

结合 ICA 和 PCA 方法的胎儿心电提取

李文娟, 郭晓静, 吴小培

(安徽大学 计算智能与信号处理教育部重点实验室, 安徽 合肥 230039)

摘 要:在生物医学信号处理领域,独立分量分析(PCA)和主分量分析(ICA)是两种广泛应用的方法。但是,这两种方法各有其优缺点。提出了一种新颖的方法,将 ICA 和 PCA 相结合,通过求相关的技术,分别取 ICA 和 PCA 方法的优点。将该方法应用于从母体腹部测得的多通道信号中提取胎儿心电信号的实验,得到令人满意的结果。研究表明,这种结合 ICA 和 PCA 的方法能够比较准确地分离出所需要的胎儿心电信号,进而可以对胎儿心电进行监护,因此在临床上具有一定的实用价值。

关键词:主分量分析;独立分量分析;胎儿心电;相关

中图分类号:TN911.73

文献标识码:A

文章编号:1673-629X(2007)08-0223-03

Fetal ECG Extraction by Combining Methods of ICA and PCA

LI Wen-juan, GUO Xiao-jing, WU Xiao-pei

(Ministry of Education Key Laboratory of Intelligent Computing & Signal
Processing, Anhui University, Hefei 230039, China)

Abstract: In the field of biomedical signal processing, independent component analysis (ICA) and principal component analysis (PCA) are two widely-used methods. But they both have their advantages and disadvantages. In this paper, a novel method of combining PCA and ICA is proposed. This method is applied to the fetal ECG extraction from the multi-channel signals getting from the mother's abdomen. The experiment results are satisfying. According to the analysis and experiment results, the method can extract the fetal ECG accurately. So the new method has the practicability in the application of clinic.

Key words: principal component analysis; independent component analysis; fetal ECG(FECG); correlation

0 引言

对孕妇进行心电测量,得到的是孕妇心电与胎儿心电的混合信号,而更有用的是将胎儿心电提取出来,从而来诊断胎儿发育是否正常。胎儿信号是一种微弱的信号,总是混于比其大得多的母体心电中,要将测量所得的混合信号分离^[1,2]开来,ICA(独立分量分析)方法可以得到比较准确的结果。但是,ICA 得到的分离信号的顺序是随机的,如果没有经验知识或者当一些其他源信号跟想要的信号类似时,就很难用肉眼选出所要的信号。而 PCA(主分量分析)是按能量大小排列的,顺序较固定。对于这个混合心电的分离而言,由经验可知,能量靠前的应该是孕妇的心电信号,其次是

胎儿的心电信号,所以从 PCA 分离出来的各路信号中可以很容易地找出胎儿心电信号。然后把 PCA 分离出来的胎儿心电信号跟 ICA 得到的所有分离信号进行相关运算,相关系数最大的是 ICA 分离出来的胎儿心电信号。这就是结合 ICA 和 PCA 的胎儿心电提取原理。

1 独立分量问题的描述

在信号处理领域的复杂的背景噪声中所观测到的信号是由不同的信号源所产生的多路信号的混合,设 $x_i(t)$ 是一组用 N 个传感器观测到的 N 维时间信号,其中下标 $i = 1, 2, \dots, N$ 表示观测的通道序号, $t = 1, 2, \dots, T$ 表示观测的时间离散数据, $x_i(t)$ 是由 M 个变量 $s_j(t)$, $j = 1, 2, \dots, M$ ($M \leq N$) 线性混合产生,混合矩阵用 A 来表示,即 $x = As$ 。现在的任务是,在 s 和 A 均为未知的条件下,求取一个解混矩阵 W ,使得 x 通过它后所得输出 y ($y = Wx$) 是 s 的最优逼近。这时的 y 就是分离出来的源信号。如在心电处理中,找出的隐

收稿日期:2006-10-17

基金项目:国家自然科学基金资助项目(60271024);安徽省人才基金资助项目(2004Z028)

作者简介:李文娟(1985-),女,安徽宿州人,硕士研究生,研究方向为智能信息处理;吴小培,教授,博士生导师,研究方向为现代智能信号处理。

含信号就是心电源或其他噪声源部分。上述过程可用框图表示,如图 1 所示。

$$s \rightarrow \boxed{A} \rightarrow x = As \rightarrow \boxed{B} \rightarrow y = Wx$$

