4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

Independent Component Analysis of Electrogastrogram Data

Masashi OHATA[*1], Takahiro MATSUOMOTO[*2], Akio SHIGEMATSU[*3] and Kiyotoshi MATSUOKA[*4]

[*1]Biologically Integrative Sensors Lab., RIKEN BMC Research Center, Japan
[*2] Department of Anesthesia, Kyushu Koseinenkin Hospital, Japan
[*3] Dep. of Anesthesiology, University of Occupational and Environmental Health, Japan
[*4] Dep. of Brain Science and Engineering, Kyushu Institute of Technology, Japan

Abstract

This paper presents an application of independent component analysis to electrogastrogram (EGG), which is a gastric myoelectrical activity measured by several electrodes attached on the abdomen. The purpose is to remove the interfering signals other than the gastric activity.

Our analysis is done under the assumption that electrical activities of the organs near the electrodes are statistically mutually independent of each other and the EGG data is a convolutive mixture of them. The result shows that the proposed method is able to clearly extract the component originated from the the gastric activity.

I. Introduction

Eelectrocgastrogram (EGG) is the recording of gastric myoelectrical activity by several electrodes attached on the abdomen. A configuration of the electrodes is shown in Fig.1. A crucial problem in analyzing EGG data is that the signals detected by the electrodes do not only contain the signal of interest, namely the component originated from the stomach, but also those from other organs near the stomach. In order to use EGG data for clinical purpose, it is required to remove the interfering signals other than gastric activity itself. In healthy humans the frequency range of the stomach-originated signals is localized around 3 [cycles / min] (5×10^{-2} [Hz]).

A method for extracting the gastric component might be to apply a band-pass filter to EGG data. It is however difficult to do so because the frequency range of gastric signal overlaps with that of the signals originated from other organs and moreover the frequency characteristics of those signals are rarely known.

In such a situation the technique of independent component analysis (ICA) seems to be effective, in which no information about the frequency characteristics of the sources is required. ICA is a statistical technique to extract the set of independent components only from given data. We assume that the signals produced by different organs are statistically independent and EGG data is an observation of their mixture though, strictly speaking, it might not be true.

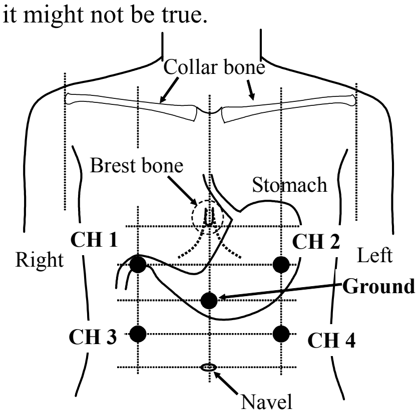


Fig. 1 Location of electrodes

In general a mixing process is classified into either of two types: instantaneous mixture and convolutive mixture. In [7] and [12], the mixing process in EEG is assumed to be instantaneous. In our approach, oppositely, we treat the mixing process as a convolutive one because the gastric myoelectrical wave is governed by gastric peristalsis and hence the signal from the stomach is measured by the electrodes with different delays.

In conventional algorithms for convolutive mixture of sources, each source signal is usually assumed to be independent and identically distributed (iid).[8] However, such approaches are not suitable for EGG data because the component produced by the stomach may have a strong periodicity. The reason of why the assumption of iid is unsuitable for convolutive mixture of periodic source signals is described in [9]. In this paper we use a new algorithm without the assumption that sources are iid. The algorithm was derived by one of the authors, based on two principles called "minimal distortion principle" and "inverse minimal distortion principle." [11]

This paper is organized as follows. In section II we describe some mathematical notations used in this paper. Section III provides a formulation of ICA for the readers unfamiliar with its concept. Section IV, V and VI show the details of derivation of our algorithm used in the analysis of EGG. Section VII shows a result of the application of the algorithm to EGG. Section IIV is devoted to the conclusion.

II. Mathematical notations

In this section we describe mathematical denotations for matrices appearing in the following sections. Below, matrix $\mathbf{X}$ and transfer function matrix $\mathbf{X}(z) = \Sigma_\tau \mathbf{X}_\tau z^{-\tau}$ are $M \times N$ matrices. $\mathbf{X}$ can be complex-valued while coefficients $\mathbf{X}_\tau$ of $\mathbf{X}(z)$ are real-valued.

The conjugate transpose of matrix $\mathbf{X}$ is represented as $\mathbf{X}^H$. The conjugate transpose of $\mathbf{X}(z)$ is defined as $\mathbf{X}^H(z) \triangleq \mathbf{X}^T(z^{-1})$.

