

**BLIND SEPARATION FOR FETAL ECG  
FROM SINGLE MIXTURE BY SVD AND ICA**

**GAO PING**

*(B.Sc., Xi'an Highway University)*

**A THESIS SUBMITTED  
FOR THE DEGREE OF MASTER OF SCIENCE  
DEPARTMENT OF COMPUTATIONAL SCIENCE  
NATIONAL UNIVERSITY OF SINGAPORE**

**2003**

---

# Acknowledgments

---

I would like to thank my supervisor, Dr. Chang Ee-Chien, who gave me the opportunity to work on such an interesting research project, paid patient guidance to me, and gave me much invaluable help and constructive suggestion on it.

It is also my pleasure to express my appreciation to Dr. Lonce Wyse and Mr. Liu Bao for their inspiring ideas.

I would also wish to thank Chia Ee Ling for providing the ECG signals.

My sincere thanks go to all my department-mates and my friends in Singapore for their friendship and so much kind help.

I would like also to dedicate this work to my parents, my brothers and my husband, for their unconditional love and support.

**Gao Ping**  
**March 2003**

---

# Contents

---

<b>Acknowledgments</b>	<b>ii</b>
<b>Summary</b>	<b>vi</b>
<b>List of Figures</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 General Introduction . . . . .	1
1.2 Previous techniques for fetal ECG extraction . . . . .	3
1.3 Outlines . . . . .	7
<b>2 Independent component analysis</b>	<b>8</b>
2.1 Motivation . . . . .	8
2.2 Mathematical model . . . . .	9
2.3 Illustration of ICA . . . . .	10
2.4 Independence . . . . .	11
2.5 Information theory background . . . . .	13
2.5.1 Entropy . . . . .	13

---

2.5.2	Negentropy . . . . .	14
2.5.3	Mutual information . . . . .	14
2.6	Approach to ICA with data model assumption . . . . .	15
2.6.1	Nongaussianity for ICA model . . . . .	15
2.6.2	Measures of Nongaussianity . . . . .	16
2.7	Approach to ICA without data model assumption . . . . .	19
2.8	Other approaches to ICA . . . . .	20
2.9	Practical Contrast Functions . . . . .	20
2.10	Conclusion . . . . .	22
<b>3</b>	<b>FastICA—an algorithm for ICA</b>	<b>23</b>
3.1	Introduction . . . . .	23
3.2	Fixed-point algorithm for one unit . . . . .	24
3.3	FastICA for several units . . . . .	25
3.4	FastICA algorithm . . . . .	26
3.5	Conclusion . . . . .	28
<b>4</b>	<b>Fetal ECG extraction</b>	<b>29</b>
4.1	Introduction . . . . .	29
4.2	Heart beats occurrence detection . . . . .	31
4.2.1	Motivation . . . . .	31
4.2.2	Problem formulation . . . . .	33
4.2.3	Proposed method for finding trends of original signal . . . . .	34
4.3	Fetal ECG complex detection . . . . .	36
4.3.1	Main idea . . . . .	36
4.3.2	Proposed method for fetal ECG extraction . . . . .	37
4.4	Refining for ECG . . . . .	41

---

4.4.1	Choice of window width of spectrogram . . . . .	41
4.4.2	Selecting the best component after ICA . . . . .	42
4.5	Conclusion . . . . .	42
<b>5</b>	<b>Programmes and experimental results</b>	<b>44</b>
5.1	Programmes structure . . . . .	44
5.2	Experimental results . . . . .	46
5.2.1	Synthetical data and results . . . . .	46
5.2.2	Experiments on real-life data . . . . .	51
5.2.3	Fetal ECG extraction . . . . .	53
5.2.4	ECG complex results . . . . .	54
5.3	Conclusion . . . . .	54
<b>6</b>	<b>Discussion and Conclusion</b>	<b>58</b>
6.1	Discussion . . . . .	58
6.2	Conclusion . . . . .	59

---

# Summary

---

In this thesis, we extract the fetal ECG from a single-channel abdominal ECG. The abdominal ECG consists of three parts: maternal ECG, fetal ECG and noise.

we propose a novel blind-source separation method to extract Fetal ECG from a single-channel signal measured on the abdomen of the mother. Our proposed method includes two parts: first is to detect the heart beats occurrence, the second part is to extract the fetal ECG and compute the ECG complex.

In the first part, the key idea is to compute the spectrogram of the original signal, and then use an assumption of statistical independence to find trends of the original signal. This is achieved by applying Singular Value Decomposition (SVD) on the spectrogram, followed by an iterated application of Independent Component Analysis (ICA) on the principle components. The SVD contributes to the separability of each component and the ICA contributes to the independence of the two components. We further refine and adapt the above general idea to ECG by exploiting a-prior knowledge of the maternal ECG frequency distribution and other characteristic of ECG. Experimental studies show that the proposed method is more

accurate than using SVD only. Because our method does not exploit extensive domain knowledge of the ECGs, the idea of combining SVD and ICA in this way can be applied to other blind separation problems.

In the second part, we construct a pure maternal ECG and then subtract it from the mixture to obtain the fetal ECG. Fetal ECG can then be produced by time domain averaging.

Experimental results on both synthetic and real-life data gives good results.

---

# List of Figures

---

- 2.1 Joint pdf for sources and mixtures . . . . . 10
  
- 4.1 The whole original signal . . . . . 30
- 4.2 Detail of the original signal . . . . . 30
- 4.3 Spectrogram of the original signal(108.raw). . . . . 33
- 4.4 Original mixture and the segments . . . . . 39
- 4.5 Large complex template . . . . . 39
- 4.6 Shift procedure . . . . . 40
- 4.7 Purely large complex signal . . . . . 40
- 4.8 Small complex signal(after removing the large complex signal) . . . . . 40
- 4.9 Small complex . . . . . 41
- 4.10 Frequency (108.raw). . . . . 42
  
- 5.1 Programme Structure . . . . . 45
- 5.2 Synthetic maternal ECG complex . . . . . 47
- 5.3 Synthetic fetal ECG complex . . . . . 47
- 5.4 Synthetic data: Constructed by Figure.5.2 and Figure.5.3 . . . . . 48



---

5.5	Comparison for the results from SVD and SVD+ICA on synthetic data	48
5.6	Synthetical data detection result for strength ratio=4 . . . . .	49
5.7	Synthetical data detection result for strength ratio=5 . . . . .	49
5.8	Synthetical data detection result for strength ratio=6 . . . . .	49
5.9	Detection accuracy for different strength ratio between maternal and fetal ECG . . . . .	50
5.10	Synthetical data detection result when noise level= 10 . . . . .	50
5.11	Original recorded data:108.raw . . . . .	51
5.12	Original recorded data:292.raw . . . . .	51
5.13	Comparison of results by SVD and SVD+ICA for maternal heart beats occurrence detection(108.raw) . . . . .	52
5.14	Comparison of results by SVD and SVD+ICA for maternal heart beats occurrence detection(292.raw) . . . . .	52
5.15	Another example: fetal heart beats occurrence detection by SVD + ICA. Arrows indicates heart beats that are difficult to detect. . . . .	52
5.16	Fetal trend comparison of SVD and ICA for 292.raw. Arrows indicates heart beats that are difficult to detect. . . . .	53
5.17	Fetal Trend by ICA for 108.raw after removing maternal ECG . . . . .	54
5.18	Fetal Trend by ICA for 292.raw after removing maternal ECG . . . . .	54
5.19	Maternal ECG complex for 108.raw . . . . .	55
5.20	Maternal ECG complex for 292.raw . . . . .	55
5.21	Fetal ECG complex for 108.raw . . . . .	55
5.22	Fetal ECG complex for 292.raw . . . . .	56
5.23	Original signal: 108.raw . . . . .	56
5.24	Fetal ECG for 108.raw . . . . .	57
5.25	Original signal: 292.raw . . . . .	57

---

5.26 Fetal ECG for 292.raw . . . . .	57
--------------------------------------	----

# Introduction

## 1.1 General Introduction

Fetal Electrocardiogram(ECG) plays an important role for determining the neurological status after birth[25, 42]. Even though the accurate fetal ECG may be obtained by placing an electrode on the fetal scalp, however, as long as the membranes protecting the child have not been broken, one should look for noninvasive techniques. So, the most popular approach to get fetal ECG is studying the ECG recordings measured by placing electrodes on the mother's skin.

Considering the small heart of the fetus and the low voltage current it generates compared with that of the mother, electrodes are usually placed on the abdomen of the mother(it is called abdominal ECG or the mixture) as close as possible to the fetal heart, and expect that at least one of the electrodes will have the fetal ECG with high enough SNR(signal-to-noise ratio). Thoracic ECG(measured on the thorax of the pregnant woman) is also needed for some methods which could be used to cancel out the effects of the maternal trace[3, 14, 30, 32, 35, 37].

However, signals recorded in this way are severely contaminated by the existence

of the maternal ECG which could be 5–1000 times higher than fetal ECG in its intensity. Furthermore, the weak recordings of fetal ECG may contain a relatively large amount of noise and may also be distorted by muscle and breathing contractions. Moreover, this is further complicated by the positioning of electrodes which by no means nontrivial.

Thus, we face a twofold problem: one is to separate the fetal ECG from the strong maternal trace, the other is to separate the fetal ECG from the noise.

In the past decades, engineers developed many different techniques to extract the FECG signals. In the 1960s, conventional filters and direct cancellation were used separately to remove the maternal ECG from the abdominal mixtures. Based on the Least Mean Square algorithm, Widrow in 1975 proposed an adaptive filtering technique to separate fetal ECG from maternal ECG. Later in 1977, Reichert generated three spatially orthogonal ECG signals from three linearly independent thoracic ECG signals, and then the proper coefficients with the three signals were selected to simulate the MECG component in abdominal ECG signals. In 1981, Bergveld adopted six independent abdominal signals to obtain maternal ECG interference suppression. Vandershoot in 1987 applied two matrix methods for the optimal maternal ECG elimination and fetal ECG detection. The more recent approach includes blind source separation which aims to find the sources from blind source separation(BSS) and SVD.

Most of these methods focus on multi-channel mixtures of signals [5, 6, 50, 51]. Relatively few works address the problem separating ECG signals recorded on a single-channel. Kanjilal et. al. [29] developed a method for single-channel signals by first detecting both the maternal and fetal heart beats. Next, “cut” the signal into pieces. These pieces are aligned (to form a matrix) and SVD is then performed to obtain the ECG complex.

In this thesis, we consider a single-channel recording. By projecting into a higher dimension, we can then employ a multi-channel technique. The proposed method has two unique features: 1) only a single abdominal signal is required and 2) the detection could be achieved as real-time applications. In later chapter, we will give the details on both the theoretical backgrounds and the procedure of implementations.

## 1.2 Previous techniques for fetal ECG extraction

Since 1960, many methods are proposed to extract the fetal ECG. According to the different input of each method, the methods can be classified into three categories. Two categories need one more mixture, and the difference between them is whether the thoracic signals are required, while the third category mainly focus on the fetal ECG extraction from single-channel abdominal ECG which is also the aim of our proposed method.

*Mathematical Model:*

Signals can be written as:

$$A_i^a(t) = M_i^a(t) + F_i^a(t) \quad (1.1)$$

$$T_i^t(t) = M_i^t(t) \quad (1.2)$$

where  $M_i^a(t)$ ,  $F_i^a(t)$  and  $M_i^t(t)$  are the abdominal MECG, FECG and thoracic MECG respectively.  $T_i^t(t)$  just contains thoracic MECG while  $A_i^a(t)$  is the mixture of the abdominal MECG and FECG.

The model would be more realistic to assume that there is some noise in  $A_i^a(t)$  and  $T_i^t(t)$ , however, since the estimation of the noise-free model is difficult enough itself, the noise terms are usually omitted in practice. Anyway, we could denoise before we use any methods to make sure that this model is enough.

Different methods have different assumption on the relationship between the abdominal MECG and thoracic MECG. Some simple methods assume that they are the same, some generate a new MECG for abdominal ECG by using several thoracic signals, some obtain an abdominal MECG from several abdominal signals, and single-channel fetal ECG extraction are trying to cancel out the interference of maternal ECG from the same abdominal signal.

*Subtraction:* Subtraction method was the first and simplest technique for detecting and enhancing the fetal ECG. It assumes that  $M_i^a(t) = M_i^t(t)$ . By applying the model, the fetal ECG can be obtained by:

$$F_i(t) = A_i(t) - T_i(t) \quad (1.3)$$

*Orthogonal analysis:* However, this simplest method does not produce very good results. The reason that direct subtraction fails is the mismatch between  $T_i(t)$  and  $M_i^a(t)$ . In order to overcome this problem, R.L. Longini in 1977 took three separate thoracic signals and constructed the fourth ECG signal which serves as  $M_i^a(t)$  the maternal ECG part of the abdominal ECG.

$$M_i^a(t) = \Gamma_1 T_1^t(t) + \Gamma_2 T_2^t(t) + \Gamma_3 T_3^t(t) \quad (1.4)$$

After getting  $M_i^a(t)$ , fetal ECG could be computed similarly as the subtraction method by Eq.1.3.

Orthogonal analysis is better than subtraction in the sense that it tries to avoid the mismatch between the thoracic MECG and abdominal MECG, but the orthogonalization requirement of the three thoracic ECG signals by Gram-Schmidt procedure makes it difficult to implement in practice.

*Linear combination:* Bergveld, Meijer, Kolling and Peuscher developed a linear

combination method based on the fact that any abdominal ECG may be represented by Eq.1.1.

Specifically, the abdominal ECG could be written as (note here the superscript is omitted since no thoracic ECG) :

$$A(t) = \sum_i \Gamma_i V_i(t) \quad (1.5)$$

$$V_i(t) = M_i(t) + F_i(t) \quad (1.6)$$

where  $\Gamma_i$  are optimized to produce a clear FECG. Now, rewrite the abdominal signal as:

$$A(t) = \sum_i \Gamma_i M_i(t) + \sum_i \Gamma_i F_i(t) \quad (1.7)$$

The goal is to optimize the  $\Gamma_i$  coefficients to produce an FECG from the chosen number of original signals such as:

$$\sum_i \Gamma_i M_i(t) = 0 \quad (1.8)$$

$$\sum_i \Gamma_i F_i(t) \neq 0 \quad (1.9)$$

Thus the fetal ECG could be achieved when several abdominal ECGs are combined through optimizing bounded coefficients.

Later, a lot of statistical methods are employed. The most popular one is the Blind Source Separation or Blind Signal Separation (BSS). Independent Component Analysis (ICA) is one of the most important approach for BSS. ICA needs at least the same number of mixtures as the number of the sources. Recently, Lathauwer et al. [11, 12, 13, 16, 38, 48], Zarzoso et al. [53] have attempted to separate maternal and fetal ECGs from cutaneous 8 – 32 channel recordings, by using ICA which assumes that the sources are statistically independent. For all the methods which need more than one mixtures, one aspect often ignored is the problem of eliminating the effects

of different interferences of extraneous reasons (e.g. the influence of respiratory activity), all the methods for multi-channel extraction suffer from this problem. However, few works address the fetal ECG extraction on single channel abdominal ECG.

*Single-channel extraction:* P. P. Kanjilal[29, 31] exploits the nearly-periodic feature for separating M-ECG and F-ECG components by using SVD. Firstly, the data are arranged in the form of a matrix  $A$  such that the consecutive maternal ECG cycles occupy the consecutive rows, and the peak maternal component lies in the same column. SVD is performed on  $A : A = U\Sigma V$ , and  $A_M = u_1\sigma_1v_1^t$  is separated from  $A$ (where  $w_1$  and  $v_1$  are the first columns of the matrix  $U$  and  $V$  respectively), forming  $A_{R1} = A - A_M$ .

After separating the MECG component from composite signal, the time series formed from the successive rows of  $A_{R1}$  will contain FECG component along with noise; this series is rearranged into a matrix  $B$  such that each row contains one fetal ECG cycle, with the peak value lying in the same column. SVD is performed on  $B$ , from which the most dominant component  $u_1\sigma_1v_1^t$  is extracted, which will give the desired FECG component.

One point should be noted here is that the aligning is required in advance. In fact, even though the MECG peaks is easy to find, it is quite difficult to align the FECG which makes the algorithm difficult to implement.

There are still many other methods for fetal ECG extraction, such as subspace projection[46], nonlinear recursive algorithm[47] and wavelet-based method[33] etc.. Here, we will not introduce them one by one.



---

## 1.3 Outlines

In this thesis, we propose a novel method to extract fetal ECG from single-channel abdominal signal. This method is made up of two parts: one to detect the heart beats occurrence and the other is to extract the fetal ECG and detect the ECG complex.

By working on single-channel abdominal signal, the proposed method avoids the multi-interferences of extraneous reasons which all the multi-channel extraction suffer.

Results show that the proposed method works well not only for synthetic data but also for real-life data.

This thesis includes six chapters:

Chapter 2 introduce the Independent Component Analysis and the FastICA algorithm. Chapter 3 gives the algorithm for ICA. In Chapter 4, our proposed method on how to detect the heart beats occurrence and the ECG complex will be described. Chapter 5 are the experimental results on synthetic data and real-life data. The last chapter is the conclusion.