图 1 独立分量问题的描述模型

2 主分量分析(PCA)方法

在统计分析中,主分量分析方法^[3]是一种很重要的数据处理方法,在信号处理、模式识别领域又被称为 KL 变换。研究资料表明,PCA 最早由 Pearson 引入,并应用于生物学领域,后来 Hotelling 等将其应用于心理测试。近几年来,PCA 的应用越来越广泛,在数据挖掘、特征提取、信号恢复和数据压缩等领域都有了很好的应用。

主分量分析的输入输出模型即为图 1 所示的模型。其输入 x 是 N 个随机向量,每一个随机向量又是由 T 次实现变量构成的,如下式所示:

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \Leftrightarrow \begin{bmatrix} x_1(1) & \cdots & x_1(T) \\ \vdots & & \vdots \\ x_N(1) & \cdots & x_N(T) \end{bmatrix} \quad (1)$$

假设这里的 x 的均值 $E(x) = 0$, 则其协方差矩阵为:

$$C_x = E[xx^T] = \begin{bmatrix} E[x_1^2] & \cdots & E[x_1x_n] \\ \vdots & & \vdots \\ E[x_nx_1] & \cdots & E[x_n^2] \end{bmatrix} \quad (2)$$

PCA 的目的就是寻找一正交变换矩阵 W , 对输入的随机向量 x 进行正交变换,使得输出的随机向量 y 中的各随机向量间彼此不相关,也就是相互正交。即:

$$y = Wx \Rightarrow \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad (3)$$

这里 y 的协方差矩阵为对角阵

$$C_y = E[yy^T] = \begin{bmatrix} E[y_1^2] & \cdots & E[y_1y_n] \\ \vdots & & \vdots \\ E[y_ny_1] & \cdots & E[y_n^2] \end{bmatrix} = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \quad (4)$$

所以我们可以看出,主分量分析是基于信号的二阶统计量的线性分析方法,其作用是消除信号间的相关性。因此 PCA 分离出的源信号是互不相关的信号,并且这些互不相关的源信号是按照能量从大到小的顺序进行排列的。所以在我们的研究过程中,一般只需要取前面的若干个主要源信号,而把后面一些含很少信息的成份舍弃掉。这是主分量分析算法的一个显著

优势。

3 独立分量分析(ICA)方法

独立分量分析^[4,5]是近几年发展起来的一种新的盲信号分离技术,与主分量分析相比,它不仅去除了信号的相关性,而且要求各高阶统计量独立,从而实现了统计意义上的独立分析。1994 年,Comon 系统地分析了瞬时混叠信号盲信号分离问题,提出了 ICA 的概念与基本假设条件,并基于累积量直接构造了目标函数。在独立分量分析算法的研究中,最早实现的算法是 20 世纪 90 年代中期, Bell 和 Sejnowski 提出的随机梯度下降学习算法,即最大熵 ICA 算法(Infomax ICA)。此后很多学者从不同的角度对独立分量分析算法进行了深入的研究,提出了一些对 Infomax 算法的改进思想和新算法。这儿就 ICA 的固定点算法进行阐述。固定点算法^[6]是对一组给定数据的递推计算,属于批处理。但是计算思路的导出却和自适应处理有些联系。固定点算法中应用比较广泛的是基于负熵的固定点算法。这里仍然采用图 1 所示的独立分量的描述模型。采用负熵为判据时,目标函数为:

$$J_G(w) = [E\{G(wx)\} - E\{G(v)\}]^2 \quad (5)$$

其中, $G(\cdot)$ 是非二次函数, v 是满足标准高斯分布的随机变数。由此可得出一个独立分量的迭代公式如下:

$$w_i(k+1) = \frac{E\{zf[w_i^T(k)z]\} - E\{f[w_i^T(k)z]\}w_i(k)}{E\{zf[w_i^T(k)z]\} - E\{f[w_i^T(k)z]\}w_i(k)} \quad (6)$$

$$w_i(k+1) \leftarrow \frac{w_i(k+1)}{\|w_i(k+1)\|_2} \quad (7)$$

其中, z 是 x 对进行白化后得到的。从上面可以看出,迭代分两步,第二步为归一运算。总之,采用负熵的固定点算法的步骤可总结如下:

