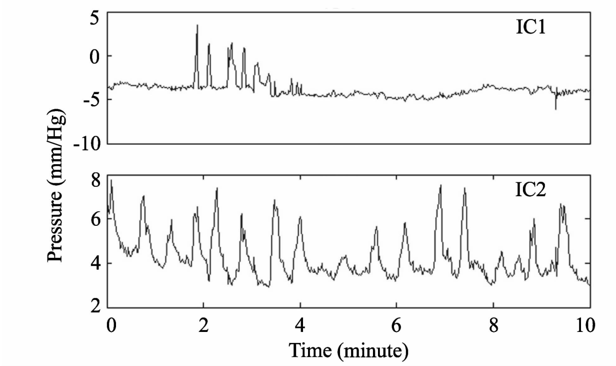

Fig. 3 Two source signals extracted by over-determined ICA

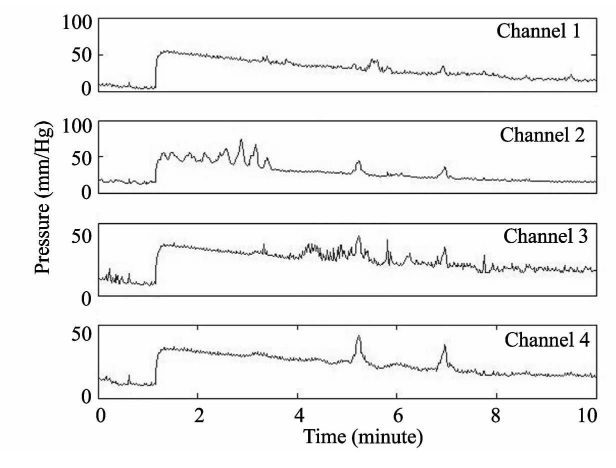

Fig. 4 Four-channel observed signal

proposed algorithm to the 4-channel signals. Eigenvalues of the covariance matrix is shown in Fig. 5. Similarly, assume that the eigenvalues which are smaller than 0.5% of the sum of all eigenvalues is minor eigenvalues. The first 3 eigenvalues account for 99.579 9% of the sum of all the eigenvalues. Thus the number of source signals is 3. The estimations of source signals are shown in Fig. 6. The waveform of IC1 could be found in all the 4-channel observed signals, which means that IC1 is the waveform of the pattern simultaneous pressure waves. IC2 shows a group of contractions between the 2nd and the 4th minute. In Fig. 4, it can be seen that IC2 occurs only in Channel 2. This is a typical example of the pattern-bursts

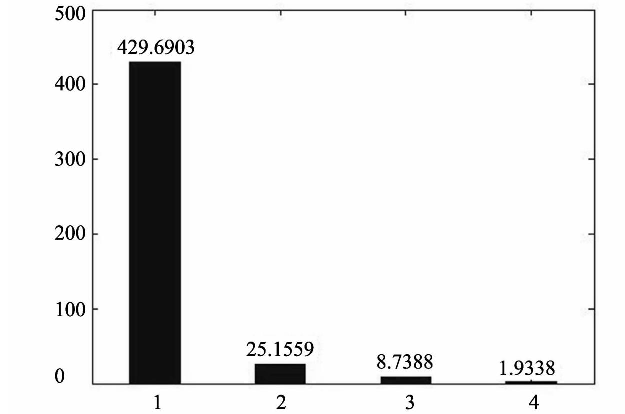

Fig. 5 Eigenvalues of observed signals' covariance matrix

of contractions. We can see that two colonic motor patterns have occured simultaneously in Channel 2 in that 10 min. IC3 is a step signal which also ccurs in all the 4 channels, but IC3 and IC1 are not extracted as one IC, whichmeans the physical source of IC1 and IC3 are not the same. It can be presumed fom the waveform that IC3 is the response of colon to some sudden stimulation in the measurement like movement of the catheter, filling of the balloon, etc. From this example, we can see that the overdetermined ICA extracts ICs that derive from different sources. It also collects the interference factors into a single IC so that the interference to other components is greatly weakened.