# Independent component analysis

## 2.1 Motivation

*Cock-tail party problem:* In a room, two people are speaking simultaneously, and two microphones are putting in different locations which are used to provide two recorded mixtures of the two speech signals. Denote the two mixture signals as  $x_1(t)$  and  $x_2(t)$ , the two speech signals as  $s_1(t)$  and  $s_2(t)$ . Here,  $t$  is the time index, and  $x_1, x_2, s_1$  and  $s_2$  are the amplitudes of the signals.

Since  $x_1(t)$  and  $x_2(t)$  are the weighted sum of  $s_1(t)$  and  $s_2(t)$ , this relation could be expressed as a linear equation:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \quad (2.1)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \quad (2.2)$$

where  $a_{11}, a_{12}, a_{21}$  and  $a_{22}$  are some parameters which rely on the distances of the microphones from the speakers. If the two speech signals  $s_1(t)$  and  $s_2(t)$  could be estimated based only on  $x_1(t)$  and  $x_2(t)$ , such estimation will be quite useful. For simplicity, any time delay or other extra factors are not be taken into account.

If the parameters  $a_{ij}$  are known,  $s_1(t)$  and  $s_2(t)$  would be obtained by solving the linear equation. However, the point is, if  $a_{ij}$  are unknown, how to solve the problem?

Such a problem is often called Blind Source Separation or Blind Signal Separation(BSS). There are many approaches to the BSS problem.

Several approaches are to exploit some information on statistics properties of  $s_1(t)$  and  $s_2(t)$  to estimate  $a_{ij}$ . Independent Component Analysis(ICA) is the approach which assumes that  $s_1(t)$  and  $s_2(t)$ , at each time instant  $t$ , are statistically independent. Amazingly, it proves to be enough to solve the cock-tail party problem by such assumption.

ICA was first developed to solve problems which are closely related to the cock-tail party problem. In recent years, due to the increase interest in ICA, ICA is found to be useful in many other applications[24, 34], such as feature extraction, EEG separation and data analysis etc. .

## 2.2 Mathematical model

Assume we have  $n$  linear mixtures  $x_1, x_2, \dots, x_n$  of  $n$  independent components  $s_1, s_2, \dots, s_n$ . Noting that the time index  $t$  is dropped in ICA model. Here, we assume each mixture  $x_j$  or each source  $s_k$  is a random variable.

Under such assumption,  $x_j(t)$  is a sample of the random variable  $x_j$ . Furthermore, we assume that all  $x_j$  and  $s_k$  are zero-mean(We can always preprocessing the mixtures to satisfy this requirement).

For convenience, we will use vector matrix notation from now on. All vectors are column vectors. Then the above model could be written as:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2.3}$$

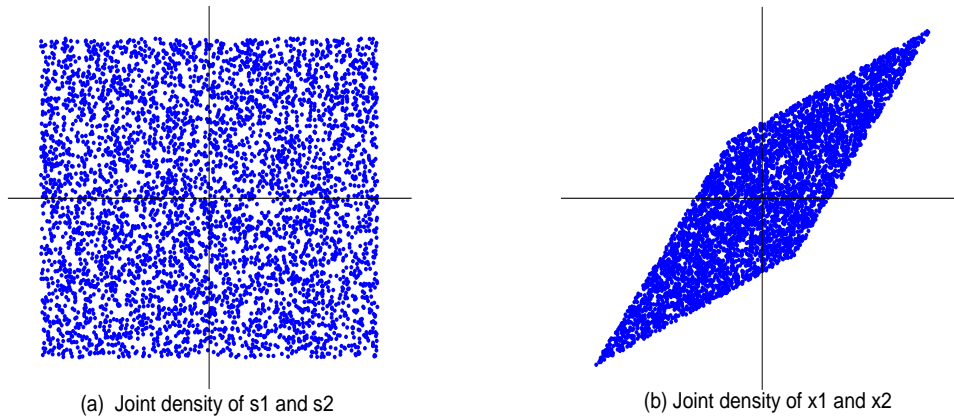


Figure 2.1: Joint pdf for sources and mixtures

Here,  $\mathbf{A}$  is the mixing matrix with elements  $a_{ij}$ ,  $\mathbf{x} = [x_1 x_2 \dots x_n]^t$  and  $\mathbf{s} = [s_1 s_2 \dots s_n]^t$ .

In ICA model, the independent components (or the sources) can not be directly observed, and the mixing matrix  $A$  is also assumed to be unknown. In another word, ICA estimates both  $\mathbf{s}$  and  $\mathbf{A}$  only when the mixture  $\mathbf{x}$  are given. Such a problem must be done under as general assumptions as possible.

## 2.3 Illustration of ICA

Consider the cock-tail party problem, if we assume the sources  $s_i$  have the following uniform distribution:

$$p(s_i) = \frac{1}{2\sqrt{3}} \quad \text{if } |s_i| \leq \sqrt{3} \quad (2.4)$$

Such distribution could guarantee the zero-mean and unit variance as was assumed in the section 2.2. Since the joint density of two independent components are the product of their marginal density, the square in Figure.2.1(a) shows the joint density of  $s_1$  and  $s_2$ .

Now let's mix  $s_1$  and  $s_2$  using the following mixing matrix:

$$\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$

Then we can get the two mixtures  $x_1$  and  $x_2$  and also their joint density (Figure 2.1(b) is their joint density). Clearly, the random variables  $x_1$  and  $x_2$  are not independent any more.

The problem of ICA is now to estimate the mixing matrix  $A$  when only information for  $x_1$  and  $x_2$  are available. Actually, an intuitive way to estimate  $A$  is to compute the edges of the parallelogram in Figure 2.1(b). This implies that we could estimate the ICA model by first estimating the joint density of the mixtures, and then locating the edges.

Here, one point should be noted is for the gaussian variables. Since the joint density of two gaussian variables are symmetric, no information could be obtained from locating the edges. Therefore,  $A$  could not be estimated by ICA for gaussian variables. More rigorously, for two gaussian independent components  $(s_1, s_2)$ , the distribution of any orthogonal transformation of  $(s_1, s_2)$  has exactly the same distribution of  $(s_1, s_2)$ . Therefore, for gaussian variables, the matrix  $A$  is not identifiable for gaussian independent components.

So now, it seems there is a solution for ICA model for variables except the gaussian case. However, in reality, such method only works with variables which have uniform distribution, and even for these variables, the computation could be very complicated. Some practical approaches to ICA model will be given in later sections.

## 2.4 Independence

The main concept for Independent Component Analysis is statistical independence.

Basically, independence between two different scalar random variables  $x$  and  $y$  means that information on the value of  $x$  does not give any information on the value of  $y$  and vice versa.

Technically, it is defined by the probability densities:

*Definition:* Denote the joint density of two random variables  $x$  and  $y$  as  $p_{xy}(x, y)$ , then the marginal density functions are:

$$p_x = \int p_{xy}(x, y) dy \quad (2.5)$$

$$p_y = \int p_{xy}(x, y) dx \quad (2.6)$$

$x$  and  $y$  are said to be independent if the following relation holds:

$$p_{xy}(x, y) = p_x(x)p_y(y) \quad (2.7)$$

In other words, if the joint density of the two variables is the product of their marginal densities, the two variables are called independent.

Independent random variables satisfy the basic property:

$$E\{g(x)h(y)\} = E\{g(x)\}E\{h(y)\} \quad (2.8)$$

Here,  $g(x)$  and  $h(y)$  are any absolutely integrable functions of  $x$  and  $y$ .

Uncorrelation between  $x$  and  $y$  means

$$E\{xy\} = E\{x\}E\{y\} \quad (2.9)$$

Let  $g(x) = x$  and  $h(y) = y$  in Eq.2.8, we could obtain Eq.2.9. Therefore, statistical independence is a much stronger property than uncorrelatedness.

Independent variables must be uncorrelated, but uncorrelated variables are not necessarily independent. For this reason, many ICA methods constrain the estimation procedure so that it always gives uncorrelated estimates of the independent components. This could help to reduce the number of free parameters and simplify the problem.

## 2.5 Information theory background

### 2.5.1 Entropy

Entropy is a basic concept in information theory[10]. The entropy of a random variable can be interpreted as the degree of randomness. The more “random”, i.e. the more unpredictable and unstructured the variable is, the larger the entropy is.

For a discrete random variable  $Y$ , entropy  $H$  is defined as:

$$H(Y) = -\sum_i P(Y = a_i) \log P(Y = a_i) \quad (2.10)$$

$$= \sum_i g(P(Y = a_i)) \quad (2.11)$$

Where  $a_i$  is the possible value of  $Y$  and  $P(Y = a_i)$  is the probability of  $Y = a_i$  and  $g(p) = -p \log p$   $0 \leq p \leq 1$ .

For a continuous random vector  $\mathbf{y}$ , the entropy  $H(\mathbf{y})$  is often called differential entropy, it is defined as:

$$H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y} \quad (2.12)$$

$$= \int g(f(\mathbf{y})) d\mathbf{y} \quad (2.13)$$

Here,  $f(\mathbf{y})$  is the probability density function(pdf) of  $\mathbf{y}$  and  $g(p) = -p \log p$   $p \geq 0$ .

A fundamental result in information theory is: *a gaussian variable has the largest entropy among all other random variables of equal variance*, for a proof, see [10, 43].

This also indicates that entropy could be a measure of nongaussianity.

More rigorously, entropy could be connected with coding length of the random variables. Actually, under some simplified assumptions, entropy gives roughly the average minimum code length of the random variable.

### 2.5.2 Negentropy

Negentropy comes from the concept of entropy, it is defined as a slight modification version of entropy. Negentropy of a random variable  $y$  is:

$$J(y) = H(y_{gauss}) - H(y) \quad (2.14)$$

where  $H(y_{gauss})$  is the entropy of a gaussian random variable of the same covariance matrix as  $y$  and  $H(y)$  is the entropy of  $y$ . Thus, negentropy is always non-negative and it is zero if and only if  $y$  is gaussian. Negentropy is an important measure of nongaussianity. Since it is well justified by statistics, negentropy could be considered the optimal estimator of nongaussianity in some sense as far as statistical properties are concerned.

As above stated, negentropy is a principled measure of nongaussianity. However, since the integral involves the probability density, it is quite difficult to compute the differential entropy or negentropy. Even though the density may be estimated by basic density estimation methods such as kernel estimators, whether the simple approach would be correct depends heavily on the correct choice of the kernel parameters. Furthermore, it would also become computationally rather complicated. Therefore, in practice, some approximations have to be used for computing negentropy.

### 2.5.3 Mutual information

Mutual information is defined based on the concept of the entropy. Given  $m$  (scalar) random variables  $y_i, i = 1, 2, \dots, m$ , the mutual information between them are:

$$I(y_1, y_2, \dots, y_m) = H(\mathbf{y}_{gauss}) - H(\mathbf{y}) \quad (2.15)$$

where  $\mathbf{y} = [y_1, y_2, \dots, y_m]$ ,  $\mathbf{y}_{gauss}$  is a Gaussian random variable of the same covariance matrix as  $\mathbf{y}$ .



By using the interpretation of entropy as code length, mutual information indicates what code length reduction is obtained by coding the whole vector  $\mathbf{y}$  instead of the separate components  $y_i$ . Generally, better codes could be produced if coding the whole vector. However, if the components are independent, they give no information on each other, and consequently, coding the whole vector will give the same length as coding its components individually.

## 2.6 Approach to ICA with data model assumption

One popular way of formulating the ICA problem is to consider the estimation of the following generative model for the data([1, 2, 4, 7, 19, 20, 27, 28, 41]).

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (2.16)$$

where  $\mathbf{x}$  is an observed  $m$ -dimensional vector,  $\mathbf{s}$  is an  $n$ -dimensional random vector whose components are assumed mutually independent, and  $\mathbf{A}$  is a constant  $m \times n$  matrix to be estimated. The matrix  $\mathbf{W}$  defining the transformation as in

$$\mathbf{s} = \mathbf{W}\mathbf{x} \quad (2.17)$$

is obtained as the (pseudo) inverse of the estimate of the matrix  $\mathbf{A}$ .

### 2.6.1 Nongaussianity for ICA model

*“Nongaussian is Independence[24]:”* Let  $\mathbf{y} = \mathbf{w}^t\mathbf{x}$ ,  $\mathbf{x}$  is the mixture vector and  $\mathbf{w}$  is a vector to be determined. (For simplicity, we assume in this section that all the independent components have identical distribution). If  $\mathbf{w}$  were one of the rows

of  $\mathbf{A}^{-1}$ , then the linear combination  $\mathbf{y}$  should be equal to one of the independent components.

Define  $\mathbf{z} = \mathbf{A}^T \mathbf{w}$ , then  $\mathbf{y} = \mathbf{w}^t \mathbf{x} = \mathbf{w}^t \mathbf{A} \mathbf{s} = \mathbf{z}^t \mathbf{s}$ . Now we can see that  $\mathbf{y}$  is a linear combination of  $s_i$ . From the Central Limit Theorem, we know the distribution of a sum of independent random variables are more Gaussian than any of the original random variable. Thus,  $\mathbf{y}$  is least gaussian when it in fact equals to one of the  $s_i$ . Here, obviously only one of the elements  $z_i$  of  $\mathbf{z}$  is nonzero (Note that  $s_i$  were assumed to be i.i.d).

Therefore,  $\mathbf{w}$  can be determined by maximizing the nongaussianity of  $\mathbf{w}^t \mathbf{x}$ . After that, a vector with only one nonzero component could be obtained, that is,  $\mathbf{w}^t \mathbf{x} = \mathbf{z}^t \mathbf{s}$  is one of the independent component.

Actually, since there are  $2n$  local maximum during optimizing for nongaussianity in the  $n$ -dimensional space of vector  $\mathbf{w}$ ,  $s_i$  and  $-s_i$  for one independent component  $s_i$ . Considering the uncorrelation between the different independent components, it is not difficult to find all the sources. Therefore, nongaussianity of the independent components is necessary for the identifiability of the model.

## 2.6.2 Measures of Nongaussianity

*Kurtosis*

*Kurtosis* is the classical measure of nongaussianity, it is defined as:

$$kurt(y) = E(y^4) - 3(E(y^2))^2 \quad (2.18)$$

$$= E(y^4) - 3 \quad \text{because } y \text{ is unit variance} \quad (2.19)$$

If  $y$  is a gaussian variable, then  $E(y^4) = 3(E(y^2))^2$ , and thus  $kurt(y) = 0$ . For most (not all) nongaussian random variables, kurtosis is nonzero, either positive or negative. Variables with positive kurtosis have typically “spiky” probability density

function(*pdf*) and they are called supergaussian. Those with a negative kurtosis are called subgaussian whose distributions are more “uniform” than that of gaussian variables.

Usually, the absolute value or the square value of kurtosis are used to measure the nongaussianity. Thus, the kurtosis is zero for a gaussian variable and greater than zero for most nongaussianity random variables. (There are still some other random variables with zero kurtosis, but they are quite rare).

*kurtosis* has two main characteristics:

1. kurtosis could be estimated by simply calculating the fourth moment of the sample data.
2. kurtosis has the linearity property, that is: if  $x_1$  and  $x_2$  are two independent random variables,

$$kurt(x_1 + x_2) = kurt(x_1) + kurt(x_2) \quad (2.20)$$

$$kurt(\alpha x_1) = \alpha kurt(x_1) \quad (2.21)$$

Such properties make *kurtosis* easy to use for its computational and theoretical simplicity, and thus become a popular measure of nongaussianity.

Even though *kurtosis* gives a simple ICA estimation, it is very sensitive to the outliers since it has to be estimated from a measured sample, and thus the value of kurtosis may depend heavily on few observations. That means kurtosis is not a robust measure of nongaussianity.

## Negentropy

As we have stated in section 2.5.1 that *a gaussian variable has the largest entropy*[34] among all random variables with equal variance. This means that the gaussian distribution is the “most random” or the least structured of all distributions. Entropy is small for distributions that are clearly concentrated on certain values, i.e., when the variable is clearly clustered, or has a pdf that is very “spiky” and entropy is large when the pdf is “uniform”.

Negentropy is a slightly modified version of entropy. Negentropy is zero for a gaussian variable and always nonnegative, thus, it can be a measure of nongaussianity and is the optimal measure of nongaussianity as far as the statistical performance is concerned. Negentropy is defined in Eq.2.14.

However, as we have stated in section 2.5.2, the problem of negentropy is its computational complexity. Methods to approximate negentropy is necessary for practical use. Many methods have been proposed to approximate. Among them, the classical approximating method is using higher-order cumulants[26], this gives the approximation:

$$J(y) \approx \frac{1}{12}E\{y^3\}^2 + \frac{1}{48}kurt(y)^2 \quad (2.22)$$

The random variable  $y$  is assumed to be zero-mean and unit variance. Actually, when the random variables have approximately symmetric distributions (this is often the case),  $E\{y^3\} = 0$  and then  $J(y) \approx \frac{1}{48}kurt(y^2)$ . This indicates that such approximation will often leads to the use of *kurtosis*.