- (1) 将 x 去均值,然后加以白化得 z 。
- (2) 任意选择 w_i 的初值 $w_i(0)$, 要求 $\|w_i(0)\|_2 = 1$ 。
- (3) 令 $w_i(k+1) = \frac{E\{zf[w_i^T(k)z]\} - E\{f[w_i^T(k)z]\}w_i(k)}{E\{zf[w_i^T(k)z]\} - E\{f[w_i^T(k)z]\}w_i(k)}$ 。
- (4) 归一化: $w_i(k+1) \leftarrow \frac{w_i(k+1)}{\|w_i(k+1)\|_2}$ 。
- (5) 如未收敛,返回步骤(3)。

4 结合 PCA 和 ICA 处理孕妇心电信号

上面分别介绍了 PCA 和 ICA 算法,对于测得孕妇的心电信号,可以认为测量所得的心电信号实际上就是母体心电、胎儿心电以及其他各种相互独立的干扰信号的合成。所以可以看出,用 ICA 的方法对这些独立分量进行分离可以达到更精确的效果。但是,ICA

方法分离出来的各独立分量的顺序是不固定的,要想从分离出的各分量中提取出所需要的分量,需要进行肉眼观察,对所需要的分量波形要有一些先验的知识。如果先验知识不够,可能导致错误的结论。而 PCA 的方法恰恰可能弥补这一不足。因为 PCA 分离出来的各分量是按能量大小顺序排列的,顺序基本固定。所以可将 ICA 和 PCA 结合起来以便更准确地提取胎儿心电信号。