Frequency transfer function $\mathbf{X}(e^{j2\pi f})$ associated with $\mathbf{X}(z)$ is denoted by $\tilde{\mathbf{X}}(f)$. If a square matrix $\tilde{\mathbf{X}}(f)$ is nonsingular for every frequency f, $\mathbf{X}(z)$ is said to be nonsingular. If $M \geq N$ ($M \leq N$) and $\mathbf{X}(z)\mathbf{X}^H(z)$ ($\mathbf{X}^H(z)\mathbf{X}(z)$) is nonsingular, then $\mathbf{X}(z)$ is said to be full column (row) rank. Furthermore if $\mathbf{X}(z)$ is full column (row) rank, its pseudoinverse $\mathbf{X}^\dagger(z)$ is defined as $(\mathbf{X}^H(z)\mathbf{X}(z))^{-1}\mathbf{X}^H(z)$ $(\mathbf{X}^H(z)(\mathbf{X}(z)\mathbf{X}^H(z))^{-1})$.

tr $\mathbf{X}$ represents the trace of square matrix $\mathbf{X}$. The Frobenius norm of matrix $\mathbf{X}$ is defined as $\|\mathbf{X}\| \triangleq (\mathrm{tr}\mathbf{X}\mathbf{X}^H)^{1/2}$. Also the Frobenius norm of transfer function $\mathbf{X}(z)$ is defined as $\|\mathbf{X}(z)\| \triangleq (\Sigma_\tau \|\mathbf{X}_\tau\|^2)^{1/2}$ or equivalently $\|\mathbf{X}(z)\| \triangleq (\int_{-1/2}^{1/2} \|\tilde{\mathbf{X}}(f)\|^2 \, df)^{1/2}$.

diag $\{d_1, ..., d_N\}$ represents the diagonal matrix with diagonal entries $d_1, ..., d_N$. Given a square matrix $\mathbf{X}$, diag $\mathbf{X}$ (off-diag $\mathbf{X}$) sets its off-diagonal (diagonal) entries to zeros.

III. Mixing process and demixing process

Let $\mathbf{x}(t) = [x_1(t), ..., x_M(t)]^T$ be a vector signal measured by M sensors at discrete time t. Assume that it is a sum of a convolutive mixture of statistically independent signals $\mathbf{s}(t) = [s_1(t), ..., s_N(t)]^T$ and a noise signal $\mathbf{n}(t) = [n_1(t), ..., n_M(t)]^T$. Namely, $\mathbf{x}(t)$ is given by

$$\mathbf{x}(t) = \Sigma_k \mathbf{A}_k \mathbf{s}(t-k) + \mathbf{n}(t) = \mathbf{A}(z)\mathbf{s}(t) + \mathbf{n}(t), \qquad (1)$$

where $\mathbf{A}(z) \triangleq \Sigma_k \mathbf{A}_k z^{-k}$ is an unknown $M \times N$ transfer function matrix ($M \geq N$). In this paper we refer to $\mathbf{A}(z)$ as a mixing process. $\mathbf{A}(z)$ is assumed to be full column rank and $\mathbf{n}(t)$ is independent of $\mathbf{s}(t)$. If the magnitude of $\mathbf{n}(t)$ is small compared with that of $\mathbf{s}(t)$, the source $\mathbf{s}(t)$ can adequately be estimated by a process taking the form

$$\mathbf{y}(t) = \Sigma_l \mathbf{W}_l \mathbf{x}(t-l) = \mathbf{W}(z)\mathbf{x}(t), \qquad (2)$$

where $\mathbf{y}(t) = [y_1(t), ..., y_N(t)]^T$ and $\mathbf{W}(z) = \Sigma_l \mathbf{W}_l z^{-l}$. Hereafter we refer to $\mathbf{W}(z)$ as a demixing process or a separator.

If $\mathbf{A}(z)$ is known beforehand, by putting its pseudoinverse $\mathbf{A}^\dagger(z)$ to $\mathbf{W}(z)$, we obtain independent components $\mathbf{y}(t) = \mathbf{s}(t)$ (strictly speaking, for $\mathbf{n}(t) = \mathbf{0}$). Thus the task of ICA is basically to estimate $\mathbf{A}^\dagger(z)$ and a set of independent components by eqn. (2) only from the observation. Unfortunately, since the only information given is that $\mathbf{s}(t)$ is mutually independent, $\mathbf{W}(z)$ cannot uniquely be determined.

Let $d_i(z)$ denote an arbitrary linear filter. Then, it is impossible to distinguish $\{s_i(t)\}$ from $\{d_i(z)s_i(t)\}$ only by independency among the signals. Namely, for an arbitrary diagonal matrix $\mathbf{D}(z) = \mathrm{diag}\{d_1(z), ..., d_N(z)\}$, $\mathbf{W}(z) = \mathbf{D}(z)\mathbf{A}^\dagger(z)$ can be considered valid as a demixing process.