## Conclusion

Usually, *kurtosis* and *negentropy* are thought to be two important measures of non-gaussianity. From the above analysis, *Kurtosis* is in fact an approximation form of *negentropy*. In practice, many other approximations of negentropy instead of kurtosis have been proposed. In section 2.9, we will give another important, more generative and practical approximate form of negentropy for measuring the non-gaussianity.

## 2.7 Approach to ICA without data model assumption

Comon [9] showed how to obtain a more general formulation for ICA that does not need to assume an underlying data model. This definition is based on the concept of mutual information.

As defined in last section, the differential entropy of a random vector  $\mathbf{y} = (y_1, \dots, y_n)^T$  with density  $f(\cdot)$  is Eq.2.12. The negentropy is given in Eq.2.14 and Eq.2.15 is the mutual information  $I$  between the  $n$ (scalar) random variables  $y_i, i = 1, 2, \dots, n$  [9, 10].

If we constrain the variables to be uncorrelated, the mutual information could be expressed as following[9]:

$$I(y_1, y_2, \dots, y_n) = J(y) - \sum_i J(y_i) \quad (2.23)$$

AS the information-theoretic measure of independence of random variables, mutual information could be used as the criterion for finding the ICA transform. Therefore, the ICA of a random vector  $\mathbf{x}$  as an invertible transformation  $\mathbf{s} = \mathbf{W}\mathbf{x}$  where

the matrix  $\mathbf{W}$  is determined so that the mutual information of the transformed components  $s_i$  is minimized.

Because negentropy is invariant for invertible linear transformations[9], it is obvious from Eq.2.23 that finding an invertible transformation  $W$  that minimizes the mutual information is roughly equivalent to finding directions in which the negentropy is maximized.

Therefore, the two approaches to ICA is equivalence to each other and negentropy is their common contrast function.

## 2.8 Other approaches to ICA

Besides the two main approaches to ICA, Maximum Likelihood estimation[40] and the Infomax principle[2, 39] are always used as another two approaches. Even though all of the approaches seem to be different in the notations, several authors have demonstrated that these approaches could be equivalent under some conditions for the parameter functions. For details, see [8, 44].

## 2.9 Practical Contrast Functions

There are several contrast functions for ICA models based on the different approaches, such as the *kurtosis*, *negentropy*, *maximum likelihood*, *mutual information* and *infomax*(maximum of the output entropy) and etc.. However, as we have analyzed above, *kurtosis* is one form of *negentropy*, approaches of maximum likelihood and infomax principle are equivalent to mutual information estimation which uses negentropy as the contrast function. So here, we will focus on the practical negentropy contrast function.

Usually, the computational complexity makes the negentropy impossible to use

without approximation. There have been many methods to approximate the negentropy. Here, we will introduce one class of new approximations developed in [21]. In [21] it was shown that these approximations are often considerably more accurate than the conventional, cumulant-based approximations in [1, 9, 26]. In the simplest case, these new approximations are of the form:

$$J(y_i) \approx c[E\{G(y_i)\} - E\{G(v)\}]^2 \quad (2.24)$$

Where  $G$  is practically any nonquadratic function,  $c$  is an irrelevant constant, and  $v$  is a Gaussian variable of zero mean and unit variance (i.e. standardized). The random variable  $y_i$  is assumed to be of zero mean and unit variance. For symmetric variables, this is a generalization of the cumulant-based approximation in [9], which is obtained by taking  $G(y_i) = y_i^4$ .

The approximation of negentropy given above gives readily a new objective function for estimating the ICA transform. First, to find one independent component, or projection pursuit direction as  $y_i = \mathbf{w}^t \mathbf{x}$ , we maximize the function  $J_G$  given by

$$J_G(w) = [E\{G(w^t x)\} - E\{G(v)\}]^2 \quad (2.25)$$

for practically any nonquadratic function  $G$ . Here  $\mathbf{w}$  is an  $m$ -dimensional vector constrained so that  $E\{(\mathbf{w}^t \mathbf{x})^2\} = \mathbf{1}$  (we can fix the scale arbitrarily). Several independent components can then be estimated one-by-one.

If the function  $G$  could be wisely chosen, such approximations in Eq.2.25 would be better than the higher-order cumulants approximation given in Eq.2.22. Especially when choosing a  $G$  that does not grow too fast, a robust estimator could be expected. The following choices of  $G$  have proved very useful:

$$G_1(y) = \frac{1}{a_1} \log \cosh a_1 y \quad (2.26)$$

$$G_2(y) = -\exp(-y^2/2) \quad (2.27)$$

where  $1 \leq a_1 \leq 2$  is some suitable constant, often taken equal to one.

## 2.10 Conclusion

ICA is a very general-purpose statistical technique. In ICA, the observed random data(mixtures) are linearly transformed into sources which are as independent as possible from each other. The intuitive way to estimate ICA model is to maximize nongaussianity, and furthermore, different ways which are approximately equivalent could also be derived. Finally, a class of negentropy approximations are given for practical use.

When using ICA for single-channel fetal ECG extraction, we have two problems:

1. Since ICA requires the number of the mixtures can not be less than the number of the sources, which, in our case, only one mixture available for obtaining at least three sources(maternal ECG, fetal ECG and noise).
2. Another problem is that ICA gives random components and we could not know which component is the one for maternal ECG, fetal ECG or for noise.

In later chapters, we will give the algorithm and our novel method which could provide a good way to solve these problems and leads to a promising extraction.



# FastICA—an algorithm for ICA

## 3.1 Introduction

The current algorithms for ICA can be roughly divided into two categories. In the first category (Cardoso, 1992; Comon, 1994), the algorithms rely on batch computations minimizing or maximizing those contrast functions. The requirement of very complex matrix or tensorial operations of these algorithms makes this kind of algorithm difficult to implement. The second category contains adaptive algorithms often based on stochastic gradient methods, which may have implementations in neural networks (Amari et al., 1996; Bell and Sejnowski, 1995; Delfosse and Loubaton, 1995; Hyvärinen and Oja, 1996; Jutten and Herault, 1991; Moreau and Machi, 1993; Oja and Karhunen, 1995). The main problem with this category is the convergence which is very slow and crucially dependent on the correct choice of the learning rate parameters. A bad choice of the learning rate can, in practice, destroy convergence. Therefore, it would be important in practice to make the learning faster and more reliable. This can be achieved by the following algorithm—FastICA [17, 18, 19].

FastICA uses the fixed-point iteration scheme and it is very simple but highly efficient in finding the extrema for ICA. Meanwhile, the fixed-point algorithms have

very appealing convergence properties which make them a very interesting alternative to adaptive learning rules.

In this thesis, FastICA was used for our ICA model. The following is a detailed discussion for this algorithm.

## 3.2 Fixed-point algorithm for one unit

To begin with, we firstly show the one-unit version of FastICA. A “unit” is referred to a computational unit, eventually an artificial neuron which has a weight vector  $\mathbf{w}$  that the neuron is able to update by a learning rule. FastICA learning rule finds a direction, i.e. a unit vector  $\mathbf{w}$  such that the projection  $\mathbf{w}^T \mathbf{x}$  maximizes nongaussianity or minimizing the mutual information. Here we used the approximation of negentropy we introduced in Eq.2.25 as the contrast function. The variance of  $\mathbf{w}^T \mathbf{x}$  must here be constrained to unity; for whitened data this is equivalent to constraining the norm of  $\mathbf{w}$  to be unity.

The derivations of FastICA is as follows: first note that the maxima of the approximation of the negentropy  $\mathbf{w}^T \mathbf{x}$  are obtained at certain optima of  $\mathbf{E}\{\mathbf{G}(\mathbf{w}^T \mathbf{x})\}$ . According to the Kuhn-Tucker conditions[36], the optima of  $\mathbf{E}\{\mathbf{G}(\mathbf{w}^T \mathbf{x})\}$  under the constraint  $\mathbf{E}\{\mathbf{G}(\mathbf{w}^T \mathbf{x})^2\} = \|\mathbf{w}\|^2 = 1$  are obtained at points where

$$\mathbf{E}\{\mathbf{xg}(\mathbf{w}^T \mathbf{x})\} - \beta \mathbf{w} = 0 \quad (3.1)$$

Solve this equation by Newton’s method. The Jacobian matrix of the above equation is:

$$\mathbf{JF}(\mathbf{w}) = \mathbf{E}\{\mathbf{xx}^T \mathbf{g}'(\mathbf{w}^T \mathbf{x})\} - \beta \mathbf{I} \quad (3.2)$$

To simplify the inversion of this matrix, the first term is approximated in the following. Since the data is sphered, a reasonable approximation seems to be:

$\mathbf{E}\{\mathbf{x}\mathbf{x}^T\mathbf{g}'(\mathbf{w}^T\mathbf{x})\} \approx \mathbf{E}\{\mathbf{x}\mathbf{x}^T\}\mathbf{E}\{\mathbf{g}'(\mathbf{w}^T\mathbf{x})\} = \mathbf{E}\{\mathbf{g}'(\mathbf{w}^T\mathbf{x})\}\mathbf{I}$ . Thus, the jacobian matrix becomes diagonal, and can easily be inverted. Therefore, the following approximative Newton iteration is obtained:

$$\mathbf{w}^+ \leftarrow \mathbf{w} - [\mathbf{E}\{\mathbf{x}\mathbf{g}(\mathbf{w}^T\mathbf{x})\} - \beta\mathbf{w}]/[\mathbf{E}\{\mathbf{g}'(\mathbf{w}^T\mathbf{x})\} - \beta] \quad (3.3)$$

Multiplying both sides by  $\beta - \mathbf{E}\{\mathbf{g}'(\mathbf{w}^T\mathbf{x})\}$ , the following FastICA iteration could be obtained after algebraic simplification,

1. Choose an initial(e.g. random) weight vector  $\mathbf{w}$
2. Let  $\mathbf{w}^+ \leftarrow \mathbf{E}\{\mathbf{x}\mathbf{g}(\mathbf{w}^T\mathbf{x})\} - \mathbf{E}\{\mathbf{g}'(\mathbf{w}^T\mathbf{x})\}\mathbf{w}$
3. Let  $\mathbf{w} \leftarrow \mathbf{w}^+ / \|\mathbf{w}^+\|$
4. If not converged, go back to 2.

Note that convergence means that the old and new values of  $\mathbf{w}$  point in the same direction, i.e. their dot-product is (almost) equal to 1. It is not necessary that the vector converges to a single point, since  $-\mathbf{w}$  and  $\mathbf{w}$  define the same direction. This is again because the independent components can be defined only up to a multiplicative sign. Note also that it is here assumed that the data is prewhitened.

In practice, the expectations in FastICA must be replaced by their estimates. The natural estimates are the corresponding sample means.

### 3.3 FastICA for several units

The one-unit algorithm of the preceding subsection estimates just one of the independent components, or one projection pursuit direction. To estimate several independent components, it is necessary to run the one-unit FastICA algorithm using several units(e.g. neurons) with weight vectors  $\mathbf{w}_1, \dots, \mathbf{w}_n$ .

One problem here is to avoid different vectors from converging to the same maxima. Therefore, decorrelation should be done on the outputs  $\mathbf{w}_1^T \mathbf{x}, \dots, \mathbf{w}_n^T \mathbf{x}$  after every iteration. Usually three methods are widely used for achieving this.

The simple way is the deflation scheme based on a Gram-Schmidt-like decorrelation. This means that the independent components is estimated one by one. When  $p$  independent components have been estimated, or  $p$  vectors  $\mathbf{w}_1, \dots, \mathbf{w}_p$  are known, run the one-unit fixed-point algorithm for  $\mathbf{w}_{p+1}$ , and after every iteration step subtract from  $\mathbf{w}_{p+1}$  the “projections”  $\mathbf{w}_{p+1}^T \mathbf{w}_j \mathbf{w}_j$ ,  $j = 1, \dots, p$  of the previously estimated  $p$  vectors, and then renormalize  $\mathbf{w}_{p+1}$ :

$$\mathbf{w}_{p+1} \Leftarrow \mathbf{w}_{p+1} - \sum_{j=1}^p \mathbf{w}_{p+1}^T \mathbf{w}_j \mathbf{w}_j \quad (3.4)$$

$$\mathbf{w}_{p+1} \Leftarrow \mathbf{w}_{p+1} / \sqrt{\mathbf{w}_{p+1}^T \mathbf{w}_{p+1}} \quad (3.5)$$

Another two methods are all used for certain applications where a symmetric decorrelation may be desired. In such cases, no vectors are “privileged” over others. For details, see [17, 28].

## 3.4 FastICA algorithm

The main steps of FastICA includes:

1. *Preprocessing*:
  - (a) Center the data matrix by subtracting the mean of each column of the data matrix.
  - (b) Whiten the data matrix by projecting the data onto its principle component directions.
2. *Algorithms*

- 
- (a)  $i \leftarrow 0$ ;
  - (b)  $i \leftarrow i + 1$ ; if  $i > n$ , stop.
  - (c) Choose initial weight vectors:  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$
  - (d) Let  $\mathbf{w}_i^+ \leftarrow E\{\mathbf{x}g(\mathbf{w}_i^T \mathbf{x})\} - E\{g'(\mathbf{w}_i^T \mathbf{x})\}\mathbf{w}_i$
  - (e) Let  $\mathbf{w}_i^+ \leftarrow \mathbf{w}_i^+ / \|\mathbf{w}_i^+\|$
  - (f) If not converged, go back to 2c.
  - (g) Let  $\mathbf{w}_i \leftarrow \mathbf{w}_i - \sum_{j=1}^i \mathbf{w}_i^T \mathbf{w}_j \mathbf{w}_j$
  - (h) Let  $\mathbf{w}_i \leftarrow \mathbf{w}_i / \sqrt{\mathbf{w}_i^T \mathbf{w}_i}$
  - (i) if  $i < n$ , go to 2b.

In FastICA, if we select the derivative  $g$  as the fourth power as in kurtosis, it will lead to the method for maximizing kurtosis by fixed-point algorithm, while if the nonquadratic function  $G$  used Eq.2.26 and Eq.2.27, FastICA will give robust approximations of negentropy.

Note, the derivatives of the nonquadratic functions in Eq.2.26 and Eq.2.27 are:

$$g_1(u) = \tanh(a_1 u) \quad (3.6)$$

$$g_2(u) = u \exp(-u^2/2) \quad (3.7)$$

FastICA algorithm was derived for optimization of  $E\{G(\mathbf{w}^T \mathbf{z})\}$  under the constraint of the unit norm of  $\mathbf{w}$ . FastICA also works for maximum likelihood estimation. Actually, if the estimates of the independent components are constrained to be white, maximization of likelihood gives an almost identical optimization problem. See[22]

## 3.5 Conclusion

Compared to the stochastic gradient descent methods, FastICA has the following properties[23]:

1. FastICA has a very fast convergence which is at least quadratic.
2. Since no step-size parameters are needed, FastICA is very easy to use.
3. FastICA could estimate the independent components one by one, this makes FastICA quite useful in exploratory data analysis and decreases the computational load of the method.
4. Performance of FastICA could be optimized by choosing a suitable nonlinearity function  $g$ , especially when concerning the robust and/or the minimum variance of the algorithm. Actually, the two nonlinearities  $\mathbf{G}$  in Eq.3.6 and 3.7 have some optimal properties.

Such properties make FastICA a very popular algorithm for ICA model. In this thesis, FastICA is the algorithm we used and it proves to be very efficient.

# Fetal ECG extraction

## 4.1 Introduction

In this work[15], we are given a single-channel abdominal ECG and we are expected to extract the fetal ECG from this mixture. Like the adults, among all the information from fetal ECG, the fetal ECG complex and the heart rate variability are two important measures.

In our case, each given signal is about 10-minute long, with a sampling rate 300HZ(roughly  $1.8 \times 10^5$  samples. Figure.4.1 shows one whole signal. For clarity, Figure.4.2 gives a half-minute part of Figure.4.1. In the figures, the prominent repeating peaks are the maternal R-wave(the peak of the ECG complex), while the less visible peaks are from the fetus.

Our aim is to detect the fetal heart rate and extract the fetal ECG complex. In this chapter, we will introduce our approach to this two aspects. The main challenge is the detection of the occurrence of fetal heart beats, then it is trivial to find the ‘beat-to-beat’ heart rate. In the mean time, once the locations of the fetal heart beats are detected, the fetal ECG complex could be obtained by averaging, SVD or ICA.