原始信号是在孕妇的腹部测量得到的 5 路心电信号,波形如图 2 所示。

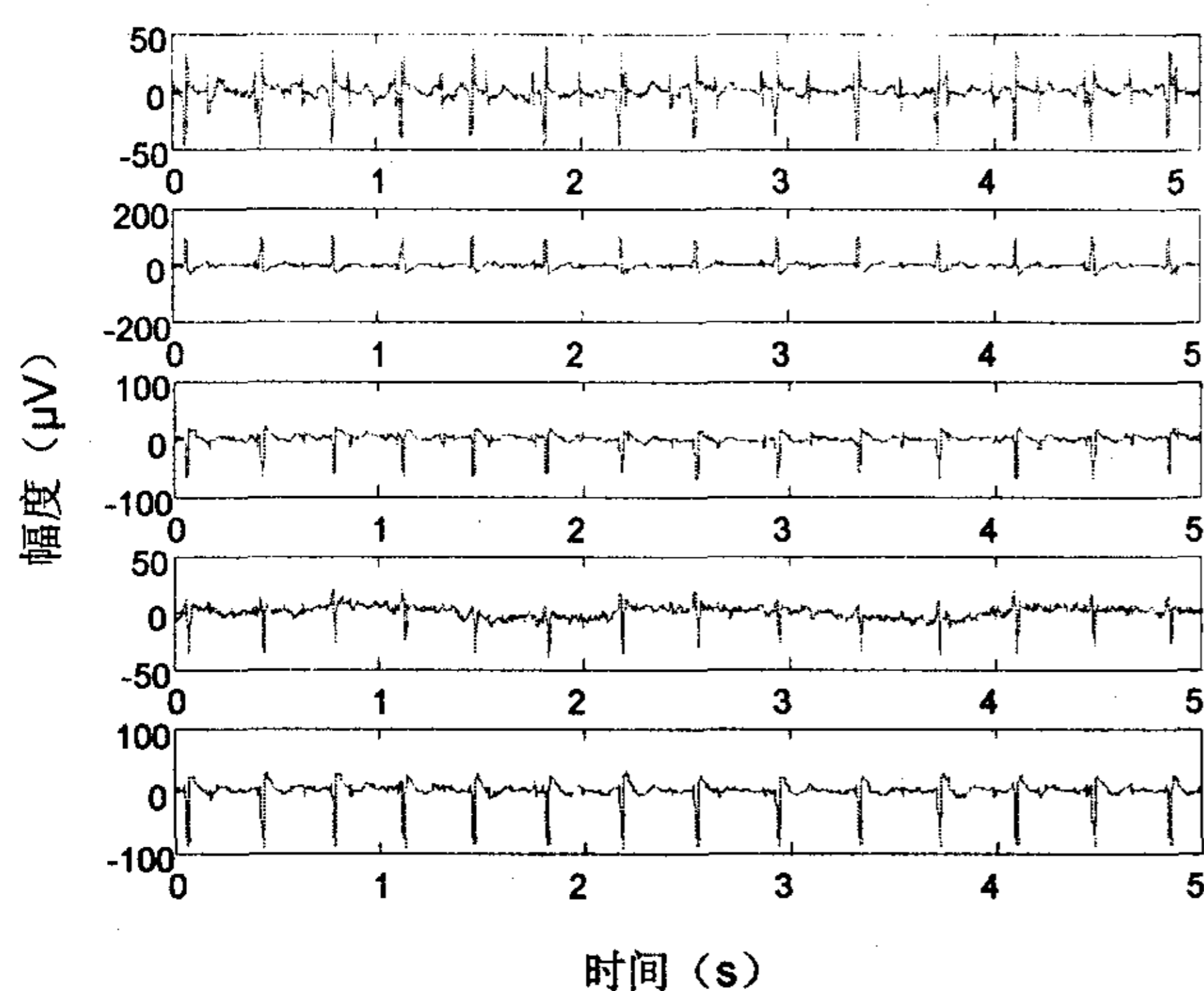


图 2 原始测量得到的心电信号

对于所用的孕妇心电信号,用 PCA 的方法分离出的分量是按能量大小顺序排列的,显然,能量靠前的应该是母体的心电信号,接下来的才是胎儿的心电信号,剩下的是一些干扰噪声信号。用 PCA 方法分离出的信号如图 3 所示。从 PCA 分离的结果中可以明显看出,第一路信号 PCA1 和第二路信号 PCA2 是母体心电成分。而第三路信号 PCA3 所示的心率明显高于 PCA1 和 PCA2,可以清晰地分辨出是胎儿心电,用 PCA_FECG 表示。