In the case where $N < M$, furthermore, a different type of indeterminacy arises. There exists an $N \times M$ non-zero matrix $\mathbf{V}(z)$ satisfying $\mathbf{V}(z)\mathbf{A}(z) = \mathbf{O}$. When using $\mathbf{W}(z)+\mathbf{V}(z)$ as a demixing process, we have

$$(\mathbf{W}(z)+\mathbf{V}(z))\mathbf{A}(z) = \mathbf{D}(z). \qquad (3)$$

This implies that $\mathbf{W}(z)+\mathbf{V}(z)$ is also valid as a separator. Let $\mathcal{S}$ represent the set of all valid separators. Then, any element in $\mathcal{S}$ proves to take the form

$$\mathbf{W}(z) = \mathbf{D}(z)\mathbf{A}^\dagger(z)+\mathbf{V}(z) \ (\mathbf{V}(z)\mathbf{A}(z) = \mathbf{O}). \qquad (4)$$

Substituting this and eqn.(1) into eqn.(2), we have

$$\mathbf{y}(t) = \mathbf{D}(z)\mathbf{s}(t) + (\mathbf{D}(z)\mathbf{A}^\dagger(z)+\mathbf{V}(z))\mathbf{n}(t). \qquad (5)$$

It can be proved that if the noise $\mathbf{n}(t)$ is white temporally as well as spatially, noise term $(\mathbf{D}(z)\mathbf{A}^\dagger(z)+\mathbf{V}(z))\mathbf{n}(t)$ can be made minimum by choosing $\mathbf{V}(z) = \mathbf{O}$.

As mentioned above, we have certain freedoms in determining $\mathbf{D}(z)$ and $\mathbf{V}(z)$. In this paper the number of arbitrarily designable transfer functions is referred to as 'the degree of freedom' for a transfer function matrix. In the present case, the degrees of freedom in $\mathbf{D}(z)$ and $\mathbf{V}(z)$ are N and $(MN-N^2)$, respectively.

IV. Minimal distortion principle and inverse minimal distortion principle

As describe above, in the case where the number of observed signals is more than that of source signals, a valid separator has two indeterminacies: $\mathbf{D}(z)$ and $\mathbf{V}(z)$. In order to obtain a unique demixing process, we must design $\mathbf{D}(z)$ and $\mathbf{V}(z)$ in accordance with some rational principles (constraints). The number of the constraints have to be equal to the degrees of freedom in $\mathbf{D}(z)$ and $\mathbf{V}(z)$. In order to eliminate the indeterminacies, one of the authors proposes two principles: minimal distortion principle (MDP) and inverse minimum distortion principle (IMDP) in [11].

At first, we give an idea for eliminating the indeterminacy in $\mathbf{D}(z)$. Here we suppose that $\mathbf{n}(t) = \mathbf{0}$. In order to determine the diagonal matrix $\mathbf{D}(z)$ uniquely, Matsuoka [11] proposes the following principle.

(P1) An optimal separator should minimize $P_1(\mathbf{W}(z)) = E[\|\mathbf{y}(t) - \mathbf{Q}\mathbf{x}(t)\|^2]$ in the set $\mathcal{S}$.

Here $\mathbf{Q}$ is an $N\times M$ matrix and is given by a designer. This principal is referred to as minimum distortion principal (MDP).

Substituting $\mathbf{W}(z) = \mathbf{D}(z)\mathbf{A}^{\dagger}(z) + \mathbf{V}(z)$ into $P_1(\mathbf{W}(z))$ and reexpressing the equation in frequency domain, we have

$$P_1(\mathbf{W}(z)) = \int_{-1/2}^{1/2}\left\{\mathrm{tr}(\tilde{\mathbf{D}}(f) - \mathbf{Q}\tilde{\mathbf{A}}(f))\Phi(f)(\tilde{\mathbf{D}}(f) - \mathbf{Q}\tilde{\mathbf{A}}(f))^H\right\}df \tag{6}$$

where $\Phi(f)$ is the Fourier transform of the autocorrelation matrix $E[\mathbf{s}(t)\mathbf{s}^T(t-\tau)]$. Since $\Phi(f)$ and $\tilde{\mathbf{D}}(f)$ are both diagonal matrices, eqn.(6) takes minimum value 0 if and only if $\tilde{\mathbf{D}}(f) = \mathrm{diag}(\mathbf{Q}\tilde{\mathbf{A}}(f))$ holds for every frequency. This condition is equivalent to

$$\mathbf{D}(z) = \mathrm{diag}(\mathbf{QA}(z)). \tag{7}$$

Let $\mathbf{J}$ be the $N\times M$ matrix whose entries are all unity and set as $\mathbf{Q} = \mathbf{J}$. Then, the output signals of the optimal separator become $y_i(t) = \Sigma_k\, a_{ki}(z)s_i(t)$ $(i = 1, ..., N)$.