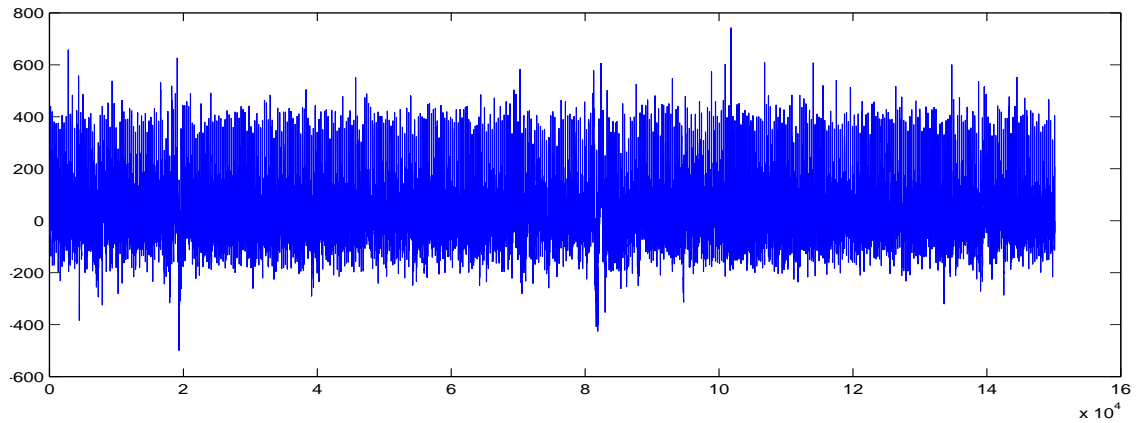


Figure 4.1: The whole original signal

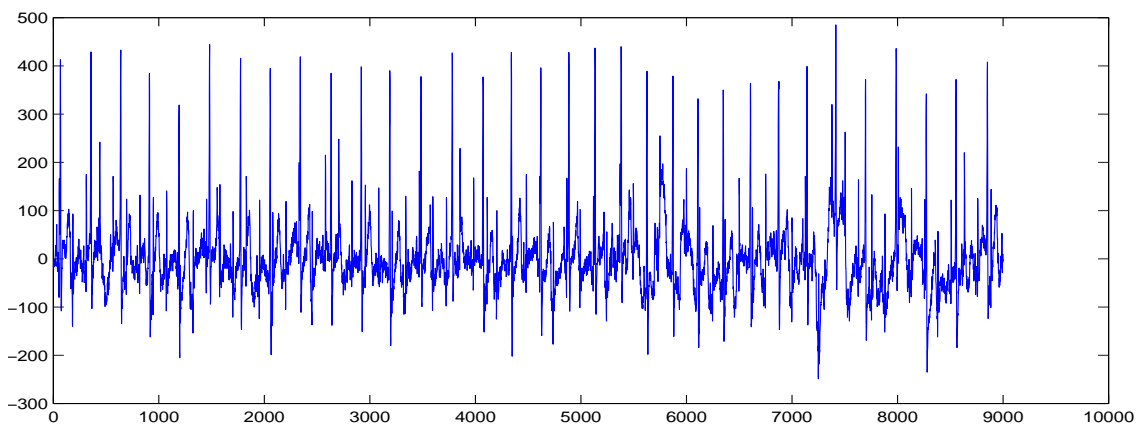


Figure 4.2: Detail of the original signal

For fetal heart beats detection, we propose a blind-source separation method using a SVD of the spectrogram, which is followed by an iterative application of ICA on both the spectral and a temporal representations of the ECG signals. This proposed method could give us a heart beats trend which is a sinusoidal with each cycle corresponding to a heart beat. Using this sinusoidal, the heart beats could be located by simple search routines. Next, time domain averaging is employed to compute the fetal ECG complex.

This chapter includes three main parts: the first part is on the heart beats



occurrence detection, the second part mainly focus on how to compute the fetal ECG complex, and the last section introduces two refining method which is used in the proposed method.

## 4.2 Heart beats occurrence detection

### 4.2.1 Motivation

Consider a series  $X$ , rearrange it into a matrix  $\mathbf{B}$  such as

$$\mathbf{B} = \begin{bmatrix} x(1) & x(2) & \dots & x(L) \\ x(L-m+1) & x(L-m+2) & \dots & x(2L-m) \\ \vdots & \vdots & \vdots & \vdots \\ x(nL-nm+1)x(nL-nm+2)\dots x(nL-(n-1)m) \end{bmatrix}$$

where  $L$  is the segment length and  $L - m$  is the overlap.  $\mathbf{B}$  is an  $n \times L$  matrix. The SVD on  $\mathbf{A}$  is given by  $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  where  $\mathbf{U}$  and  $\mathbf{V}$  are  $n \times n, L \times L$  matrix respectively.  $\mathbf{\Sigma}$  is a diagonal matrix and  $\mathbf{\Sigma} = [\mathbf{diag}(\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_r), \mathbf{0}]$ ,  $r$  is the number of non-zero elements in  $\mathbf{\Sigma}$  and  $\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_r$  are the non-zero singular values.

When  $\mathbf{x}$  is a strictly periodic series,  $L$  is the periodic length and  $m = 0$  (that means the rows are identical), only  $\sigma_1$  is non-zero, and  $\mathbf{B} = \mathbf{u}_1 * \sigma_1 * \mathbf{v}_1^T$  where  $\mathbf{u}_1$  and  $\mathbf{v}_1$  are the first column of  $\mathbf{U}$  and  $\mathbf{V}$  respectively, here,  $\mathbf{B}$  is one-rank matrix. The information energy will be concentrated in the unique dyad  $\mathbf{u}_1 * \sigma_1 * \mathbf{v}_1^T$ .

When  $\mathbf{x}$  is a nearly periodic series,  $L$  is the average period length and  $m = 0$ , then even though  $r$  will be bigger than 1, but  $\frac{\sigma_1}{\sigma_2} \gg 1$ , the most dominant information energy will still be concentrated in the dyad  $\mathbf{u}_1 * \sigma * \mathbf{v}_1^T$ . The most dominant periodic component present in the series  $\mathbf{x}$  is given by  $\mathbf{B}_{d1} = \mathbf{u}_1 * \sigma * \mathbf{v}_1^T$ . The time series of

$\mathbf{B}_{d1}$  will have the same repeating pattern given by  $\mathbf{v}_1^T$  up to a scaling factor  $\mathbf{u}_{1j}\sigma_1$  where  $\mathbf{u}_{1j}$  is the  $j^{th}$  element of  $\mathbf{u}_1$ .

When  $\mathbf{x}$  is a random series, no matter what is  $L$  and  $m$ ,  $\mathbf{B}$  will be a full rank matrix and all the singular values will be almost the same, and all the information energy is distributed uniformly among all the singular values.

Now we define a matrix  $\mathbf{S}$  based on the fourier transform of  $\mathbf{B}$ :

$$S = \text{fft}^2(Bw) \quad w \text{ is a window function with length } L. \quad (4.1)$$

When  $\mathbf{x}$  is a random series and  $\mathbf{B}$  is a full-rank matrix, let  $m = 1$ , then the consecutive rows of  $\mathbf{S}$  will be a little different, since the overlap  $L - 2$  elements are same before transformation(Here, we assume  $L \gg 1$ ). It is reasonable to expect any two consecutive rows are almost identical when we use the window:  $w = \text{blackman}(L)$  for the weight is nearly zero for the first element and the last element which are the different elements between the two rows.

Since  $\mathbf{S}$  has repetitive frequency patterns between consecutive rows, that means we have transformed the random signal into a matrix which has certain basic patterns.

For any source signal  $\mathbf{x}$ , with a large enough overlap, its spectrogram could actually serve as  $\mathbf{S}$ . Therefore,  $\mathbf{S}$  is a matrix with each row corresponding to the spectrum at a particular time(Figure.4.3(a)).

Consider a signal which consists of a repeating ECG complex. Its spectrogram also consists of repeating patterns(in this case, the overlap length could be decreased since the ECG is a nearly periodic signal). This can be seen in Figure.4.3(a)(Here,  $m$  is 10,  $L$  is 301).

Therefore, the problem now becomes how to find the pattern and how the pattern changes along with time.

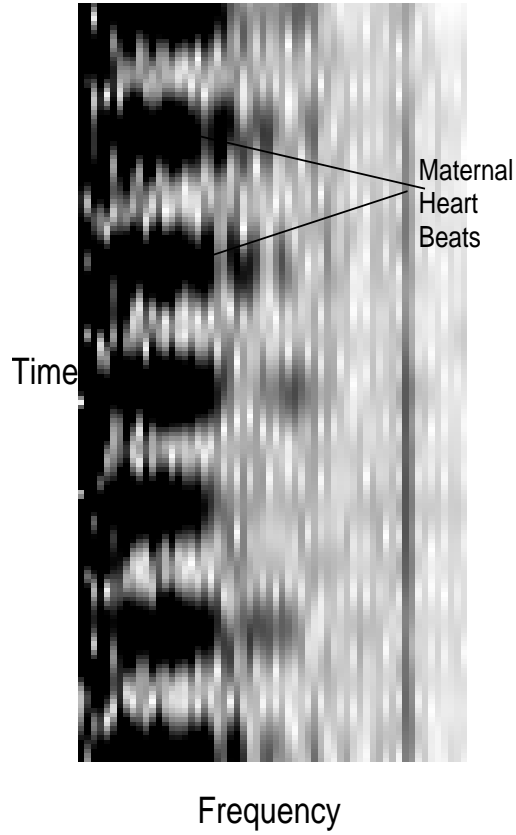


Figure 4.3: Spectrogram of the original signal(108.raw).

### 4.2.2 Problem formulation

We assume that  $\mathbf{S}$  is the mixture of the column vector  $\mathbf{u}_m, \mathbf{v}_m$  and  $\mathbf{u}_f, \mathbf{v}_f$  in the following way,

$$S = \mathbf{u}_m \mathbf{v}_m^T + \mathbf{u}_f \mathbf{v}_f^T + \mathbf{n}, \quad (4.2)$$

where  $\mathbf{n}$  is the noise. We call the vector  $\mathbf{u}_m$  and  $\mathbf{u}_f$  the maternal and fetal *heartbeat trend* respectively.

Consider a signal which consists of a repeating ECG complex. Its spectrogram also consists of repeating patterns. This can be seen in Figure.4.3(a). By carefully choosing the right window width for the spectrogram, the spectrogram of a ECG

complex could be separable. In this case, we would expect the heartbeat trends  $\mathbf{u}_m$  and  $\mathbf{u}_f$  to be approximately sinusoidal with each cycle corresponding to a heart beat, and expect  $\mathbf{v}_m$  and  $\mathbf{v}_f$  to approximate the spectrum of the ECG complex. Therefore, an accurate estimation of  $\mathbf{u}_f$  is sufficient to determine the heartbeat, which in-turn can be used to obtain the ECG complex.

Now, given  $\mathbf{S}$ , our problem is to estimate  $\mathbf{u}_m, \mathbf{v}_m, \mathbf{u}_f$ , and  $\mathbf{v}_f$ . If we attempt to minimize the energy of  $\mathbf{n}$ , then this amounts to finding the two best separable functions whose sum approximates  $\mathbf{S}$ , which can be obtained using SVD. However, numerical experiment on the synthetic signal (Figure.5.5) gives disappointing results.

Alternatively, we can borrow the idea of ICA. Besides minimizing the noise, we propose finding the components such that  $\mathbf{u}_m$  and  $\mathbf{v}_m$  are respectively statistically independent from  $\mathbf{u}_f$  and  $\mathbf{v}_f$ . In next section, we describe a method that attempts to find such components.

### 4.2.3 Proposed method for finding trends of original signal

Given the source signal  $\mathbf{x}$ , we first compute its spectrogram,  $\mathbf{S}$  (the choice of window width will be discussed in Section 4.4.1).

1. Perform SVD on  $S$ . Let  $S = U\Sigma V^T$ .

Here,  $\mathbf{S}$  is the spectrogram with rows representing time slices.  $\Sigma$  is a square diagonal matrix with weights corresponding to the significance of the related spectral vector in  $\mathbf{V}$ ,  $\mathbf{U}$  is oriented the same way as the spectrogram with columns that are orthonormal time-indexed weights associated with a given spectral vector from  $\mathbf{V}$  which sum to create spectral slices of  $\mathbf{S}$ .

2. Based on the property of SVD, the first  $k$  columns of  $\mathbf{U}$ , and  $\mathbf{V}$  are the  $k$  most significant components,  $\mathbf{S}$  then could be written as:

$$S \approx U_k \Sigma_k V_k^T$$

where  $\Sigma_k$  is the diagonal matrix whose elements are the first  $k$  singular values of  $\mathbf{S}$ . Here  $k > 2$  is a fixed constant.

3. Apply ICA on the  $k$  most significant spectral components  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  (columns in  $\mathbf{V}_k$ ), the corresponding independent components are  $\mathbf{v}_1^0, \dots, \mathbf{v}_k^0$  (columns in  $\mathbf{V}_k^0$ ) and the “mixing” matrix for  $V_k$  is  $\mathbf{A}$ . That is:

$$V_k^T = \mathbf{A} V_k^{0T}$$

4. Update the time vectors to recover the one-to-one correspondence between  $\mathbf{U}$  time vectors and  $\mathbf{V}$  spectral vectors.

That is, compute  $[\mathbf{u}_1^0, \mathbf{u}_2^0, \dots, \mathbf{u}_k^0]$  by

$$[\mathbf{u}_1^0, \mathbf{u}_2^0, \dots, \mathbf{u}_k^0] = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k] \Sigma_k \mathbf{A}.$$

where  $\mathbf{A}$  is the same “mixing” matrix determined in the previous ICA step for  $\mathbf{V}_k$ , and  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k$  are columns of  $\mathbf{U}_k$ .

By doing so, the independence of  $\mathbf{v}_1^0, \dots, \mathbf{v}_k^0$  is guaranteed and the energy of  $S$  is kept constant, which are helpful for the solution stability.

5. Make the time vectors as independent as possible.

This is achieved by performing ICA on the  $\mathbf{u}_1^0, \mathbf{u}_2^0, \dots, \mathbf{u}_k^0$ . Let  $\mathbf{u}_1^1, \dots, \mathbf{u}_k^1$  be the independent components.

6. Select and output the two best components as  $u_m$  and  $u_f$  from  $\mathbf{u}_1^0, \mathbf{u}_2^0, \dots, \mathbf{u}_k^0$ . (see Section 4.4.2)

The above algorithm requires a parameter  $k$ , which we take it as 10 in our experiment. That is, we choose the 10 most significant components from the much larger set corresponding to the number of frequency channels in the spectrogram. The

number is chosen to be large enough to retain the significant information from the original signal, but is reduced for fast computation and so that we have a reasonable number of channels for the ICA algorithm to work on.

## 4.3 Fetal ECG complex detection

Fetal ECG complex is another important measure for clinical diagnosis. In this section, we will provide a way to compute fetal ECG complex when the maternal and fetal heart beats occurrences are known.

### 4.3.1 Main idea

In this thesis, we adopt the most straightforward way—subtraction to extract the fetal ECG and then the time domain averaging is employed to get the ECG complex.

Even though the direct subtraction of the maternal ECG (usually the thoracic ECG could be a reasonable assumption) from the mixture does not give a good result, this method is not necessarily useless. Actually, if the suitable maternal ECG is available, a pure fetal ECG could be expected by this simple method. Hence, our approach will mainly focus on finding the appropriate maternal ECG which could match the abdominal mixture as well as possible.

In order to generate a pure maternal ECG, aligning, correlation, shifting and scaling are all used in our method. Here, we will give a brief introduction to the concept of correlation.

Correlation: Cross-correlation between two real random process  $y$  and  $z$  is defined as:

$$R_{yz}(m) = E\{y_{n+m}z_n\} = E\{y_nz_{n-m}\} \quad (4.3)$$

When  $y$  is equal to  $z$ , the cross-correlation is also called the autocorrelation.

In practice, we often use:

$$\hat{R}_{yz}(m) = \begin{cases} \sum_{n=0}^{N-m-1} y_{n+m} z_n & m \geq 0 \\ \hat{R}_{zy}(-m) & m \leq 0 \end{cases}$$

When  $y = z$ ,  $R_{yz}(0) \geq R_{yz}(m)$  for any  $m \neq 0$ .

Signals  $y$  and  $z$  are said to be correlated if the shapes of the waveforms of the two signals match one another. Here, we define a correlation coefficient  $r$  between  $y$  and  $z$  as:

$$r = \frac{R_{yz}(m)}{R_{yy}(0)} \quad (4.4)$$

Such a ratio determines the degree of match between the shapes of  $y$  and  $z$ .

### 4.3.2 Proposed method for fetal ECG extraction

#### Method

1. Segment the original ECG signal such that each segment contains the maternal ECG complex.
2. Select the 'good' maternal ECG complex segments, average them to get the maternal ECG complex template.
3. Compare each segment in the original signal with the template, shift it if needed to make sure the location of ECG peak is the same as the template.
4. Compute all the correlation coefficients between the template and each segment. Then scale the segments by their correlation coefficients and construct a purely maternal ECG by connecting the segment-templates.
5. Subtract the purely maternal ECG from the original ECG to obtain the fetal ECG.

6. Segment the fetal ECG and average to get the fetal ECG complex.

Even though a large sampling rate may indicate high precision, it is impossible and unnecessary to adopt a very large sampling frequency. To overcome the mismatch between the template and the composite due to a relatively small sampling rate, shifting is adopted when aligning the template and the composite. Furthermore, the energy for each ECG complex wave may vary greatly, scaling could help to cancel out such influence. By carefully subtracting the maternal ECG which matches the mixture, a pure fetal ECG is then obtained.