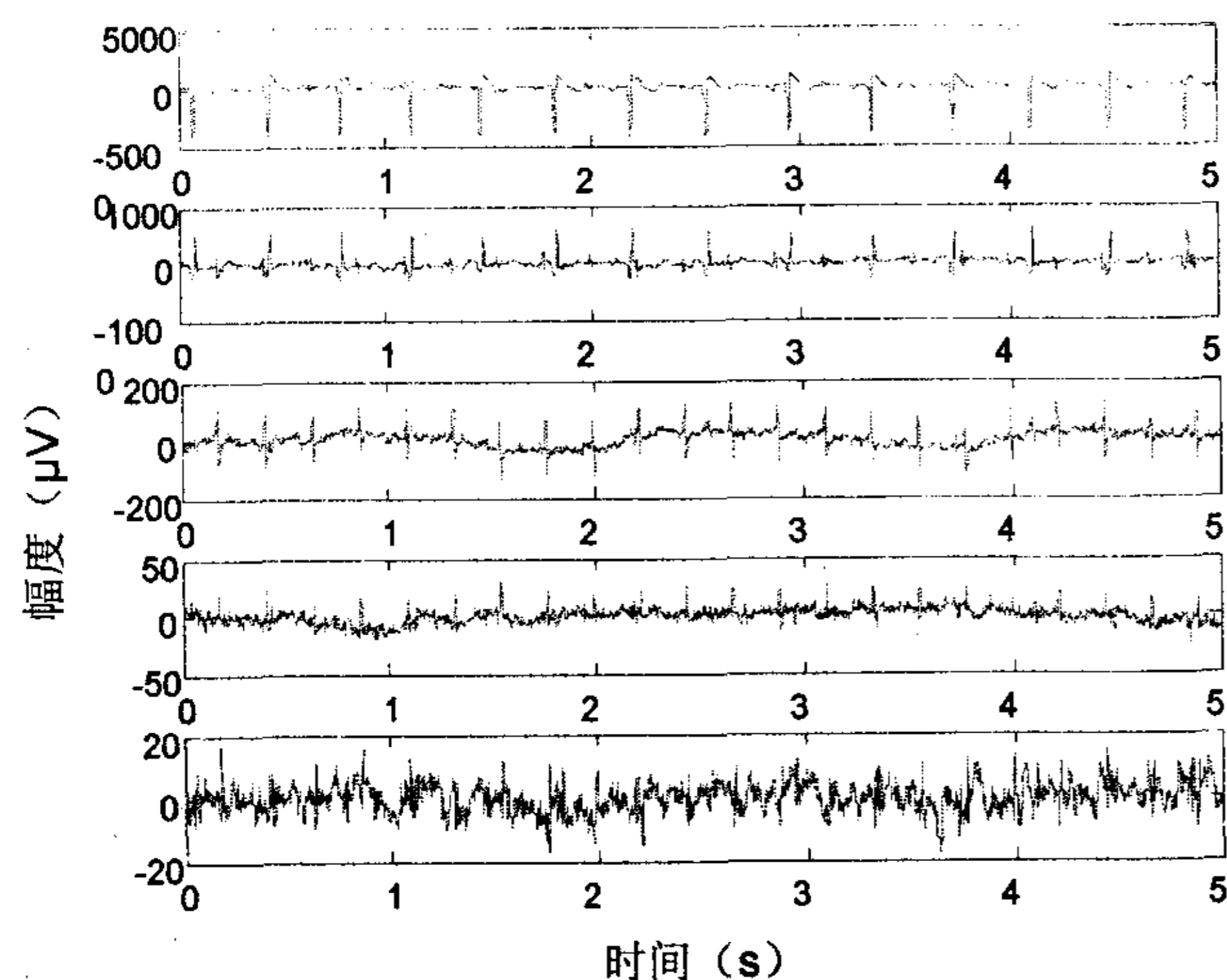


图 3 用 PCA 方法分离的结果

所以可以用 ICA 分离出的各路信号与 PCA3 做相关运算,相关系数最大的就是所要的用 ICA 方法提取出来的胎儿心电信号 ICA_FECG。用 ICA 方法分离出的各独立信号波形如图 4 所示。

