DNA 测序信号小波去噪分析的新方法*

郑华[1]，唐磊[1]，陆祖康[2]

（1 福建师范大学 激光与光电子技术研究所，福建省光子技术重点实验室，
医学光电科学与技术教育部重点实验室，福州 350007）

（2 浙江大学 现代光学仪器国家重点实验室，国家光学仪器工程技术研究中心，杭州 310027）

摘 要：为了真实构建信号模型并准确评价去噪算法的有效性，实验中通过实际系统中采集到的DNA荧光信号经前期研究中优选的小波去噪后，叠加随机噪音构建DNA测序仿真信号.去噪分析的结果表明：选择db8小波基函数、分解层数（*lev*＝5）与使用固定格式软阈值，有效去除了DNA测序信号的噪音.将其用于处理实际的DNA电泳荧光信号，相比基于高斯荧光信号峰模型筛选的算法，去噪后的信号更加真实可靠.

关键词：DNA测序；荧光信号；小波分析；去噪

中图分类号：TH74; TP391 **文献标识码**：A

0 引言

在DNA荧光测序分析中，毛细管电泳（CE）系统的样品进样量少且浓度低，噪音的存在很可能会导致错误的结果.因此，需要采用滤波的方法去除信号中的噪音.对于含噪信号的滤波，传统的方法有：曲线拟合法、移动平均法、样条函数拟合和Fourier变换等，这些滤波方法的共同特点是根据信号的特征设计最佳的滤波器，滤除噪音.但对于非平稳过程信号、含宽带噪音信号，采用传统方法处理有一定的局限性.小波分析由于具有良好的时频域分辨能力已成为信号处理的一种强有力工具[1]，在数学、物理、化学、通信、医学和地质等领域得到了广泛的应用.小波分析具有检测信号奇异性和突变结构的优势：信号和噪音在小波变换下表现出截然不同的性质，它能更准确地得到信号上特定点的奇异性信息[2~4]，因此小波分析已成为电泳荧光信号去噪的重要方法.

由于小波分析中用到的小波函数具有多样性，用不同的小波基函数分析同一个问题会得到不同的结果，因此，如何选择最优的小波基函数非常重要；此外，如何选择合适的小波分解层数与去噪阈值也直接关系到信号去噪处理的质量.本文将实际系统中采集到的DNA荧光信号经前期研究[5]中优选的sym7小波去噪后，作为理想荧光信号，叠加随机噪音构建DNA测序仿真信号.通过仿真分析，寻找适合于DNA测序CE荧光信号去噪的小波基函数与去噪处理的方法，并应用于实际的DNA测序荧光信号，获得了很好的实验结果.

*福建省青年科技人才创新项目(2008F3038)资助

+Tel: 0591-83465373 E-mail: hzheng@fjnu.edu.cn

收稿日期：2008-11-03

1 小波去噪原理

对于时域信号$f(t)$进行离散小波变换

$$C_{m,n}(t)=\int_{-\infty}^{+\infty} f(t)\psi(m,n,t)\,\mathrm{d}t \tag{1}$$

可得到一系列频率通带空间，即小波空间.目的是获得信号$f(t)$的离散小波分解在不同尺度下的带通项.Mallat提出了一种快速分解算法[4]，利用小波变换将信号分解成不同的频段成分

$$f(t)=cA_1(t)+cD_1(t) \tag{2}$$

其中$cA_1(t)$和$cD_1(t)$分别称之为$f(t)$在分辨率2^1下的离散逼近和离散细节，即$cA_1(t)$为频率不超过2^{-1}的部分，而$cD_1(t)$为频率介于2^{-1}和2^0之间的部分.$cA_1(t)$还可以继续分解

$$cA_1(t)=cA_2(t)+cD_2(t) \tag{3}$$

$$cA_2(t)=cA_3(t)+cD_3(t) \tag{4}$$

$$cA_{n-1}(t)=cA_n(t)+cD_n(t) \tag{5}$$

其中，$cD_n(t)$为信号中的高频部分，且分解次数越多，高频噪音的成分越少.因为噪音的能量总是少于有用信号的能量，所以可以从细节部分去除噪音，而不会影响信号中的有用成分.

经过上述快速分解算法的处理，可以得到信号$f(t)$的离散小波分解在不同尺度下的带通项，即

$$f(t)=cD_1(t)+cD_2(t)+\ldots+cD_n(t)+cA_n(t) \tag{6}$$

式中每一项$cD_n(t)$代表信号在某一频率下的信号大小，可以从各自所属的小波空间重构.在信号重构前，对上述小波空间中的向量进行不同的处理可得到信号的平滑和去噪的结果.平滑方法是设计一个低通滤波器，即选择一个截断尺度，使频率高于该尺度下的小波空间的向量全置为零；而去噪是选择一个去噪参数，使得在所有尺度下的小波空间中，模小于去噪参数的向量全置为零.对处理后的小波空间向量进行小

波重构便可以分别得到平滑和去噪的结果[6].