### A simple example

For illustrating the above method, we give a simple example. Note that this is not a synthetic ECG signal, it is only one signal with quasiperiodic ‘peaks’. Time domain averaging only works when there are enough periods. For simplicity, we use “Large complex” and “Small Complex” to refer to the stronger and weaker patterns which is similar as “maternal ECG complex”, “fetal ECG complex”.

Figure.4.4 is the example signal.

1. Segment the signal: 1, 2, 3, 4, 5, 6, ... are the segments.
2. Average the 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup> and 6<sup>th</sup> segments(those are the ‘good’ ones) to obtain the maternal ECG complex template, that is Figure.4.5.
3. Compare Figure.4.5 with each segments in Figure.4.4. Do shifting for the segments which do not match the shape of the template(it often occurs due to not large enough sampling rate). After that, scaling the shifted segments based on their corresponding correlation coefficients with the template.

Figure.4.6 is an example for shifting the second segment.  $CVBA$  is the original segment part. Firstly, shift  $CVBA$  one pixel left to  $C'B'V'A'$ .  $V''$  is obtained



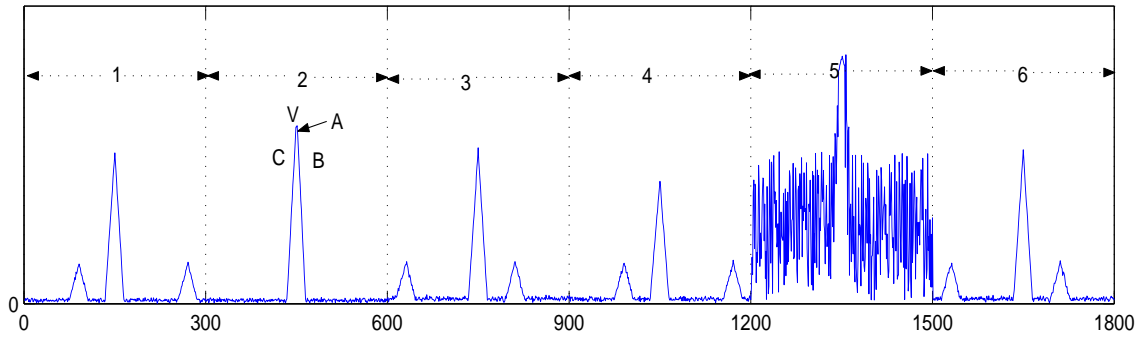


Figure 4.4: Original mixture and the segments

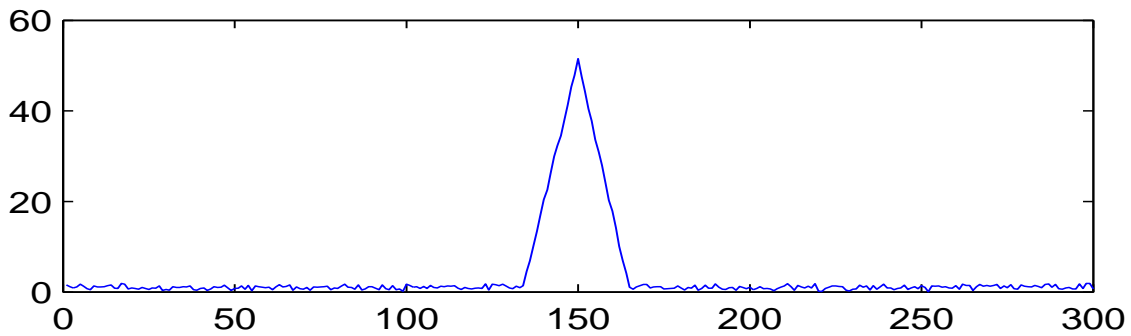


Figure 4.5: Large complex template

by extrapolation.  $V'''$  has the same magnitude as  $V''$ . Then  $C'V'V'''A'B'$  becomes the part of the segment-template. Scaling the  $C'V'V'''A'B'$  by its correlation coefficient with the template  $V_0C_0B_0$ . Note the above shifting and scaling is done on the whole second segment which including  $CVBA$  as a part.

4. Connect all the segment-templates to construct a purely maternal ECG signal.
5. Subtract the maternal ECG from the original mixture. Figure.4.8 is the result.
6. Average the fetal ECG complex segments in Figure.4.8, the fetal ECG complex could be obtained, that is Figure.4.9

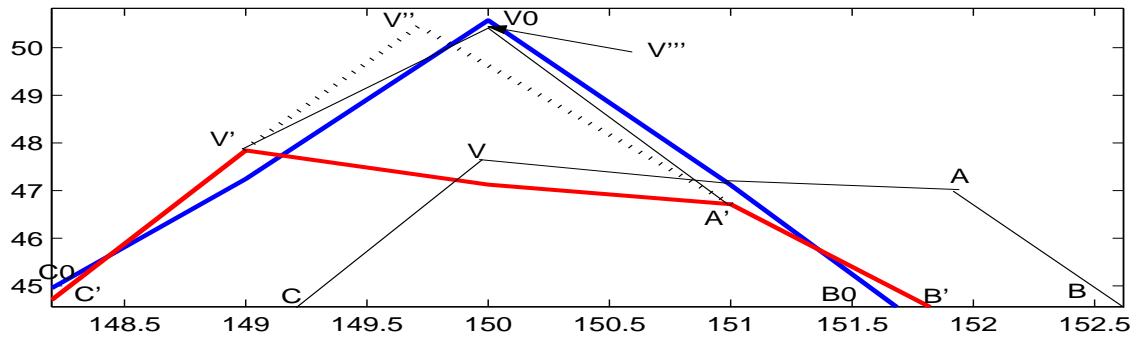


Figure 4.6: Shift procedure

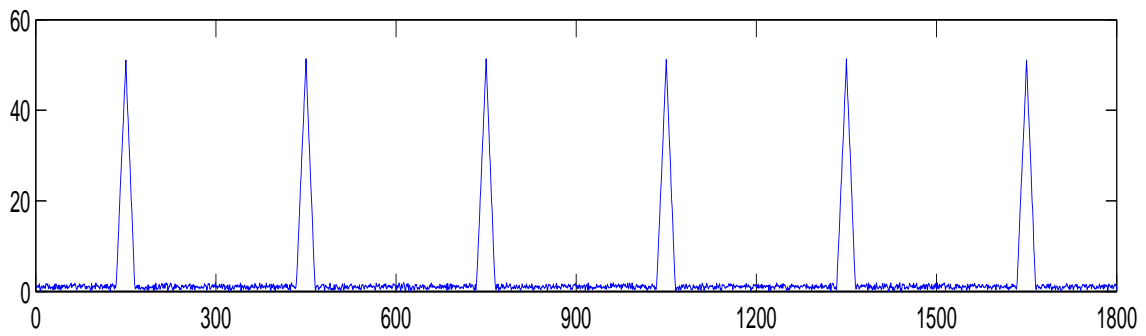


Figure 4.7: Purely large complex signal

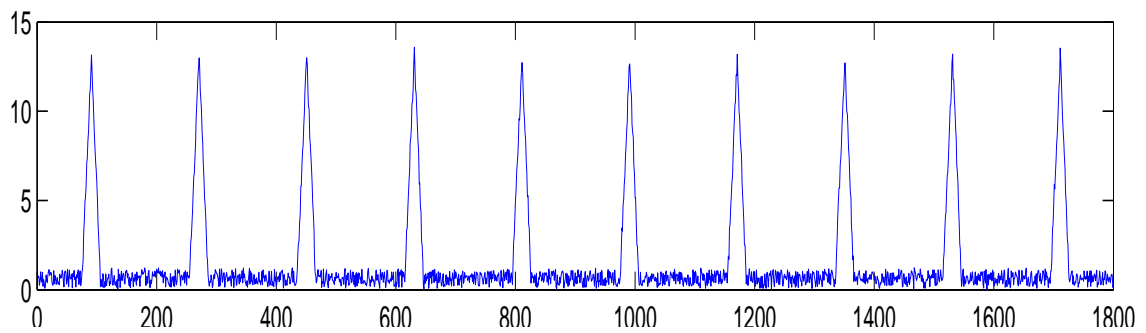


Figure 4.8: Small complex signal(after removing the large complex signal)

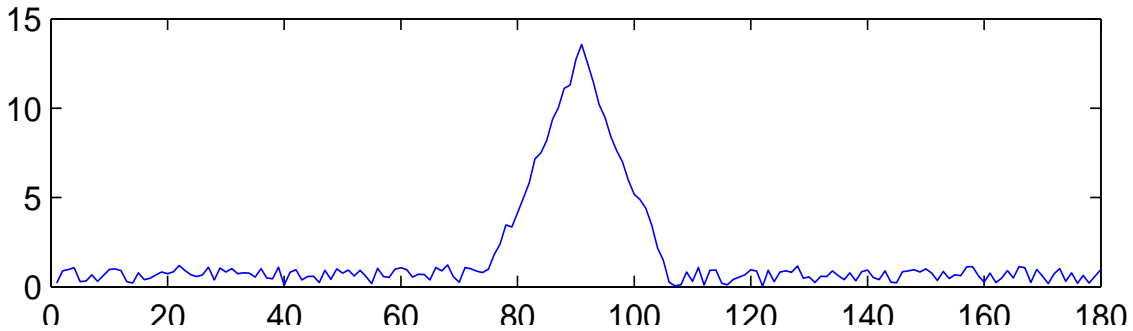


Figure 4.9: Small complex

Note: for clarity, only six segments are shown here. Actually, 600 segments are used for averaging.

## 4.4 Refining for ECG

During the procedure to detect the occurrence of heart beats, there are two main problems where special attention is needed.

### 4.4.1 Choice of window width of spectrogram

The choice of window width is essential to retain sufficient information in the spectrogram, and at the same time gives the nice separability property. If the window is too long, say triple the duration of one ECG complex, then the spectrogram is smooth along the time and no interesting heartbeat trend can be obtained. On the other hand, if the width is small, say only a fifth of the duration of one ECG complex, then the spectrogram captures the fine details of the non-stationary ECG complex. Due to these details, its spectrogram is no longer separable. In our experiment, we use the Blackman window with the width of a healthy maternal ECG complex.

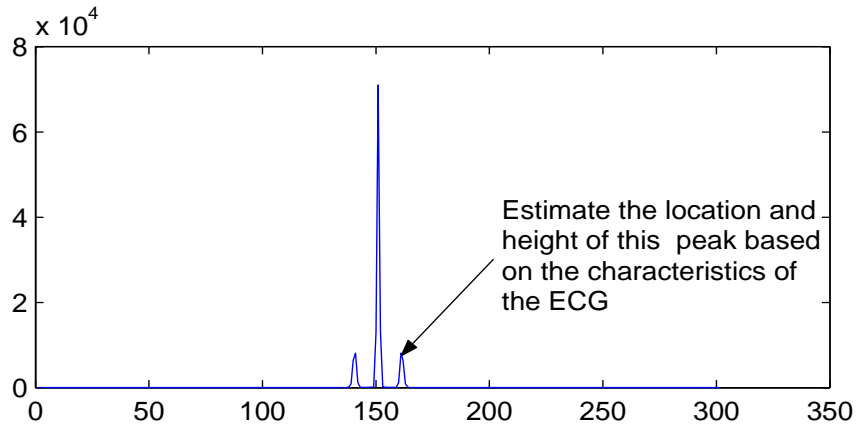


Figure 4.10: Frequency (108.raw).

#### 4.4.2 Selecting the best component after ICA

ICA yields the components in arbitrary order. In order to find which component is for maternal heartbeats and which is for fetal heartbeats, we take the frequency characteristic into account. Since the ECG signal is quasi-periodic, the expected spectrum should have only one peak whose location and height can be estimated by the approximate heart rate. Therefore, the sampling frequency will be enough for us to select the correct heart beats trend. Assigning maternal and fetal labels is facilitated by the *a priori* knowledge that the fetal heartbeat frequency is higher than the maternal.

## 4.5 Conclusion

In this chapter, we give a method which combines SVD and ICA to detect the fetal heart beats occurrence and the fetal ECG complex.

By using the spectrogram of the single-channel ECG signal, we can use the multichannel segregation techniques of ICA. Furthermore, by using frequency domain knowledge, we overcome the ambiguities of ICA and could determine which is the

expected component automatically.

For the fetal ECG complex detection, we first subtract a suitable maternal ECG from the mixture, next, time domain averaging is employed to get the fetal ECG complex. In this procedure, the main challenge is to generate the matched maternal ECG. To produce a ‘good’ match, the template are produced carefully, then scaling and shifting help to refine the maternal ECG.

In the last section of this chapter, two aspects which are important in the implementation of the proposed method are given.

Results in chapter 5 show the proposed method works well for detecting the heart beats occurrence and extract the fetal ECG from single-channel abdominal ECG.

## Programmes and experimental results

### 5.1 Programmes structure

Programmes are written in Matlab scripts, and are tested in Matlab 6.1. The running time depends mostly on how many iterations ICA need to find the heart beat trend. Normally, one iteration for maternal trend and two or three iterations for fetal trend, time range from two minutes to four minutes under Pentium III 700 with a 256M RAM.

Figure.5.1 shows the structure of the programmes.

The analysis includes seven parts:

1. Using SVD for maternal heart beats occurrence detection.
2. Using SVD for fetal heart beats occurrence detection.
3. Using ICA and SVD for maternal heart beats occurrence detection.
4. Using ICA and SVD for fetal heart beats occurrence detection.
5. Computing fetal ECG when the maternal heart beats are given.
6. Apply ICA and SVD on fetal ECG for its heart beats detection.

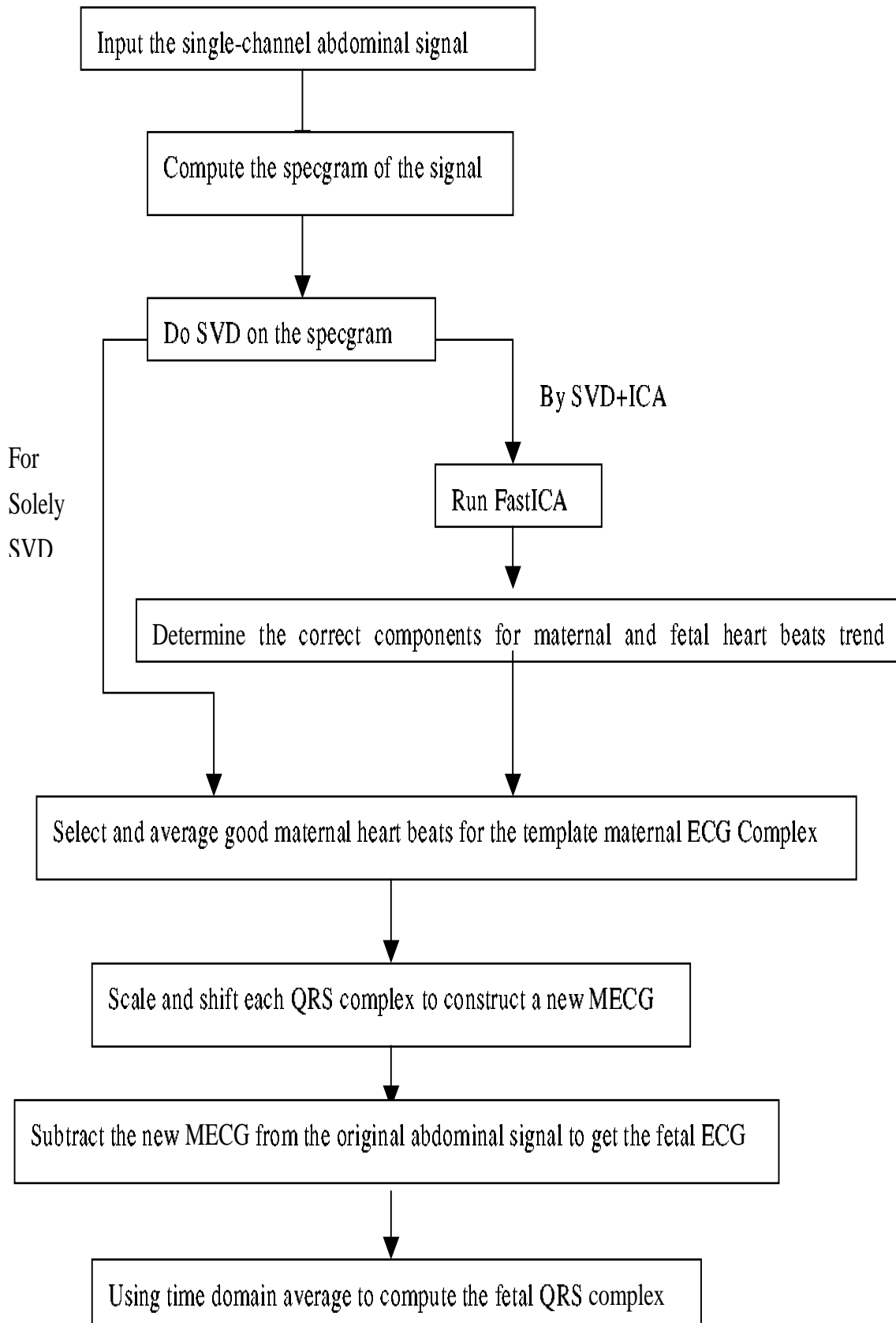


Figure 5.1: Programme Structure

7. Compute the fetal ECG complex.

Using the programmes, the following experiments have been done:

1. Compare the heart beats occurrence detection for synthetic data, by SVD and ICA+SVD.
2. Compare the maternal heart beats occurrence detection for the real life data sets, by SVD and ICA+SVD.
3. Compare the fetal heart beats occurrence detection for the real life data sets, by SVD and ICA+SVD, without knowing the maternal ECG complex.
4. Compute the fetal heart beats occurrence detection for the real life data sets after knowing the maternal ECG complex.
  - (a) Extract the fetal ECG from the mixture
  - (b) Detect the fetal heart beats occurrence by working on the fetal ECG instead of the mixture
5. Compute the maternal ECG complex and fetal ECG complex for the real life data sets.