In the particular case where $N = M$, setting $\mathbf{Q}$ to the identity matrix $\mathbf{I}$ gives the optimal diagonal matrix as $\mathbf{D}(z) = \mathrm{diag}(\mathbf{A}(z))$. Accordingly the optimal demixing process and its output signal are respectively given by

$$\mathbf{W}(z) = (\mathrm{diag}\,\mathbf{A}(z))\mathbf{A}^{-1}(z)\,, \tag{8}$$

$$y_i(t) = a_{ii}(z)s_i(t) \quad (i = 1, ..., N). \tag{9}$$

Eqn. (9) is a source signal contained in $x_i(t)$. In [10], an ICA algorithm for this particular case, $\mathbf{Q} = \mathbf{I}$, is described. Thus the principle (P1) is a generalization of the principle proposed in [10] for the case where $M \geq N$.

Next we show how to eliminate the remaining indeterminacy in $\mathbf{V}(z)$. As described in the last section, $\mathbf{V}(z) = \mathbf{O}$ is an optimal choice in a certain sense. It can be proved that $\mathbf{V}(z) = \mathbf{O}$ is equivalent to

$$\mathbf{W}^{\dagger}(z)\mathbf{W}(z)\mathbf{A}(z) = \mathbf{A}(z). \tag{10}$$

From this, the optimal separator could be defined as a valid separator satisfying eqn.(10). Such a separator could be found by minimizing the distance between $\mathbf{W}^{\dagger}(z)\mathbf{W}(z)\mathbf{A}(z)$ and $\mathbf{A}(z)$. Unfortunately, however, the minimization cannot be done because $\mathbf{A}(z)$ is unknown in advance. Instead, we therefore introduce the following principle:

(P2) the optimal separator must minimize $P_2(\mathbf{W}(z)) \triangleq E[\|\mathbf{x}(t) - \mathbf{W}^{\dagger}(z)\mathbf{y}(t)\|^2]$ in the set $\mathcal{S}$.

Substituting eqn.(4) into $P_2(\mathbf{W}(z))$, we have in the frequency domain

$$P_2(\mathbf{W}(z)) = \int_{-1/2}^{1/2}\mathrm{tr}\left\{\tilde{\mathbf{A}}^H(f)\tilde{\mathbf{A}}(f) - \left((\tilde{\mathbf{A}}^H(f)\tilde{\mathbf{A}}(f))^{-1} + \tilde{\mathbf{D}}^{-1}(f)\tilde{\mathbf{V}}(f)\tilde{\mathbf{V}}^H(f)\tilde{\mathbf{D}}^{-H}(f)\right)^{-1}\right\}\Phi(f)df$$

This equation takes the minimum 0 if and only if $\tilde{\mathbf{V}}(f) = \mathbf{O}$ $(-1/2 \leq f \leq 1/2)$, namely $\mathbf{V}(z) = \mathbf{O}$. For valid separators, this principle provides the same number of constraints as the degree of freedom in $\mathbf{V}(z)$. This principle is referred to as 'inverse minimal distortion principle'.

V. Natural gradient method and nonholonomic constraints

Amari [1] proposes the natural gradient learning method. The learning rule is efficient for minimization or maximization of a function with respect to its parameters. Although Amari describes the natural gradient adaptation for square matrices, it is easy to extend the idea to rectangular transfer matrices.

At each point $\mathbf{W}(z)$ on the manifold formed by all $N\times M$ transfer matrices of full rank, we introduce the following two kinds of Riemannian metric for tangent vector $d\mathbf{W}(z) = \Sigma_k\, d\mathbf{W}_k z^{-k}$, that is, we define two norms for $d\mathbf{W}(z)$:

(a) $\|d\mathbf{W}(z)\|_Y \triangleq \|d\mathbf{W}(z)\mathbf{W}^{\dagger}(z)\|$,

(b) $\|d\mathbf{W}(z)\|_X \triangleq \|\mathbf{W}^{\dagger}(z)d\mathbf{W}(z)\|$.

In definition (a), note that the rows of $d\mathbf{W}(z)$ must be constrained on the space spanned by the row vectors of $\mathbf{W}(z)$.