2 荧光信号去噪的仿真分析

2.1 含噪荧光信号的模拟

在目前的文献中，大多以Gaussian函数模拟毛细管电泳的激光诱导荧光（CE-LIF）信号

$$Y(t) = H\exp\left[-\frac{(t-t_R)^2}{2\sigma^2}\right] \tag{7}$$

其中，H和σ 分别为信号峰的峰高和半峰宽，t_R是信号峰的中心位置.

然后，在式(7)模拟的光谱信号上叠加随机噪音，作为含噪的仿真信号.随机噪音用式(8)产生[7]

$$v(i) = \frac{(0.5-\mathrm{RND})I_{\max}}{S/N} \tag{8}$$

RND为0~1之间的随机数，用来模拟白噪音，$I_{\max}$为真实信号中的最大值，S/N为信噪比.由于电泳荧光信号峰存在展宽现象，并不是严格的高斯型，因此在这个模式下优选的去噪方法对于实际信号的去噪效果不一定最优.因此我们以图1所示实验系统[8-10]（1：氩离子激光器；2：扩束器；3：平面镜；4：二色镜；5：平面镜；6：物镜；7：毛细管；8：滤色镜；9：汇聚透镜；10：共焦小孔；11：PMT；12：A/D转换

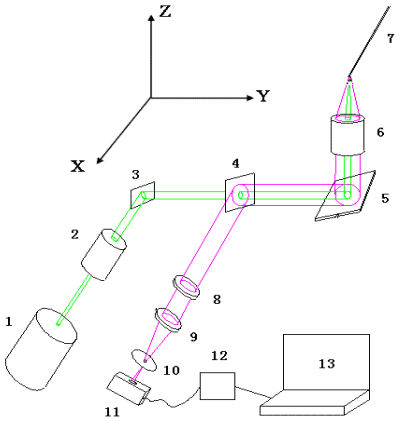
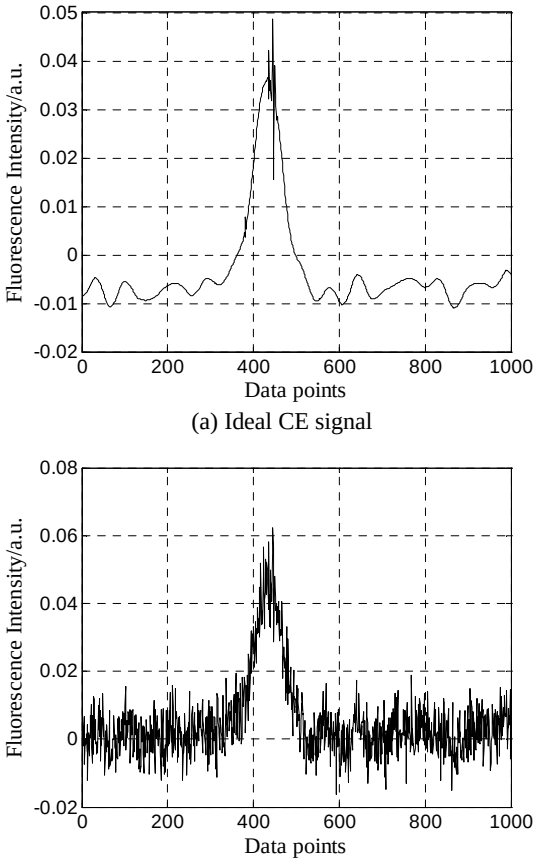


图1 毛细管电泳DNA测序系统

Fig. 1 Capillary electrophoresis DNA sequencing system

器；13：计算机）中，使用标准DNA样品pBR322/Hae III电泳时采集到的信号，经前期研究中优选的sym7小波去噪后，作为理想荧光信号，叠加式(8)产生的随机噪音来构建DNA测序仿真信号，对去噪效果进行更加客观和准确地评估.

由于目前毛细管电泳激光诱导荧光信号的信噪比在3以上被认为是可靠的信号，因此在仿真研究中，着重分析最弱的荧光信号峰，设置其信噪比为3[11]，理想信号和含噪信号如图2.