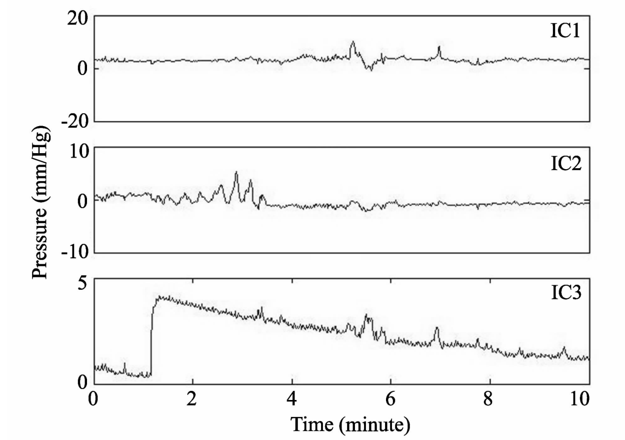

Fig. 6 Three source signals extracted by over-determined ICA

Besides simultaneous pressure waves and bursts of pressure waves mentioned above, signals of other patterns of colonic activity were also extracted by analyzing all the 26 colonic pressure recordings, such as periodic colon motor activity (PCMA)[10], rectal motor complex (RMC)[13] and isolated pressure waves. However, limited to our experimental conditions, the measurement in each segment of the colon lasted only 20 minutes. This made it impossible for us to observe some patterns of colonic motility that occurred only a few times per day.

4 Conclusion

This paper proposed an algorithm to solve the problem of overdetermined ICA and applied it successfully in human colonic pressure signal processing. Using PCA, the method estimated the number of sources firstly and reduced the dimension of observed signals to be the same with that of the sources; and then, Fast-ICA was used to estimate all the sources. Processing the colonic pressure signals with this algorithm, we successfully extracted waveforms of several patterns of colonic activity from the 4-channel observed signals, which could greatly facilitate further medical researches in discovering the physiological significance of these colonic motor patterns. And what's more, the paper provides a new perspective to study the colonic pressure recordings that we could first separate the data into useful components independently, and then analyze them thoroughly.

5 Ankowledgements

The authors would like to thank the reviewers for their efforts to help improve this paper. This work was supported by China National Natural Science Foundation (No. 60875061).

References

[1] A. Neumann, M. Grosse-Wentrup, M. Buss, et al, 2008. The effect of mutual information on independent component analysis in EEG/MEG analysis: A simulation study. *International Journal of Neuroscience*, 118(11): 1534-1546.

[2] D. Mantini, R. Franciotti, G. L. Romani, et al, 2008. Improving MEG source localizations: An automated method for complete artifact removal based on independent component analysis. *NeuroImage*, 40(1): 160-173.

[3] M. Joho, H. Mathis, R. H. Lambert, 2000. Overdetermined Blind Source Separation: Using More Sensors Than Source Signals in a Noisy Mixture. ICA2000, Helsinki, Finland, p. 81-89.

[4] M. Camilleri, A. E. Bharucha, C. Di Lorenzo, et al. 2008. American neurogastroenterology and motility society consensus statement on intraluminal measurement of gastrointestinal and colonic motility in clinical practice. *Neurogastroenterology and Motility*, 20(12): 1269-1282.

[5] D. Liu, Y. H. Song, F. Wang, 2005. Over-determined Blind Signal Separation Based on PCA and Infomax. Proceeding of 2005 Chinese Control and Decision Conference. Shenyang, p. 538-541.

[6] A. Hyvarinen, 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3): 626-634.

[7] A. Hyvarinen, J. Karhunen, E. Oja, 2001. Independent Component Analysis. John Wiley & Sons, Inc. New York.

[8] M. M. Schuster, 2002. Atlas of Gastrointestinal Motility in health and disease 2nd ed. et al Decker Inc..

[9] G. Bassotti, U. Germani, C. Calcara, 1996. Twenty-four hour manometric recording of colonic motility: A technique with both research and clinical identities. *Techn Coloproctol*, 4: 142-144.

[10] S. S. C. Rao, P. Sadeghi, J. Beaty, et al, 2001. Ambulatory 24-h colonic manometry in healthy humans. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 280(4 43-4): G629-G639.

[11] P. G. Dinning, M. M. Szczesniak, I. J. Cook, 2008. Proximal colonic propagating pressure waves sequences and their relationship with movements of content in the proximal human colon. *Neurogastroenterology and Motility*, 20(5):512-520.

[12] G. Bassotti, G. de Roberto, D. Castellani, et al. 2005. Normal aspects of colorectal motility and abnormalities in slow transit constipation. *World Journal of Gastroenterology*, 11(18), 2691-2696.

[13] B. A. Orkin, R. B. Hanson, K. A. Kelly. 1989. The rectal motor complex. *Journal of Gastrointestinal Motility*, 1(1): 5-8.