Let $f(\mathbf{W}(z))$ be a scalar function to be minimized with respect to $\mathbf{W}(z)$. Then, the natural gradient learning algorithms associated with the above two metrics are respectively given by

$$\frac{d\mathbf{W}(z)}{du} = -\frac{\partial f(\mathbf{W}(z))}{\partial \mathbf{W}(z)}\mathbf{W}^H(z)\mathbf{W}(z)\,, \tag{11}$$

$$\frac{d\mathbf{W}(z)}{du} = -\mathbf{W}(z)\mathbf{W}^H(z)\frac{\partial f(\mathbf{W}(z))}{\partial \mathbf{W}(z)}\,, \tag{12}$$

where $\partial f(\mathbf{W}(z))/\partial \mathbf{W}(z) \triangleq \sum_\tau (\partial f(\mathbf{W}(z))/\partial \mathbf{W}_\tau)z^{-\tau}$. Note that parameter u in these equations represent 'time' for the updating of $\mathbf{W}(z)$. The derivation of eqn.(12) is shown in Appendix. For eqn. (11), we omit its derivation because we do not have enough pages.

To the first learning rule, we apply an idea of nonholonomic constraint, which is proposed by Choi et al.[6]. Adopting diag $d\mathbf{W}(z)\mathbf{W}^{\dagger}(z) = \mathbf{O}$ and off-diag $d\mathbf{W}(z)\mathbf{W}^{\dagger}(z) = \mathbf{O}$ as nonholonomic constraints, respectively, we have the following algorithms:

$$\frac{d\mathbf{W}(z)}{du} = -\,\text{off-diag}\,\frac{\partial f(\mathbf{W}(z))}{\partial \mathbf{W}(z)}\mathbf{W}^H(z)\cdot\mathbf{W}(z)\,, \tag{13}$$

$$\frac{d\mathbf{W}(z)}{du} = -\,\mathrm{diag}\,\frac{\partial f(\mathbf{W}(z))}{\partial \mathbf{W}(z)}\mathbf{W}^H(z)\cdot\mathbf{W}(z)\,. \tag{14}$$

VI. ICA algorithm

We adopt the same evaluation function as used in Amari et al.[3] and Bell and Sejnowski [5]. It is

$$I(\mathbf{W}(z)) \triangleq -\sum_{i=1}^N E[\log q_i(y_i(t))] - h[\mathbf{y}(t)]\,, \tag{15}$$

where $h[\mathbf{y}(t)]$ denotes entropy rate for time series $\{... , \mathbf{y}(-1), \mathbf{y}(0), \mathbf{y}(1), ...\}$. $q_i(s)$ is a function being a model for a probability density function (pdf) of signal $s_i(t)$. If $s_i(t)$ is iid and $q_i(s_i)$ is selected so as to be close to the true pdf of source signal $s_i(t)$, a separator can be obtained by minimizing $I(\mathbf{W}(z))$ with respect to $\mathbf{W}(z)$. Setting $I(\mathbf{W}(z))$ to $f(\mathbf{W}(z))$ in eqn.(13), we have

$$\frac{d\mathbf{W}(z)}{du} = -\,\text{off-diag}\,E[\varphi(\mathbf{y}(t))\mathbf{y}^H(t,z)]\cdot\mathbf{W}(z), \quad (16)$$

where $\varphi_i(y) \triangleq -d\log q_i(y)/dy$ and $\varphi(\mathbf{y}(t)) \triangleq [\varphi_1(y_1(t)), \ldots, \varphi_N(y_N(t))]^T$ and $\mathbf{y}(t, z) \triangleq \Sigma_\tau \mathbf{y}(t+\tau)z^{-\tau}$. According to Amari et al. [2] and Bell and Sejnowski [5], for sub-Gaussian source $\varphi_i(u) = u^3$ is set and for super-Gaussian source, $\varphi_i(u) = \tanh(u/2)$. Equilibrium points of eqn.(16) are $\mathbf{W}(z)$ that satisfies

$$\text{off-diag}\,E[\varphi(\mathbf{y}(t))\mathbf{y}^H(t, z)] = \mathbf{O}. \quad (17)$$

This holds if $y_i(t)$ ($i = 1, \ldots, N$) are mutually independent, namely, every element $\mathbf{W}(z)$ in $\mathcal{S}$ satisfies eqn.(17). Accordingly the equilibrium still has the indeterminacies of $\mathbf{D}(z)$ and $\mathbf{V}(z)$. To determine these indeterminacies, we use the dynamics (18) and (19).

Setting $f(\mathbf{W}(z)) = P_1(\mathbf{W}(z))$ in eqn.(13), we have

$$\frac{d\mathbf{W}(z)}{du} = -\,\text{diag}\,E[(\mathbf{y}(t)-\mathbf{Q}\mathbf{x}(t))\mathbf{y}^H(t,z))]\cdot\mathbf{W}(z). \quad (18)$$

This dynamic seeks a valid separator for which $\mathbf{D}(z)$ equals (7).