(a) Ideal CE signal

(b) Simulated noisy CE signal (*SNR*=3)

图 2 理想的 CE 信号和仿真的含噪 CE 信号

Fig. 2 Ideal CE signal and Simulated noisy CE signal

2.2 小波基函数的选择与仿真信号的去噪处理

由于小波分析中用到的小波基函数具有多样性，在DNA样品 CE-LIF信号去噪的运用中，主要从经过小波变换处理后所获得的电泳信号曲线不失真的角度来选择小波基函数.考虑到信号处理的时效性，在信号分解层数（$lev=5$）和使用固定格式软阈值的条件下，比较研究了Haar、Daubechies(1-10)、Symlets(2-8)、Coieflet(1-5)和Biorthogonal(1.1-6.8)等小波函数对信号的分解、去噪和重构的影响.发现用Daubechies8 (db8)、Symlets7 (sym7)、Coieflet（coif4, coif5）和Biorthogonal（bior3.5）小波去噪处理后，信号的峰形良好，且噪音基本被去除.

为了进一步确定适合于CE-LIF信号去噪的小波基函数，我们在实验中以去噪后的信号与理想信号的峰位置误差和峰高误差的大小、均方根误差(*RMSE*)的大小以及*SNR*的大小来确定最佳的小波基函数，数值比较见表1和表2.

表 1 不同小波基去噪效果比较（峰位与峰高）

	峰位置	峰位置误差/%	峰高	峰高误差/%
理想信号	441	0	0.0486	0
db8	443	0.45	0.0483	-0.6173
sym7	443	0.45	0.0481	-1.0288
coif4	437	-0.91	0.0479	-1.4403
coif5	439	-0.45	0.0469	-3.4979
bior3.5	438	-0.68	0.0495	1.8516

表 2 不同小波基去噪效果比较（信噪比与均方根误差）

	SNR	*RMSE*
理想信号	∞	0
db8	15.9816	0.0018
sym7	15.8567	0.0019
coif4	15.3955	0.0021
coif5	15.7562	0.0020
bior3.5	15.2697	0.0019

从表1和表2可以直观地看出：CE-LIF信号经db8小波基函数去噪处理后，峰位置、峰高的误差最小，*SNR*最高，*RMSE*最小，为最优的小波基函数，sym7小波次之.

3 实验电泳荧光信号的去噪

为了进一步确定使用db8小波基函数去噪处理的有效性，我们对DNA标样电泳荧光信号进行了处理.使用自制的毛细管电泳激光诱导荧光共焦检测DNA测序装置（图1），采用标准DNA样品pBR322/Hae III进行电泳实验，采集到的信号原始图谱如图3，图4为使用db8小波基函数去噪处理后的信号波形，从图4可以看出，经db8小波去噪后，信号的基线平稳，荧光峰清晰可辨，信噪比提高了5~6倍.

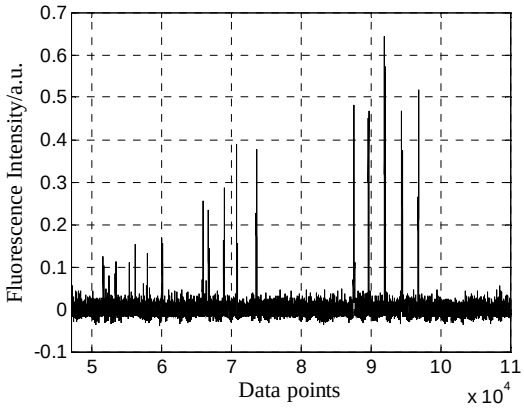


图 3 DNA 片段毛细管电泳原始信号

Fig. 3 Original CE signal of DNA fragment

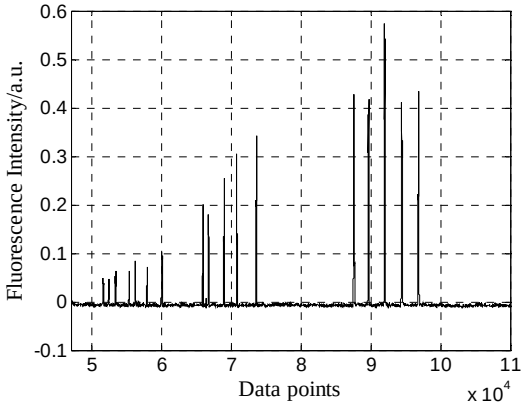


图 4 去噪后的 DNA 片段毛细管电泳信号

Fig. 4 Denoised CE Signal of DNA Fragment

4 结论

在一维信号的小波去噪处理中，选择准确的噪音模型，合适的小波基函数、分解层数和去噪阈值直接关系到信号去噪处理的质量.本文通过对真实CE荧光信号叠加随机噪音去噪的仿真研究，对典型的小波基函数进行筛选.仿真研究的结果表明：选择db8小波基函数、分解层数（*lev* = 5）与使用固定格式软阈值，可以有效地去除CE荧光信号的噪音，提高信号的信噪比；去噪处理后，信号的峰位置准确、峰高误差小，提高了CE荧光信号分析的准确度,在对实际DNA片段电泳荧光信号的处理中也直观地表现了这些优点.