## 5.2 Experimental results

### 5.2.1 Synthetical data and results

Due to the lack of ground truth, we evaluate the performance on a few signals by visual inspection. We also compose a few synthetic mixtures where ground truth are available.

*Synthetic data:* The synthetic mixture(Figure.5.5 is constructed from two simulated ECG complexes (Figure.5.2 and Figure.5.3). Note that the energy of one complex



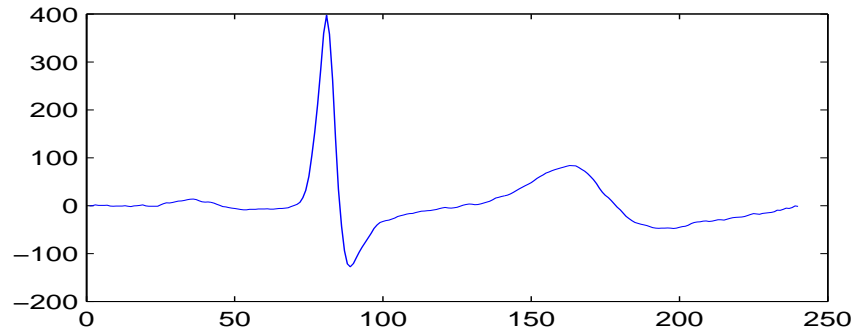


Figure 5.2: Synthetic maternal ECG complex

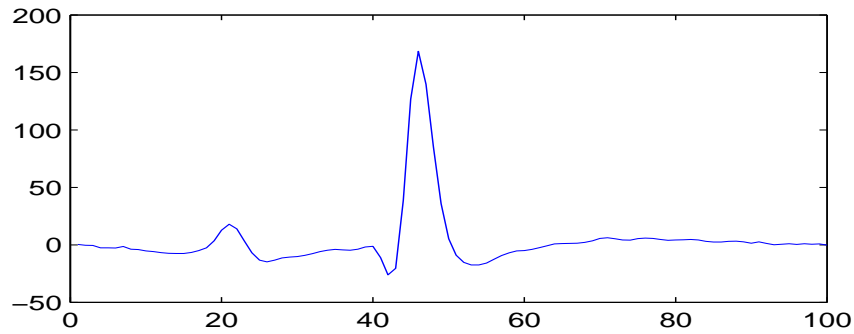


Figure 5.3: Synthetic fetal ECG complex

is higher than the other. This is to emulate the relatively strong maternal ECG and weak fetal ECG. One period of the maternal and fetal ECG complex is 240 and 100 samples respectively.

We compare the proposed method with the method that uses SVD as described in Section 4.2.1, which finds the  $\mathbf{u}_m$ ,  $\mathbf{v}_m$ ,  $\mathbf{u}_f$ , and  $\mathbf{v}_f$ , that minimize the noise.

Figure.5.5 compares the fetal heart beats detected by our method with those found using SVD. Clearly, our proposed method is much better even for the periodic synthetic signal than using solely SVD.

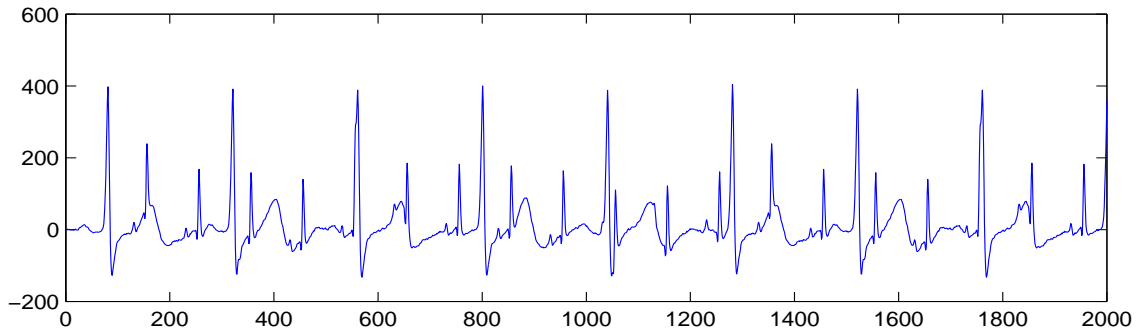


Figure 5.4: Synthetic data: Constructed by Figure.5.2 and Figure.5.3

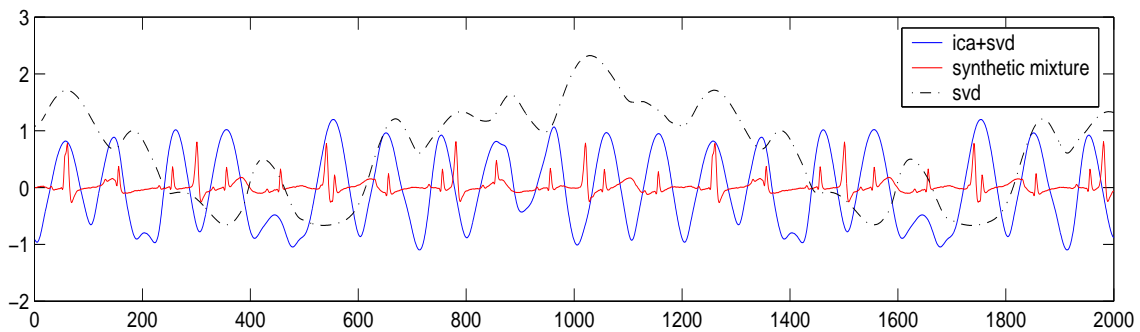


Figure 5.5: Comparison for the results from SVD and SVD+ICA on synthetic data

In order to check the suitability of this method, synthetic signals with different strength ratio of the maternal heart beats and fetal heart beats are composed and analyzed. Fig.5.6-Fig.5.8 show that for strength ratio up to 6, the proposed method works well. We can also see the detection accuracy trend in Fig.5.9. Furthermore, regarding the noise level, experiments show that the proposed method will not be affected when noise level (the variance of the noise) is less than 10. Fig.5.10 gives the result when the noise level is 10.

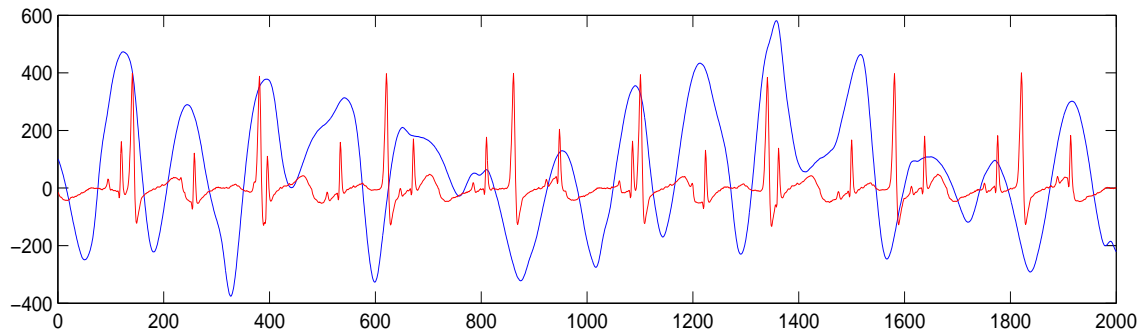


Figure 5.6: Synthetical data detection result for strength ratio=4

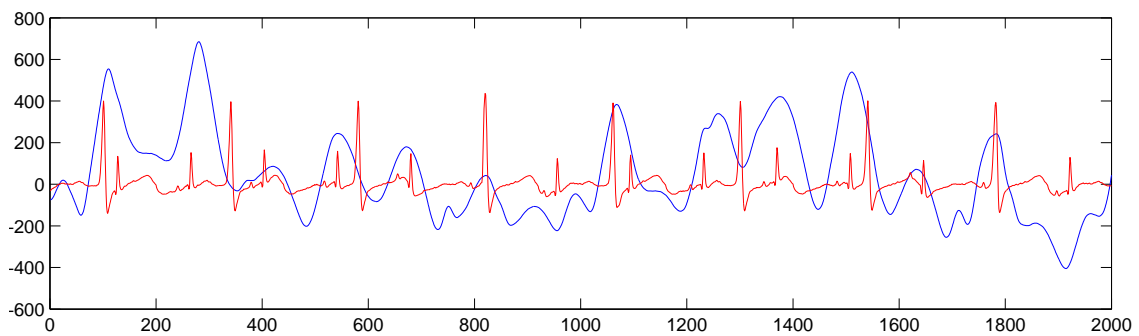


Figure 5.7: Synthetical data detection result for strength ratio=5

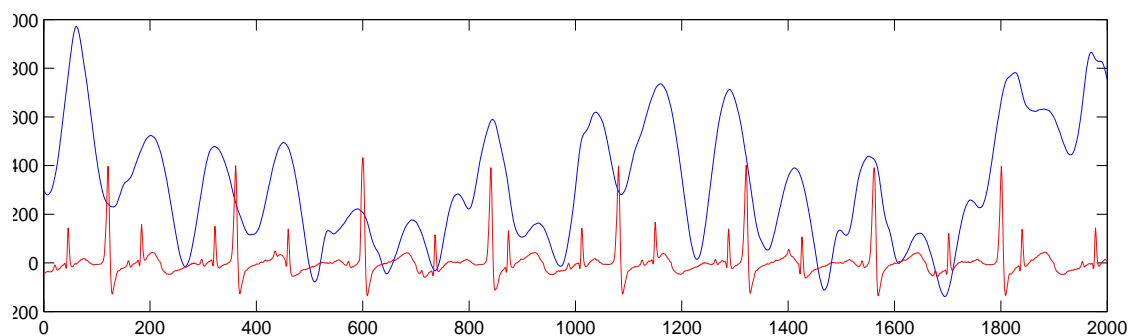


Figure 5.8: Synthetical data detection result for strength ratio=6

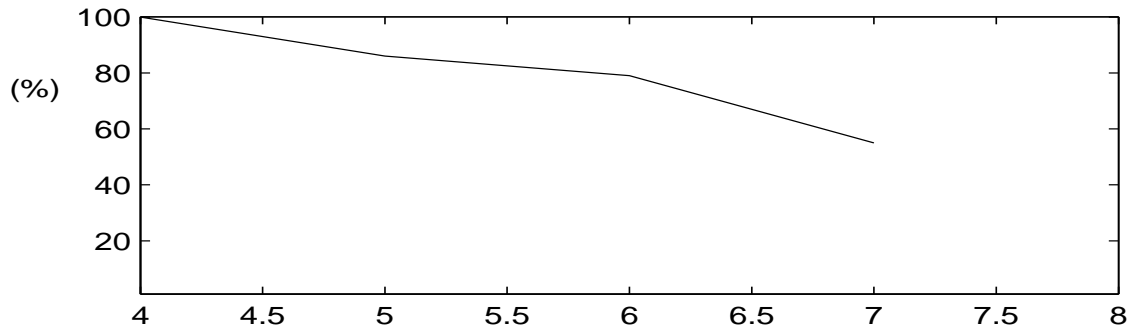


Figure 5.9: Detection accuracy for different strength ratio between maternal and fetal ECG

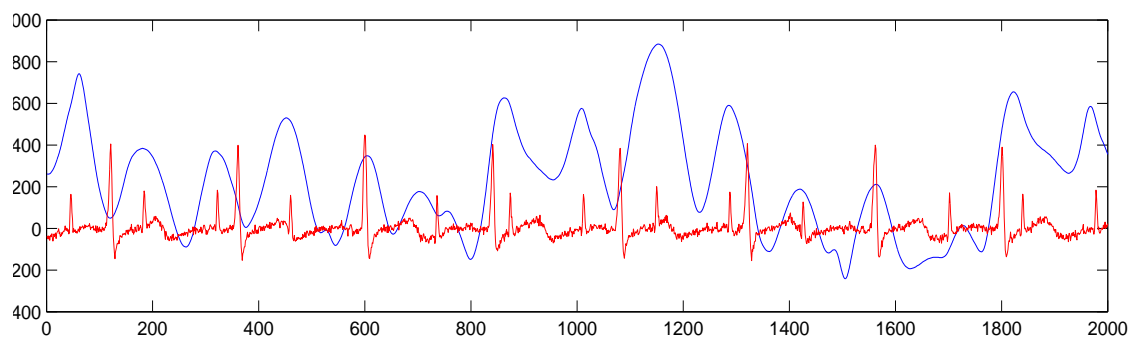


Figure 5.10: Synthetical data detection result when noise level= 10

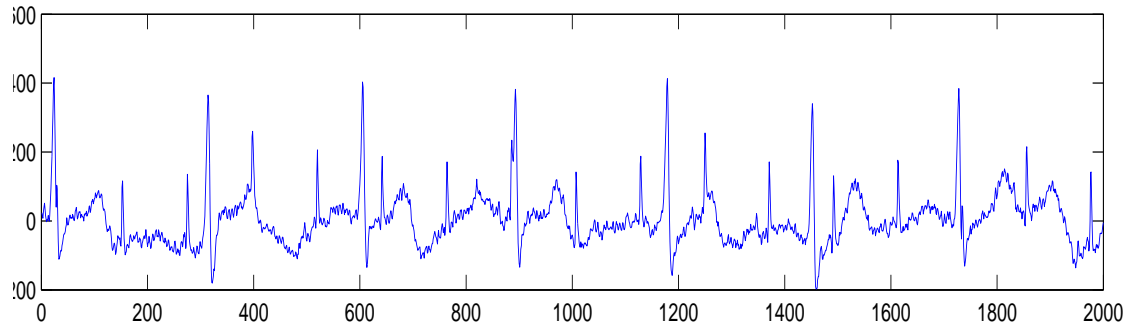


Figure 5.11: Original recorded data:108.raw

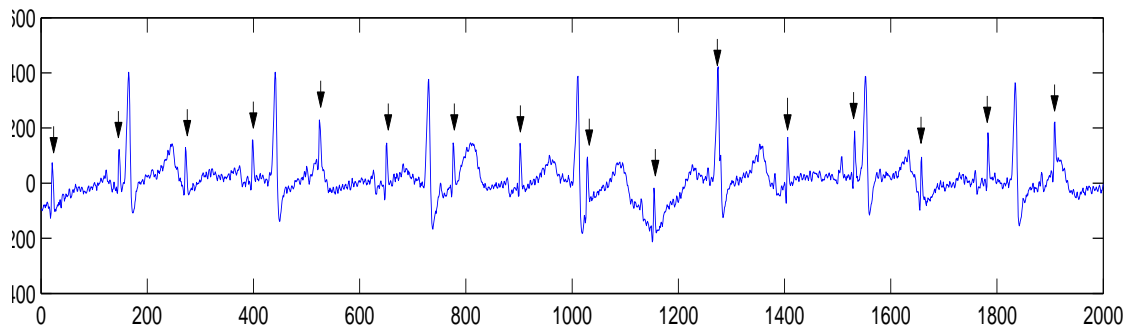


Figure 5.12: Original recorded data:292.raw

### 5.2.2 Experiments on real-life data

*Recorded signal:* In the second set of experiments, we performed the comparison for a number of recorded signals. We will present two in this section. The signals are obtained from two patients with a gestation period of 37 weeks. Each signal is about 10 minutes long, with sampling rate 300Hz (roughly  $1.8 \times 10^5$  samples). The heart rate can vary across the time, especially for the fetus who might move during the recording. Nevertheless, the maternal heart rate is slower, and ranges around 60-110 times per minute. Figure.5.11 and 5.12 shows a short part of the original signals.