Furthermore putting $P_2(\mathbf{W}(z))$ to $f(\mathbf{W}(z))$ in eqn.(12), we have

$$\frac{d\mathbf{W}(z)}{du} = E[\mathbf{y}(t)\mathbf{x}^H(t,z)] - E[\mathbf{y}(t)\mathbf{y}^H(t,z)]\bar{\mathbf{U}}(z)\mathbf{W}(z), \quad (19)$$

where $\bar{\mathbf{U}}(z) \triangleq (\mathbf{W}(z)\mathbf{W}^H(z))^{-1}$. This dynamic provides $(MN-N^2)$ constraints for $\mathbf{W}(z)$ and removes $\mathbf{V}(z)$.

Set $\mathbf{W}(z) = \sum_{\tau=-L}^{L}\mathbf{W}_\tau z^{-\tau}$ and let $\Delta\mathbf{W}_\tau$ be update values for coefficients $\mathbf{W}_\tau$ of $\mathbf{W}(z)$. Combining eqn.(16), eqn.(18) and eqn.(19), we have eqn.(22) (next page). In eqn.(22) α, β and γ are all small positive constants. The separator's output is calculated as

$$\mathbf{y}(t-L) = \sum_{\tau=-L}^{L}\mathbf{W}_\tau\mathbf{x}(t-L-\tau). \quad (20)$$

In eqn.(19) $\bar{\mathbf{U}}(z) = \sum_{\tau=-L}^{L}\bar{\mathbf{U}}_\tau z^{-\tau}$ represent the inverse of $\mathbf{U}(z) = \mathbf{W}(z)\mathbf{W}^H(z)$. The inverse is found by the following recursive rule:

$$\bar{\mathbf{U}}^{(l)}(z) = 2\bar{\mathbf{U}}^{(l-1)}(z) - \bar{\mathbf{U}}^{(l-1)}(z)\mathbf{U}(z)\bar{\mathbf{U}}^{(l-1)}(z). \quad (21)$$

Here $\bar{\mathbf{U}}^{(l)}(z)$ denotes an estimate of $\bar{\mathbf{U}}(z)$ at the l-th step. Note that the initial value $\bar{\mathbf{U}}^{(0)}(z)$ must be para-Hermitian and be chosen close to $\mathbf{U}^{-1}(z)$. If $\bar{\mathbf{U}}^{(0)}(z)$ is set to the estimated value of $\mathbf{U}^{-1}(z)$ at the previous step in the updating of $\mathbf{W}(z)$, $\bar{\mathbf{U}}^{(l)}(z)$ converges to $\mathbf{U}^{-1}(z)$ in a few steps.

VII. Results

We applied the above-mentioned algorithm to EGG data measured by 4 electrodes ($M = 4$) as shown in Fig.1. Let $\mathbf{x}(t) = [x_1(t), \ldots, x_4(t)]^T$ represent a sampled value of the four electrodes at sample time t. The data was sampled at 1[Hz] and consists of 10,000 samples. The power spectra of the EGG data are shown in Fig. 2. One can see that the component associated with the gastric activity is corrupted with other components.

Our algorithm is applicable to the case where the number of independent signals is less than that of their mixtures. We applied the algorithm to the data in each case that the number of sources is assumed to be two, three or four (N = 2, 3, or 4). We set the order of a separator to 31 (L = 15), and for each case the initial values of the separator's parameters are set as

$$\mathbf{W}(z) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \text{ (for } N = 2\text{)},$$

$$\mathbf{W}(z) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ (for } N = 3\text{)},$$

$$\mathbf{W}(z) = \mathbf{I} \text{ (for } N = 4\text{)}.$$

We assumed that the sources are all sub-Gaussian and set all nonlinear functions in eqn.(16) as $\varphi_i(u) = u^3$. Parameters α, β and γ were chosen as $\alpha = 1.0\times10^{-5}$, $\beta = \gamma = 1.0\times10^{-7}$, respectively. In all the cases, the number of iterations of the update algorithm for $\{\mathbf{x}(t)\}$ was 600 (the total number of iterations for the estimation of $\mathbf{W}(z)$ is 6×10^6) and at each update of $\mathbf{W}(z)$, the number of iterations in the recursive algorithm (21) was 2.