参考文献

[1] BARCLAY V J, BONNER R F. Application of wavelet transforms to experimental spectra: smoothing, denoising, and data set compression[J]. *Analytical Chemistry*, 1997, **69**(1): 78-90.

[2] MALLAT S, HWANG W L. Singularity detection and processing with wavelets[J]. *IEEE Transactions on Information Theory*, 1992, **38**(2): 617-643.

[3] MALLAT S, ZHONG S. Characterization of signals from multiscale edges [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, **14**(7): 1019-1033.

[4] MALLAT S. A theory for multiresolution signal decomposition: the wavelet representation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, **11**(7): 674-693.

[5] ZHENG Hua, SHI Yan, WANG Jie, *et al*. Analyzing the wavelet denoising of electrophoresis fluorescence signals for DNA sequencing[J]. *Optical Instruments*, 2007, **29**(2): 17-21.
郑华，石岩，汪洁，等. DNA测序电泳荧光信号的小波去噪分析. 光学仪器，2007, **29**(2): 17-21.

[6] WANG Li-shi, YANG Xiao-yun, XI Shao-feng, *et al*. Wavelet Smoothing and Denoising to Process Capillary Electrophoresis Signals[J]. *Chemical Journal of Chinese Universities*, 1999, **20**(3): 383-386.
王立世，杨晓云，席绍峰，等. 毛细管电泳信号的小波平滑与去噪[J]. 高等学校化学学报，1999, **20**(3): 383-386.

[7] ZHONG Hong-bo, LI Guan-bin, LIU Hui, *et al*. Application of Dyadic Wavelet Transform Modulus Maximum Method to Denoising of Capillary Electrophoresis Signals[J]. *Chemical Journal of Chinese Universities*, 2002, **23**(5): 796-800.
仲红波，李关宾，刘辉，等. 利用二进样条小波模极大方法消除毛细管电泳信号噪声的研究[J]. 高等学校化学学报，2002, **23**(5): 796-800.

[8] WANG Jie, WANG Li-qiang, SHI Yan, *et al*. Detection Limit of DNA Analyzer[J]. *Acta Photonica Sinica*, 2008, **37**(3): 543-546.
汪洁，王立强，石岩，等. DNA分析仪的灵敏度分析[J]. 光子学报，2008, **37**(3): 543-546.

[9] WANG Jie, WANG Li-qiang, SHI Yan, *et al*. An analysis of stray light in capillary array electrophoresis with laser induced fluorescence detection[J]. *Acta Photonica Sinica*, 2008, **37**(2): 360-363.
汪洁，王立强，石岩，等. 毛细管阵列电泳检测过程中的杂散光分析[J]. 光子学报，2008，**37**(2): 360-363.

[10] SHI Yan, WANG Li-qiang, ZHENG Hua, *et al*. Signal to Noise Ratio Analysis for DNA Detecting System by Capillary Electrophoresis and Laser-induced Fluorescence[J]. *Acta Photonica Sinica*, 2008, **37**(7): 1446-1449.
石岩，王立强，郑华，等. 激光诱导荧光毛细管电泳DNA检测系统信噪比分析[J]. 光子学报，2008，**37**(7): 1446-1449.

A Novel Wavelet Analyzing Method for Signals Denoising of DNA Sequencing

ZHENG Hua[1], TANG Lei[1], LU Zu-kang[2]

(1 *Institute of Laser and Optoelectronics Technology， Fujian Provincial Key Laboratory for Photonics Technology, Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education, Fujian Normal University, Fuzhou* 350007, *China*)

（2 *State Key Laboratory of Modern Optical Instrumentation, NERC for Optical Instrument, Zhejiang University, Hangzhou,* 310027, *China*）

Received date: 2008-11-03

Abstract: In order to construct the same peak model as that in experiment and evaluate the denoising algorithm precisely, a novel method was presented. The random noise was added to a real denoised DNA signal to simulate a noisy sequencing signal, thus the denoising efficiency could be evaluated accurately. The denoising results indicate that using db8 wavelet base, decomposition level at 5 and using fixed form soft threshold can effectively reduce the noise. When the same algorithm was applied to the experimental DNA sequencing data, the results were more credible than that obtained through other algorithms based on the Gaussian peak model.

Key words: DNA sequencing; Fluorescence signal; Wavelet analysis; Denoising

ZHENG Hua was born in 1975. He obtained his Ph.D. degree in optical engineering in 2007 from Zhejiang University. Now he is a lecturer at Fujian Normal University, and his main research interests focus on the optical instrument with fluorescence detection.