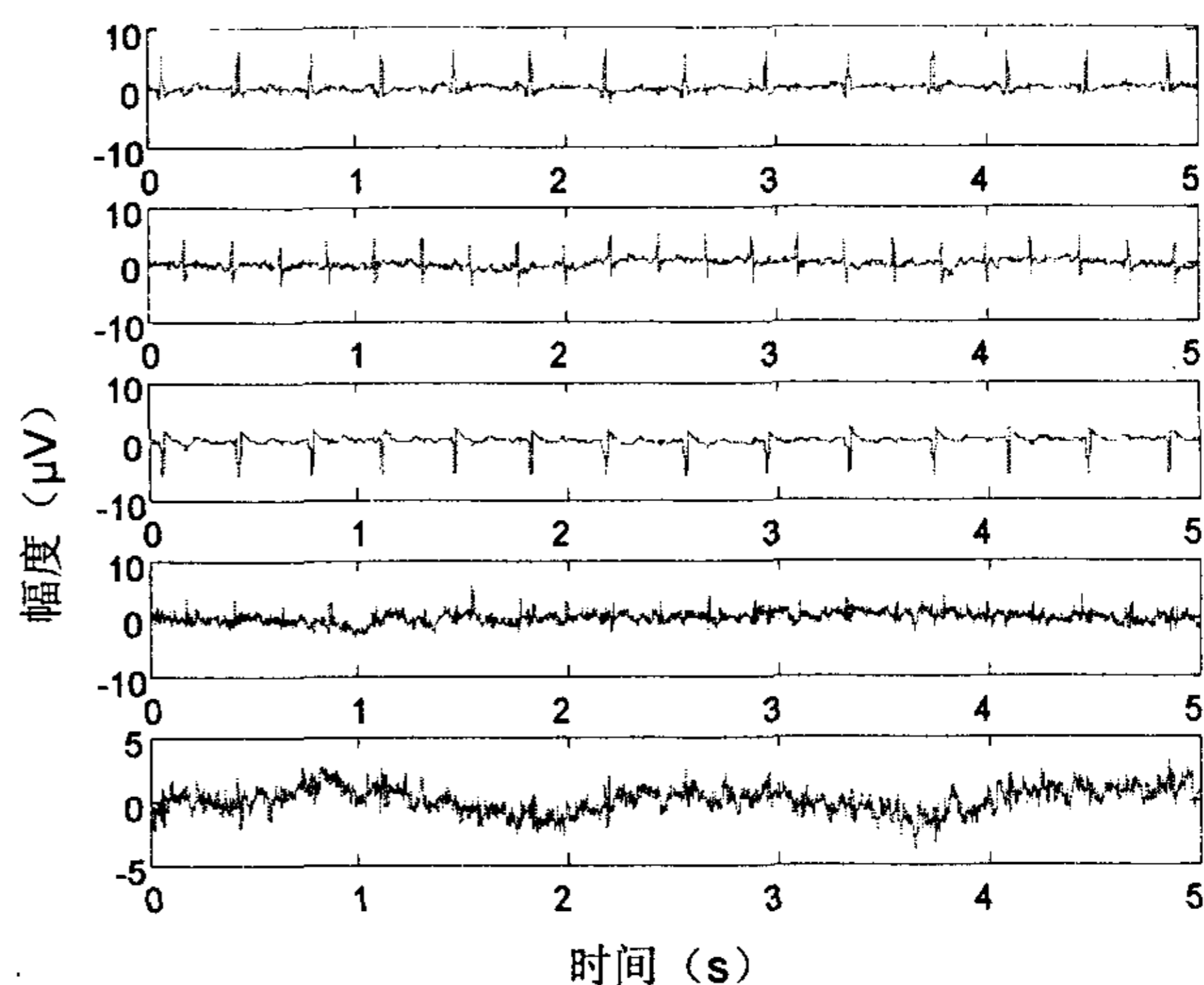


图 4 用 ICA 方法分离出的结果

现把图 4 所示的用 ICA 方法分离出的各路信号 ICA_s 与 PCA_FECG 做相关运算,结果如表 1 所示。

表 1 相关运算结果

ICAs	ICA1	ICA2	ICA3	ICA4	ICA5
与 PCA_FECG 的相关系数	0.0794	0.8201	0.0316	0.2444	0.5102

从表中可以清楚地看到,ICA 分离出的第二路信号 ICA2 与 PCA_FECG 有最大的相关系数,它就是通过 ICA 和 PCA 相结合的方法提取出来的胎儿心电信号 FECG,波形如图 5 所示。

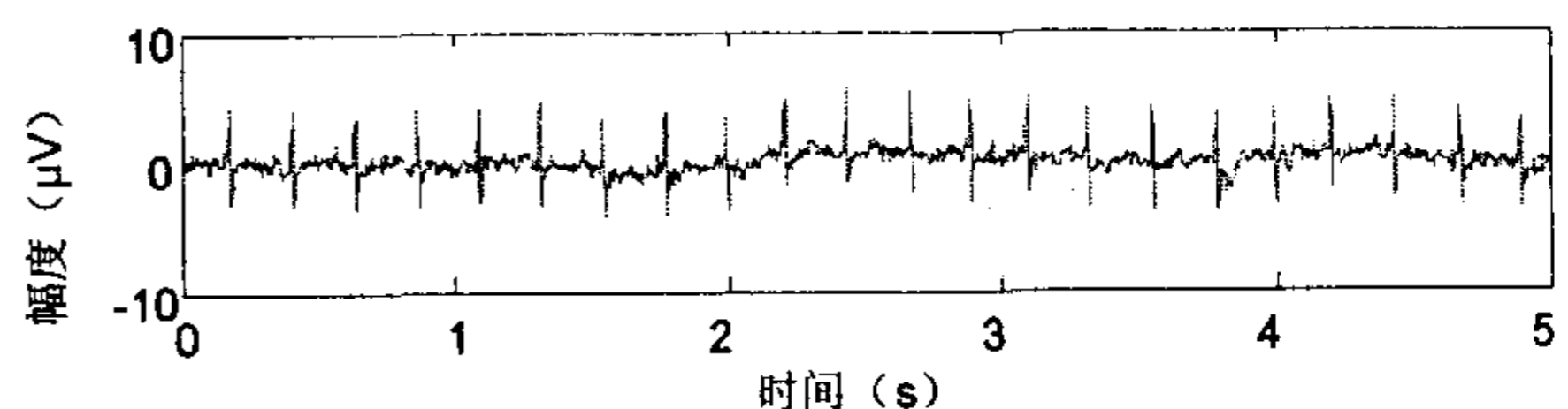


图 5 用 ICA 和 PCA 相结合的方法提取的胎儿心电信号 FECG

5 结束语

胎儿心电的提取有着重要的临床意义。通过提取胎儿心电,进行胎儿心电监护,是产期监护的重要内容。如何准确地提取出胎儿心电信号,一直是一个研究的焦点问题。文中将独立分量分析和主分量分析这两种重要的方法相结合,这样能够充分发挥这两种方法的优势,扬长避短,使信号分离能够更准确,精度更高。