The power spectra of the separated signals are shown in Fig. 3. According to the property of the electrical potential corresponding to the stomach's peristalsis, output $y_2(t)$ is guessed to be the component from the stomach.

$$\begin{aligned}\Delta\mathbf{W}_\tau = &-\alpha\textstyle\sum_r \text{off-diag}\,\varphi(\mathbf{y}(t))\mathbf{y}^T(t-3L-\tau+r)\cdot\mathbf{W}_r \\ &-\beta\textstyle\sum_r \text{diag}\big(\mathbf{y}(t-3L)-\mathbf{Q}\mathbf{x}(t-3L)\big)\mathbf{y}^T(t-3L-\tau+r)\cdot\mathbf{W}_r \\ &+\gamma\big\{\mathbf{y}(t-3L)\mathbf{x}^T(t-3L-\tau)-\textstyle\sum_r\sum_k\mathbf{y}(t-3L)\mathbf{y}^T(t-3L-\tau+r)\bar{\mathbf{U}}_{r-k}\mathbf{W}_k\big\}\end{aligned} \quad (22)$$

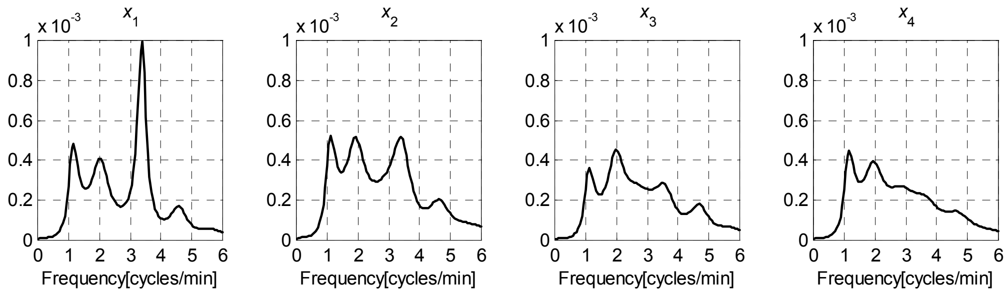
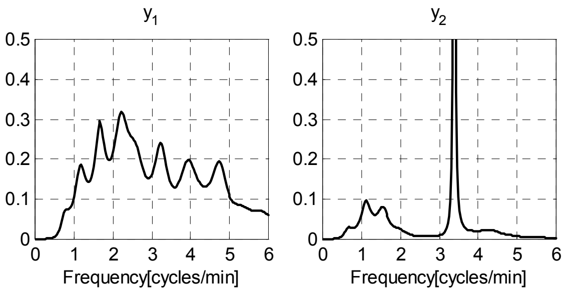
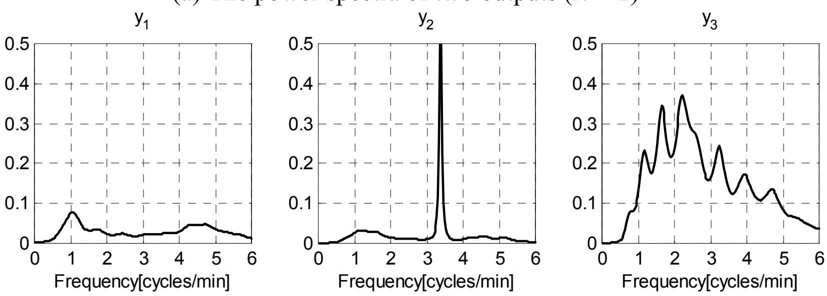


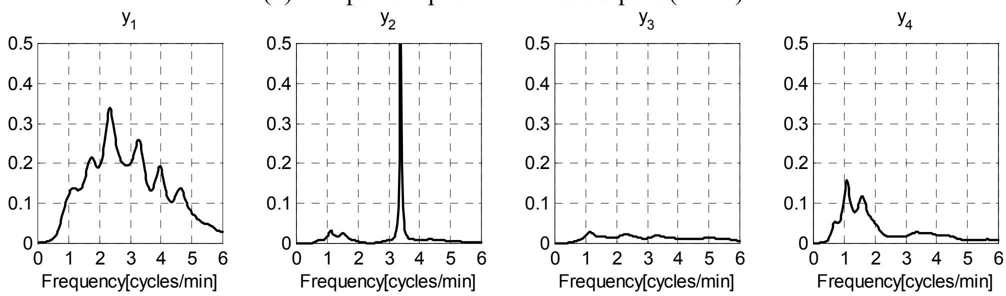
Fig.2 The power spectra of four channel EGG data



(a) The power spectra of two outputs ($N = 2$)



(b) The power spectra of three outputs ($N = 3$)



(c) The power spectra of four outputs ($N = 4$)

Fig.3 The power spectra of the outputs in four cases

IIX. Conclusion

We have shown an application of ICA to EGG data. Our algorithm is applicable to the case where the number of the observations is more than that of the sources. We applied the algorithm, assuming that the number of the sources were two, three, and four. In every case, the algorithm extracted two dominant components. This result suggests that the EGG signals detected by the electrodes seem to be mainly originated from two independents sources. One of them is most likely to be associated with the stomach's activity while the origin of the other component is unknown at present.

Reference

[1] S. Amari, "Natural gradient learning works efficiently in learning," Neural Computation, vol.10, pp. 51-276, 1998.