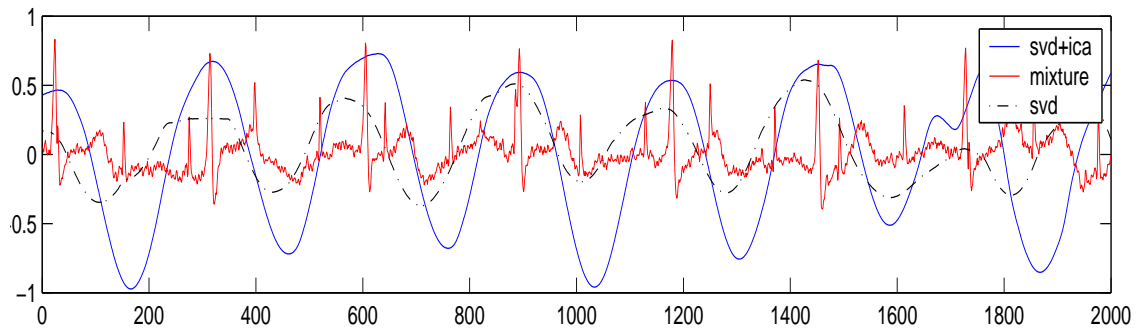


Figure 5.13: Comparison of results by SVD and SVD+ICA for maternal heart beats occurrence detection(108.raw)

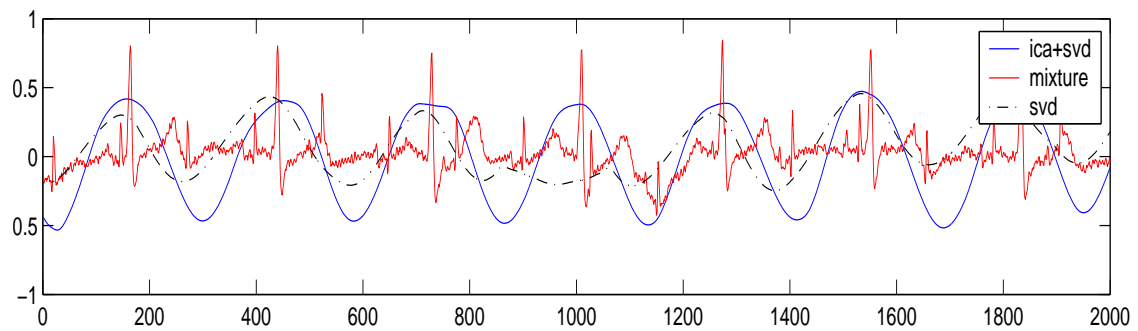


Figure 5.14: Comparison of results by SVD and SVD+ICA for maternal heart beats occurrence detection(292.raw)

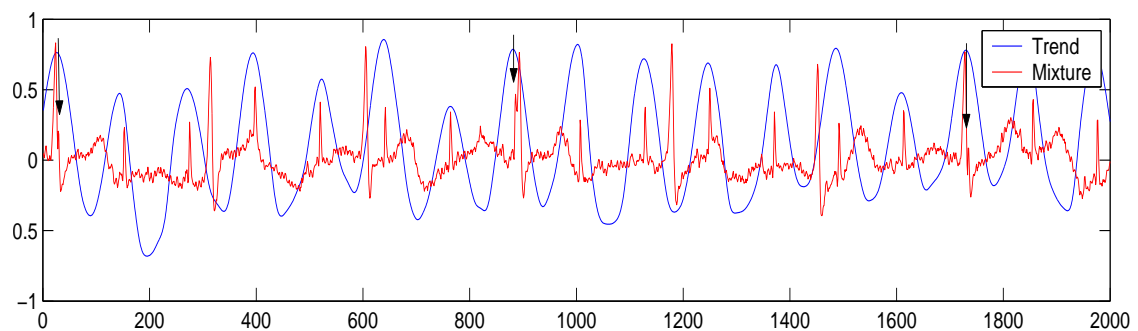


Figure 5.15: Another example: fetal heart beats occurrence detection by SVD + ICA. Arrows indicates heart beats that are difficult to detect.

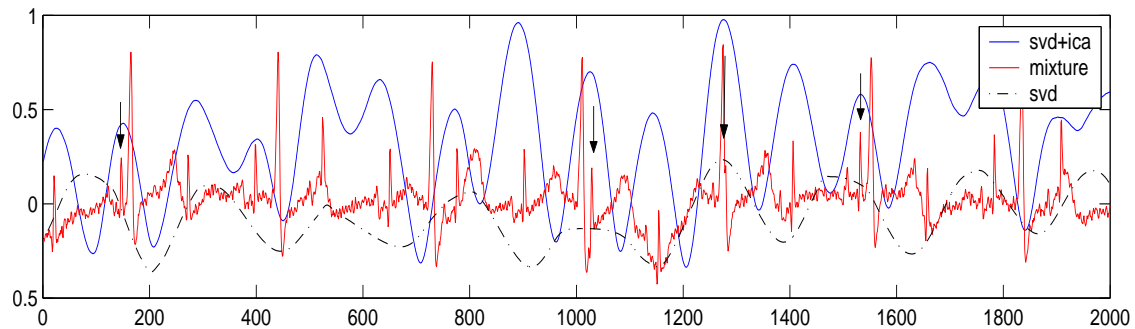


Figure 5.16: Fetal trend comparison of SVD and ICA for 292.raw. Arrows indicates heart beats that are difficult to detect.

### Comparison of results by SVD and SVD+ICA

Figure.5.13 and Figure.5.14 is a comparison for detection of maternal heart beats occurrence using SVD and our method. Both methods give good detection. However, our method is able to detect some occurrences where SVD fails.

Figure.5.15 and Figure.5.16 are comparisons for detection of fetal heart beats occurrence between the two methods. The SVD performs poorly. It gives a heartbeat trend that is seriously influenced by the maternal's. The proposed method gives good detection. It successfully detects all the heartbeat occurrences in both figures, but falsely detects two occurrences in Figure.5.15 (the false detections can be filtered out using domain knowledge). Note that it succeeds in cases where the maternal and fetal heartbeat coincide.

### 5.2.3 Fetal ECG extraction

Figure.5.3 and Figure.5.3 are the fetal ECGs after subtracting the scaled and shifted maternal ECG complex template.

Figure.5.17 and Figure.5.18 are the results of heart beats occurrence detection on the fetal ECG signal.

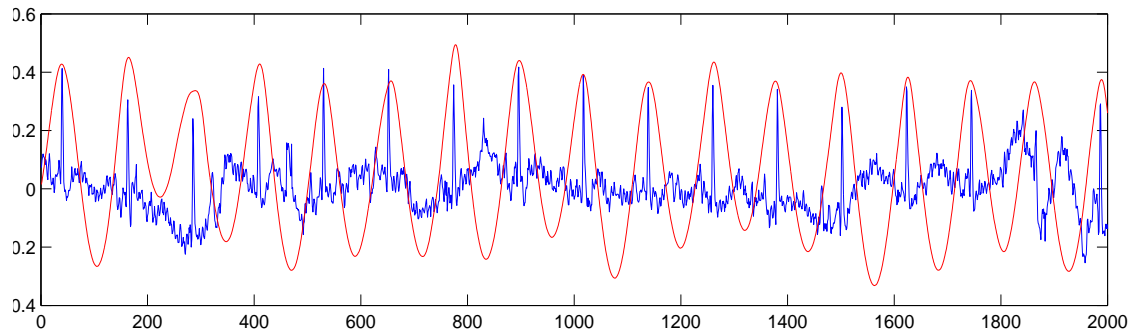


Figure 5.17: Fetal Trend by ICA for 108.raw after removing maternal ECG

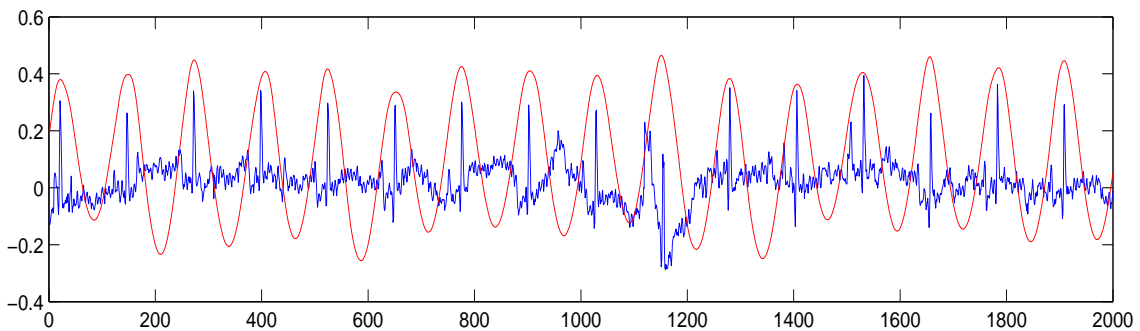


Figure 5.18: Fetal Trend by ICA for 292.raw after removing maternal ECG

#### 5.2.4 ECG complex results

Figure.5.19, Figure.5.20, Figure.5.21 and Figure.5.22 are the Maternal and Fetal ECG complex for 108.raw and 292.raw respectively which are all obtained by our proposed methods.

### 5.3 Conclusion

From the results, we can see that our proposed method works well not only for synthetic data, but also for real-life data. The comparison between SVD and SVD+ICA indicates that when independence are taken into account, more promising detection



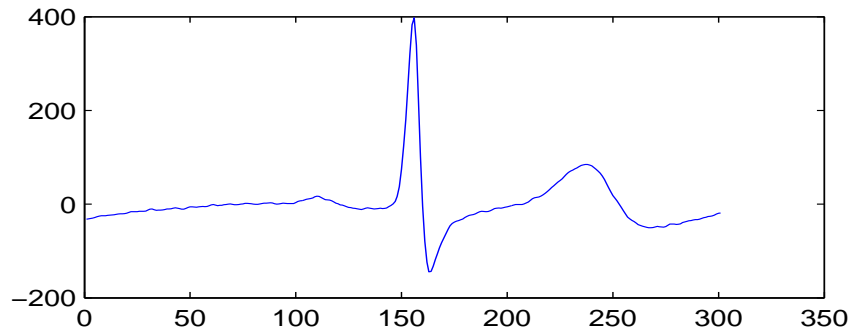


Figure 5.19: Maternal ECG complex for 108.raw

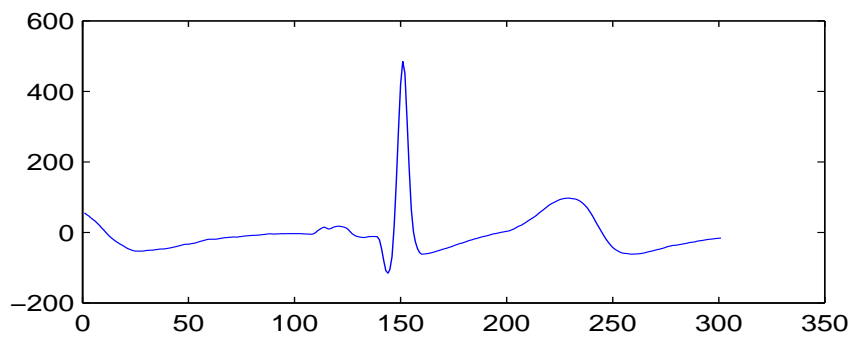


Figure 5.20: Maternal ECG complex for 292.raw

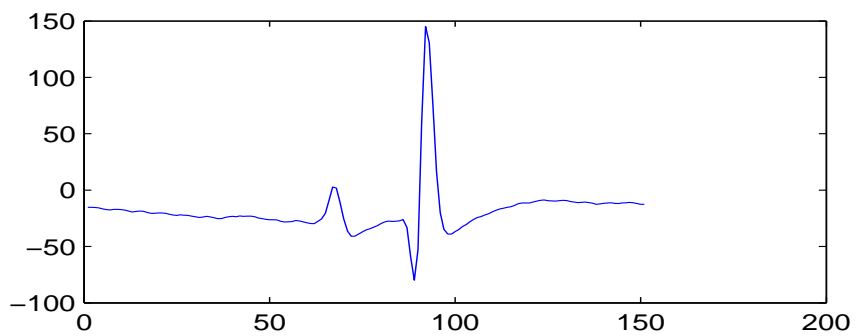


Figure 5.21: Fetal ECG complex for 108.raw

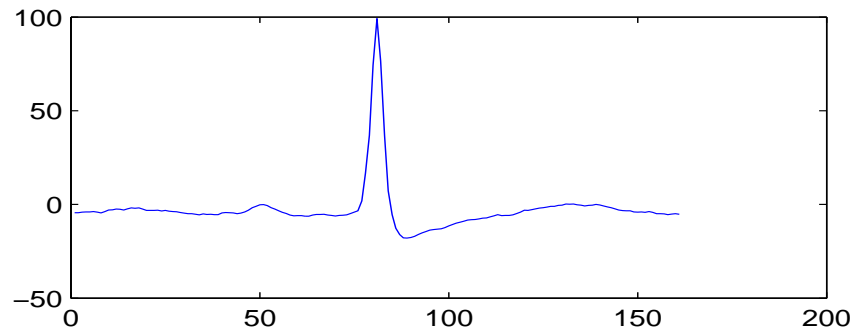


Figure 5.22: Fetal ECG complex for 292.raw

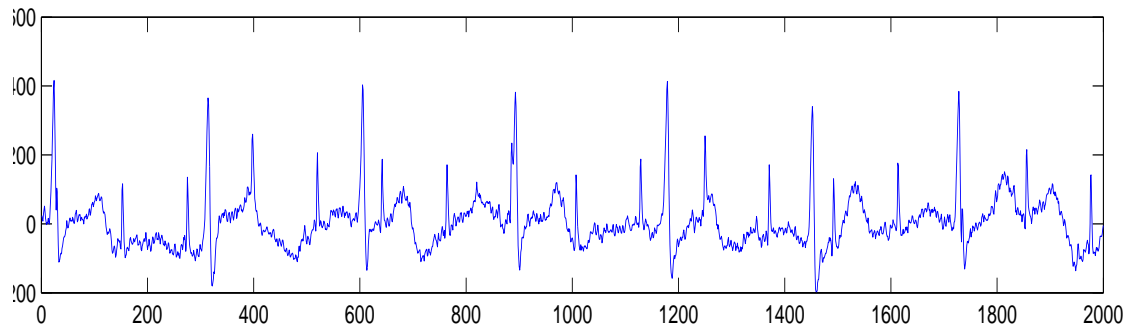


Figure 5.23: Original signal: 108.raw

could be expected.

For convenience, here, we give the original ECG and the fetal ECG we have extracted.

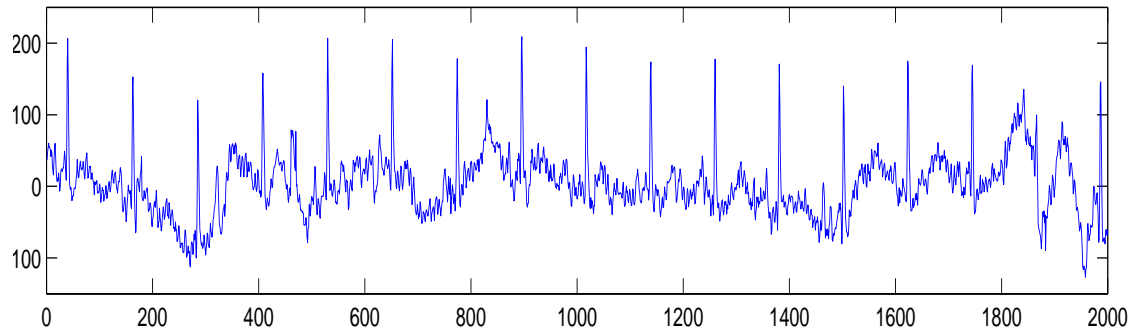


Figure 5.24: Fetal ECG for 108.raw

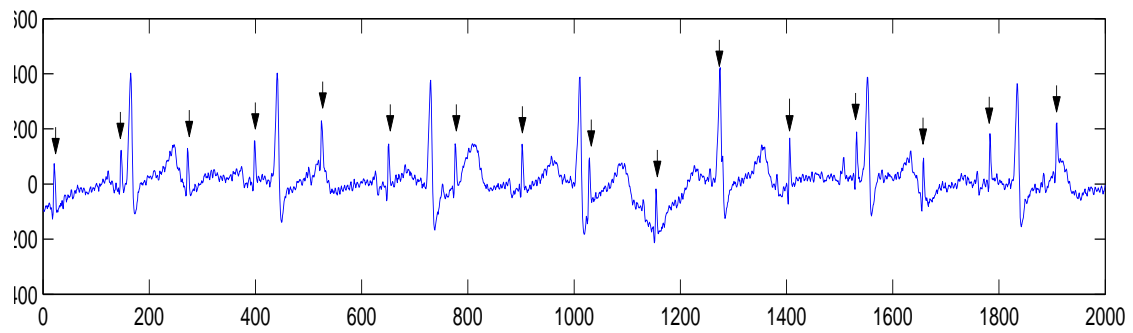


Figure 5.25: Original signal: 292.raw

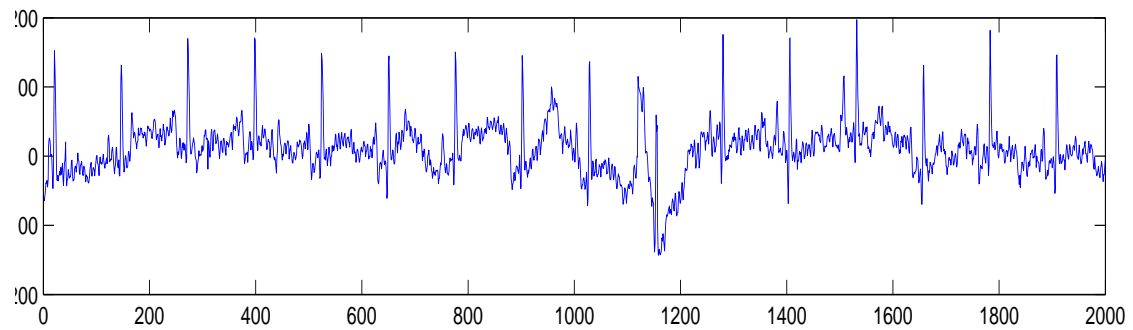


Figure 5.26: Fetal ECG for 292.raw

## Discussion and Conclusion

### 6.1 Discussion

Many methods[5, 6, 49, 50, 51, 52] proposed to extract the fetal ECG. However, most of the methods are working on multi-channel extraction. In multi-channel extraction, one aspect often ignored is the problem of eliminating the effects of differential interferences due to extraneous reasons(e.g., due to respiratory activity[45]) on the thoracic signals and on the composite abdominal ECG signals.