参考文献:

- [1] Comon P, Jutten C, Herault J. Blind Separation of Sources,

(下转第 229 页)

```

static void nodeGetChild(DataTable L_Get, string fatherbh)
{
    foreach(DataRow dr in myDataTable.Rows)
    {
        string path = dr[nodeName].ToString();
        string text = dr[nodebh].ToString();
        //找到子节点
        if ((path.Length == fatherbh.Length + 2) && path.
StartsWith(fatherbh))
        {
            DataRow MyDR = L_Get.NewRow();
            MyDR["sbbm"] = text;
            L_Get.Rows.Add(MyDR);
            nodeGetChild(L_Get, path);
        }
    }
}

```

(4)合理使用存储过程。

水电费抄表后要进行核算,会对多个表进行批量处理,这时考虑使用传参到数据库服务器用存储过程来提高效率,但是实际使用时当核算节点太多时逐一传送节点编号又加大了系统开销,反而使处理变慢,本系统把需要多次传参的操作单独拿出来在 SQL 语句中通过条件表示为集合 where in (参数列表)一次处理,其他传参少但一次处理多个表的使用存储过程,二者结合使用后速度提高了 2/3。

(5)使用多线程。

系统归档时调用存储过程处理工作周期长,独占资源无法知道当前处理进度,为了使得程序更加人性化,此时采用了多线程来改变调用存储过程的进程独占 CPU 资源的情况,一个进行调用存储过程,另外一个显示进度,二者同步进行实现了对当前进度的实时监控。

(6)数据的备份和恢复。

数据的备份和恢复重要且复杂,本系统为用户提供了良好易操作的界面,采用了一星期一次完全备份、每天一次差异备份的办法,既使数据有了安全保障,又保证了备份数据存储量不至于增长速度太快。

(7)B/S 部分采用模板列页面多级调用时进行多参数传递^[4,5]。

在 B/S 模式下实现了类似于 C/S 的页面逐级调用返回。

下面是利用模板列进行 URL 多个变量参数的传递方法的代码:

```

<ItemTemplate>
    <asp:HyperLink runat="server" Text="查看"
        NavigateUrl='<%# "usepowerdetail.aspx? person=" +
        DataBinder.Eval(Container, "DataItem.person") + "&date="
        +
        DataBinder.Eval(Container, "DataItem.monthid") + "&url =
        userpowerlist.aspx"%>'
        ID="Hyperlink_detail" NAME="Hyperlink_detail">
    </asp:HyperLink>
</ItemTemplate>

```

4 结束语

通过算法的优化、方法与新技术的巧妙组合弥补了原系统功能上的不足,提高了处理速度,精简了部分无用的功能,使本软件的行业针对性更强,又通过 B/S 部分如 IE 浏览查询和 BBS 留言,增加了与用户间的沟通,提高了服务响应速度,使得收费系统更加人性化。本系统现在库尔勒供电段试用,它的应用将极大提高铁路供电段的工作效率。

参考文献:

- [1] 甄 镭..NET 与设计模式[M].北京:电子工业出版社,2005.
- [2] 罗宾逊,内格尔.C# 高级编程[M].李敏波 译.北京:清华大学出版社,2005.
- [3] 王晓东.计算机算法设计与分析[M].北京:电子工业出版社,2001.
- [4] 章立民.用实例学 ASP.NET[M].北京:电子工业出版社,2005.
- [5] Sussman D, Homer A. ASP.NET 高级编程[M].北京:清华大学出版社,2002.

(上接第 225 页)

- Part II: Problem Statement[J]. Signal Processing, 1991, 24 (1):11-20.
- [2] Cichocki, Unbehauen R. Robust Neural Networks with On Line Learning For Blind Identification and Blind Separation of Source[J]. IEEE Trans on Circuit and Systems, 1996, 43 (11):894-906.
- [3] 吴小培,李晓辉,张道信.多变量数据分析及应用研究[J].电子测量与仪器学报,2004,18(1):51-56.

- [4] Lee T W. Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources[J]. Neural Computation, 1999, 11(2):409-433.
- [5] 杨福生,洪 波.独立分量分析的原理与应用[M].北京:清华大学出版社,2006:97-101.
- [6] Hyvarinen A. A fast fixed-point algorithm for independent analysis[J]. Neural Computation, 1997, 9:1483-1492.