[2] S. Amari, T.P. Chen, and A. Cichocki, "Stability analysis of adaptive blind source separation," Neural Networks, vol.10, no.8, pp.1345-1351, 1997.

[3] S. Amari, S.C. Douglas, A. Cichocki and H.H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in Proc. IEEE International Workshop on Wireless Communication, pp. 101-104, 1997.

[4] S. Amari, T.P. Chen, and A. Cichocki, "Nonholonomic orthogonal learning algorithms for blind source separation," Neural Computation, vol.12, no.6, pp.1463-1484, 2000.

[5] J. Bell and T.J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," Neural Computation, vol.7, no.6, pp.1129-1159, 1995.

[6] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," in Proc. International Workshop on Independent Component Analysis and Blind Signal Separation, pp.371-376, 1999.

[7] H. Liang, "Adaptive independent component analysis of multichannel electrogastrograms," Medical Engineering & Physics, vol. 23, pp.91-97, 2001.

[8] M. Ohata, T. Matsumoto, A. Shigematsu and K. Matsuoka, "Kurtosis-based algorithm for blind sources separation using the Cayley transformation and Its application to multi-channel electrogastrograms," in CD-ROM of Proceedings of Korea Automatic Control Conference, TPI10-3, 2000.

[9] M. Ohata and K. Matsuoka, "Stability analyses of information-theoretic blind separation algorithms in the case where the sources are nonlinear processes," IEEE Trans. on Signal Processing, vol.50, no.1, pp.69-77, 2002.

[10] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in Proc. International Workshop on Independent Component Analysis and Blind Signal Separation, pp.722-727, 2001.

[11] K. Matsuoka, "Principles for eliminating two kinds of indeterminacy in blind source separation," In Proc. of International Conference on Digital Signal Processing, pp. 147-150,2002.

[12] Z.S. Wang, J.Y. Cheung and J.D. Chen, "Blind separation of electrogastrograms using independent component analysis based on a neural network," Medical & Biological Engineering & Computing, vol.37, pp.80-86, 1999.

Appendix: The derivation of (12)

Defining $\mathbf{C}(z) = \Sigma_m \mathbf{C}_m z^{-m} = \mathbf{W}^{\dagger}(z)$ ($\mathbf{C}_m$ are $M\times N$ matrices), we have

$$\| d\mathbf{W}(z) \|_X^2 = \sum_{l,m,n} \mathrm{tr}\{\mathbf{C}_l d\mathbf{W}_m d\mathbf{W}_n^T \mathbf{C}_{l+m-n}^T\}.$$

When $\mathbf{W}(z)$ changes from $\mathbf{W}(z)$ to $\mathbf{W}(z) + d\mathbf{W}(z)$, the variation of the scalar function $f(\mathbf{W}(z))$ is given by

$$df(\mathbf{W}(z)) \triangleq f(\mathbf{W}(z)+d\mathbf{W}(z)) - f(\mathbf{W}(z)) = \Sigma_l \, \mathrm{tr}\left(\frac{\partial f}{\partial \mathbf{W}_l} d\mathbf{W}_l^T\right).$$

According to the natural gradient method, we obtain $\{d\mathbf{W}_k\}$ such that $df(\mathbf{W}(z))$ takes a minimum under the constraint $\| d\mathbf{W}(z) \|_X^2 = \varepsilon^2$ (= const.). We solve the problem by Lagrange's method of indeterminate multiplier. Define the Lagrangean function

$$L(\{d\mathbf{W}_k\}) = \Sigma_l \, \mathrm{tr}\left(\frac{\partial f}{\partial \mathbf{W}_l} d\mathbf{W}_l^T\right) - \frac{\lambda}{2} \sum_{l,m,n} \mathrm{tr}\{\mathbf{C}_l d\mathbf{W}_m d\mathbf{W}_n^T \mathbf{C}_{l+m-n}^T\},$$

where λ represents a Lagrangean multiplier. The optimal values of $\{d\mathbf{W}_k\}$ must satisfy

$$\frac{\partial L}{\partial (d\mathbf{W}_k)} = \frac{\partial f}{\partial \mathbf{W}_k} - \lambda \sum_{l,m} \mathbf{C}_{l+m-k}^T \mathbf{C}_l d\mathbf{W}_m = \mathbf{O} \quad (k = ..., -1, 0, 1, ...).$$

From this, we have

$$\lambda \mathbf{C}^H(z)\mathbf{C}(z) d\mathbf{W}(z) = \frac{\partial f}{\partial \mathbf{W}(z)}.$$

Since $\mathbf{C}^H(z)\mathbf{C}(z)$ becomes $(\mathbf{W}(z)\mathbf{W}^H(z))^{-1}$ and $d\mathbf{W}(z)$ must be a direction such as $df(\mathbf{W}(z)) < 0$, we have eqn. (12).