In this work, SVD and ICA have been combined to obtain the fetal ECG from single-channel composite signal. By computing the spectrogram of the original signal, we can use the multichannel segregation techniques of ICA. The ambiguities of ICA (lack of any ordering to the separated signals) is manageable with an obvious application of domain knowledge.

The computational load due to SVD is not much because 1). fast implementation are possible; 2). only partial SVD is necessary. In general, SVD-based methods ([5, 6, 50, 51]) including the proposed method are expected to be more immune to noise than others.

Most ICA algorithms, are either iterative fixed point algorithms (such as FastICA) or gradient descent algorithms, both of which optimize a solution only locally and are sensitive to initial randomization conditions that can produce quite different solutions, even for exactly the same signal. We view our technique of first separating the spectral basis vectors before submitting the remixed time domain signals to ICA as a way of setting up advantageous initial conditions that contribute to the stability of the solutions for the time domain separation. Therefore, the above mentioned interference problems do not affect the proposed method.

Results show that the proposed algorithm works well for extracting a fetal ECG from the composite signal. Since it only uses single-channel recording, there are no confounding issues that arise from having original signals that can differ more complicated ways than simply signal mixture levels.

## 6.2 Conclusion

At first when we began to work on this project, we tried the direct methods: locate the maternal heart beat by its peak, then get the template by averaging, deduct the template from each ECG complex. Then the ‘pure’ fetal ECG signal (do not include the maternal part) is obtained. Similarly for getting the Fetal ECG template complex. Even though the ‘big picture’ seems alike as our new algorithms, the original one need lots of manual interference. Furthermore, since no other characteristics of the signal are used, the only information is the magnitude of the signal, it can only work for quite limited cases and those with a larger strength ratio( $\geq 1/4$ ). However, the normal ratio is usually less than  $1/5$  which makes our original method useless for fetal heart beat detection.

When trying to improve the results, we were attracted by the popular ICA idea.

It is very natural for this project to use ICA since fetal ECG and Maternal ECG could be reasonably assumed to be independent.

Problems occurs when considering only single-channel mixture available, not that as required by the ICA—at least the same number mixtures as the components(In our case, at least three mixtures should be available for our three components: maternal ECG, Fetal ECG and noise).

Fortunately, noticing that ECG is nearly periodic, we could transfer the single-channel mixture to the multi-channel case. Such transformation makes it possible to use ICA for single-channel mixture.

Compared with the existing works, the proposed method is better in the following aspects:

1. Only one mixture is needed which makes the data collection much easier and avoids the multi-interferences of extraneous reasons which all the multi-channel extraction[5, 6, 50, 51].
2. In the single-channel fetal ECG extraction method proposed by P.P. Kanjilal [29], the locations of the fetal heart beat peaks are required to be known before doing the extraction. However, it is very difficult to do alignment for fetal ECG in practice. Our method could detect the heart beats trend as well as the locations of the heart beat peaks automatically and thus it is a feasible way for fetal ECG extraction and could serve as the preprocessing procedure for method in [29] or others which need the alignment.
3. The computational load which usually comes with SVD is avoided by using Partial SVD. Approximately, the proposed method need several minutes(2-4minutes) to extract the fetal ECG complex from the original mixture under Matlab 6.1 on PC with Pentium III 700 and a 256M RAM.

However, the method is far from perfect. there is still a big space for improvement, such as:

1. After doing experiments on dozens of mixtures, we found that even though the proposed method can detect the maternal heart beats very accurately, it fails to detect the fetal heart beats when the ratio of the fetal heart beat strength to the maternal heart beat strength is small. Experiments on synthetic data give the limit ratio as  $1/6$ .
2. Since no ground truth exists, the only way for us is to locate the Fetal ECG by its peaks. In other words, if the accuracy for peak detection is high, that means we have done a good job. This is our estimation method.

Here, one aspect should be noted: the data we use come from the FEMO—a monitor of ECG, it could detect the heart beats for both mother and fetus. However, before the data was recorded, the student (This project is cooperated with professor Ho Ting-fei in Medical department of National University of Singapore. All the data are collected by her students) who collected it removed some unknown parts which may seem not ‘good’. Therefore, alignment is impossible and no way to compare the two results. But one point for sure is that when our method could detect most of the fetal heart beats (according to our estimation method: peaks detection), the FEMO fails.

More work should be done later to set up a standard estimation system which would be much useful for comparing all the methods and help to understand the limits and advantages of each method.

3. In this project, even though all the algorithms such as averaging, SVD and ICA have some ability to denoise. We did not denoise explicitly. This should be done in future work.

---

## Bibliography

---

- [1] S. Amari, A. Cichocki, and H. H. Yang. *A new learning algorithm for blind source separation*. Advances in Neural Information Processing 8, pp.757-763. MIT press, Cambridge, MA, 1996.
- [2] A. J. Bell and T. J. Sejnowski. *An information-maximization approach to blind separation and blind deconvolution*. Neural Computation, 7:1129-1159, 1995.
- [3] P. Bergveld and W. H. J. Meijer. *A new technique for the suppression of the MECCG*, *IEEE Trans. Biomed. Eng.*, vol. ME-28, pp. 348-354, Apr. 1981.
- [4] A. Cichocki and R. Unbehauen. *Neural Networks for Signal Processing and Optimization*. Wiley, 1994.
- [5] D. Callaerts, J. Vandershoot, J. Vandewalle, W. sansen, G. Vantrappen, and J. Janssens. *An adaptive on-line method for the extraction of the complete fetal electrocardiogram from cutaneous multilead recordings*. *J. Perinatal Med.*, vol.14, pp421-433, 1986.



- 
- [6] D. Callaerts, B. De Moor, J. Vandewalle, and W.Sansen. *Comparison of SVD methods to extract the fetal electrocardiogram from cutaneous electrode signals.* Med., Biological Eng. and computing, vol. 28, pp.217-224, 1990.
- [7] J. -F. Cardoso and B.Hvarn Laheld. *Equivariant adaptive source separation.* IEEE Trans., Signal Processing, 44(12):3017-3030, 1996.
- [8] J.-F. Cardoso. *Infomax and maximum likelihood for source separation.* IEEE Letters on Signal Processing, 4:112-114,1997.
- [9] P. Comon. *Independent component analysis - a new concept?* Signal Processing, 36:287-314, 1994.
- [10] T. M. Cover and J. A. Thomas. Elements of Information Theory, John Wiley & Sons, 1991.
- [11] L. De Lathauwer, B. De Moor, and J. Vandewalle. *Fetal electrocardiogram extraction by blind source separation.* ESAT/SISTA, Leuven, Belgium, Tech. Rep.98-127, 1998.
- [12] L. De Lathauwer, B. De Moor, and J. Vandewalle. *Blind source separation by simulataneous third-order tensor diagonalisation.* Proc. EUSIPCO, Italy, vol.3, pp.2089-2092, 1996.
- [13] L. De Lathauwer, B. De Moor, and J. Vandewalle. *Fetal electrocardiogram extraction by blind source separation.* IEEE Trans. Biomed Eng., vol.47, No.5, pp. 567-572, May 2000.
- [14] A. G. Favret and A. F. Caputo. *Evaluation of Autocorrelation Techniques for Detection of the Fetal Electrocardiogram.* IEEE Trans, Biomed. Eng., vol. BME-13, pp.37-43, January 1966.

- 
- [15] Ping Gao, Ee-Chien Chang and Lonce Wyse, *Blind Separation of Fetal ECG from Singale Mixture using SVD and ICA*. 4th Int. Conf. on Information, Communications & Signal Processing and 4th Pacific-Rim Conf. on Multimedia (ICICS-PCM 2003).
- [16] G. H. Golub and C. F. Van Loan. *Matrix Computations*, 3<sup>rd</sup>ed. Baltimore, MD:Johns Hopkins Univ. Press, 1996
- [17] A. Hyvarinen. *Fast and Robust Fixed-Point Algorithms for Independent Component Analysis*. IEEE Transactions on Neural Networks 10(3):626-634, 1999.
- [18] A. Hyvarinen and E. Oja. *Independent Component Analysis: Algorithms and Applications*. Neural Networks, 13(4-5):411-430, 2000.
- [19] A. Hyvarinen and E. Oja. *A Fast Fixed-Point Algorithm for Independent Component Analysis*. Neural Computation, 9(7):1483-1492, 1997.
- [20] A. Hyvärinen, E. Oja, P. Hoyer, and J. Hurri. *Image feature extraction by sparse coding and independent component analysis*. Proc. Int. Conf. on Pattern Recognition(ICPR'98), pp. 1268-1273, Brisbane, Australia, 1998.
- [21] A. Hyvärinen. *New approximations of differential entropy for independent component analysis and projection pursuit*. In *Advances in Neural Information Processing Systems*, volume 10, pages 273-279. MIT Press, 1998.
- [22] A. Hyvärinen. *The fixed-point algorithm and maximum likelihood for independent component analysis*. Neural Processing Letters 10(1): 1-5, 1999.
- [23] A. Hyvärinen. *Independent component analysis: algorithms and Applications*. Neural networks, 13(4-5):411-430, 2000.

- 
- [24] Aapo Hyvärinen, Juha Karhunen and Erkki Oja. Independent Component Analysis. John Wiley & Sons, Inc. 2001
- [25] E. H. Hon and S. T. Lee. *Electronic evaluation of the fetal heart rate patterns preceding fetal death*. Further observations, Am. J.Obstet. Gynecol.87(1965)814-826.
- [26] M. C. Jones and R. Sibson. What is projection pursuit? J. of the Royal Statistical Society, ser. A, 150:1-36, 1987.
- [27] C. Jutten and J. Herault. *Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture*. Signal Processing, 24:1-10, 1991.
- [28] J. Karhunen, E.Oja, L. Wang, R. Vigario, and J. Joutsensalo. *A class of neural networks for independent component analysis*. IEEE Trans. Neural Networks, 8(3):486-504, 1997.
- [29] P. P. Kanjilal and S. Palit. *Singular Value Decomposition applied to the modelling of quasiperiodic processes*. IEEE Trans. on Signal Processing, vol. 35, No. 3,pp. 257-267, 1994.
- [30] Kanjilal, P. P., Palit, S. *On multiple pattern extraction using singular value decomposition, IEEE Trans. Signal Processing*. IEEE trans. on Signal Processing, vol.43, pp.1536-1540, June 1995
- [31] P. P. Kanjilal and Goutam Saha. *Fetal ECG Extraction from Single-Channel Maternal ECG using Singular Value Decomposition*. IEEE Trans. Biomed. Eng., vol. 44, No. 1, Jan, 1997.
- [32] H. P. Kunzi and W. Lrelle. Nicht Lineaire Programmierung. Berlin, West Germany: Springer Verlag. 1962, ch5, pp. 73-79.

- 
- [33] Ali. Khamene. *A New Method for the Extraction of Fetal ECG from the Composite Abdominal Signal*. IEEE Trans. Biomed. Eng., vol. 47, No. 4, April, 2000.
- [34] T. W. Lee. *Independent Component Analysis-Theory and Applications*. Kluwer, 1998.
- [35] R. L. Longini et al. *Near-Orthogonal Basis Functions: A Real Time Fetal ECG Technique*. IEEE Trans, Biomed. Eng., vol. BME-24, pp.39-43, January 1977.
- [36] D. G. Luenberger. *Optimization by Vector Space Methods*. John wiley & Sons, 1969.
- [37] W. J. H. Meijer and P. Bergveld. *The Simulation of the Abdominal MEGG*. IEEE Trans. Biomed. Eng., vol. BME-28, pp.354-357, Apr.1981.
- [38] C. L. Nikias and J. M. Mendel. *Signal processing with higher order spectra*. IEEE Signal Processing Magazine, pp.10-37, July 1993.
- [39] J.-P. Nadal and N. Parga. *Nonlinear neurons in the low noise limit: a factorial code maximizes information transfer*. Network, 5:565-581, 1994.
- [40] D.-T. Pham, P. Garrat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. In Proc. EUSIPCO, pages 771-774, 1992.
- [41] E. Oja. *The nonlinear PCA learning rule in independent component analysis*. Neurocomputing, 17(1):25-46, 1997.
- [42] F. Ori, G. Monitor, J. Weiss, X. Sayhouni and D. H. Singer. *Heart rate variability:Frequency domain analysis*. Card. Clin. 10(1992) 499-537.
- [43] A. Papoulis. *Probability, Random Variables and Stochastics Processes*. McGraw-Hill, 3rd Edition, 1991.

- [44] B. A. Pearlmutter and L. C. Parra. *Maximum likelihood blind source separation: A context-sensitive generalization of ica*. In Advances in Neural Information Processing Systems, volume 9, pages 613-619, 1997.
- [45] R. Pallas-areny, J. Colominus-balague, and F.J.Rosell. *The effect of respiration-induced heart movements on the ECG*. IEEE Trans. Biomed. Eng., vol. BME-36, pp.585-590, 1989.
- [46] M. Richter, T. Schreiber and D. T. Kaplan. *Fetal ECG Extraction with Nonlinear State-Space Projections*. IEEE. Trans. Biomed., vol. 45, No.1, January, 1998.
- [47] E. Soria, M.Martinez, J. Calpe, JV. Frances, AJ. Serrano and JF. Guerrero. *A New Non-Linear Recursive Algorithm for Obtaining the Fetal Electrocardiogram*. IEEE. Computers in Cardiology, vol. 24, 1997.
- [48] L. Tong, R. Liu, V. Soon, and Y.-F. Huang. *Indeterminacy and indetifiability of blind identification, IEEE transactions on Circuits and Systems*. vol.38, pp. 499-509, May 1991.
- [49] A. Van Oosterom. *Patial filtering of the fetal electrocardiogram*. J. Perinatal Med., vol. 14, pp. 411-419,1986.
- [50] J. Vandershoot, D.Callaerts, W.Sansen, J.Vandewalle, G.Vantrappen, and J.Janssens. *Two methods for optimal MECC elimination and FECC detection from skin electrode signals*. IEEE Trans.Biomed Eng., vol. BME-34, pp.233-243,1987
- [51] J. H. Van Bemmelen. *Detection of weak electrocardiograms by autocorrelation and cross correlation envelopes*. IEEE Trans, Biomed. Eng., vol. BME-15, pp 17-23, 1968.

- 
- [52] B. Widrow, J. M. McCool, J. Kanmitz, C. Williams, R. Hearn, J. Zeidler, E. Dong, and R. Goodlin, Adaptive noise cancelling Principles and applications, in Proc. IEEE, 1975, vol.63, no. 12, pp. 1692-1716.
- [53] V. Zarzoso and A. K. Nandi. *Blind separation of independent sources for virtually any source probability density function*. IEEE Trans. Signal Processing, vol. 47, No. 9, pp.2419-2431, September 1999.

**Name:** Gao Ping  
**Degree:** Master of Science  
**Department:** Computational Science  
**Thesis Title:** Blind Separation For Fetal ECG from Single Channel Mixture  
By SVD and ICA

### Abstract

In this thesis, we propose a novel blind-source separation method to extract fetal ECG from a single-channel signal measured on the abdomen of the mother. The signal is a mixture of the fetal ECG, the maternal ECG and noise. The key idea is to compute the spectrogram of the original signal, and then use an assumption of statistical independence between the components to find the trends of the original signal. This is achieved by applying Singular Value Decomposition (SVD) on the spectrogram, followed by an iterated application of Independent Component Analysis (ICA) on the principle components. The SVD contributes to the separability of each component and the ICA contributes to the independence of the two components. We further refine and adapt the above general idea to ECG by exploiting a-prior knowledge of the maternal ECG frequency distribution and other characteristics of ECG. Experimental studies show that the proposed method is more accurate than using SVD only. Because our method does not exploit extensive domain knowledge of the ECGs, the idea of combining SVD and ICA in this way can be applied to other blind separation problems.

**BLIND SEPARATION FOR FETAL ECG  
FROM SINGLE MIXTURE BY SVD AND ICA**

**GAO PING**

**NATIONAL UNIVERSITY OF SINGAPORE**

**2003**



BLIND SEPARATION FOR FETAL ECG FROM SINGLE MIXTURE BY SVD AND ICA      GAO PING      2003

BLIND SEPARATION FOR FETAL ECG FROM SINGLE MIXTURE BY SVD AND ICA      GAO PING      2003