

УДК 621.384.8

© И. В. Заруцкий, В. В. Манойлов

ПРЕДВАРИТЕЛЬНАЯ ОЧИСТКА МАСС-СПЕКТРОМЕТРИЧЕСКИХ СИГНАЛОВ ОТ ШУМОВ С ПОМОЩЬЮ ВЕЙВЛЕТ-ФИЛЬТРОВ

Рассмотрены возможности прямого и обратного вейвлет-преобразований для очистки масс-спектрометрических сигналов от шумов. Использовано многоуровневое преобразование исходного сигнала с целью установления оптимальной полосы низкочастотного фильтра при сохранении разрешения. Критерием оптимальности выбран уровень энтропии, который быстро уменьшается на первых этапах аппроксимации и детализации, а затем его уменьшение сильно замедляется. Приведены примеры фильтрации реального масс-спектрометрического сигнала предлагаемым методом.

ПОСТАНОВА ЗАДАЧИ

Одной из причин, затрудняющих анализ любого сигнала, является наличие шумов. Существует несколько общепринятых методов удаления помех из исходного сигнала: применение оконного Фурье-преобразования [1, 2], полиномиальное сглаживание и относительно недавно появившийся метод, основанный на использовании вейвлет-преобразования. По сравнению с другими методами вейвлет-очистка сигналов имеет ряд преимуществ, заключающихся, в первую очередь, в оптимальной частотно-временной локализации, что позволяет удалять помеху вблизи некоторой точки t , не искажая при этом самого сигнала. Строгое математическое обоснование метода очистки сигнала от шумов с помощью вейвлетов и сравнение его с другими методами было рассмотрено в работе [3].

Основная идея очистки сигнала от шума заключается в разложении сигнала на вейвлеты, идентификации компонент шума и восстановления сигнала без этих компонент.

ФИЛЬТРАЦИЯ МАСС-СПЕКТРОМЕТРИЧЕСКИХ СИГНАЛОВ МЕТОДОМ ПРЯМОГО И ОБРАТНОГО ДИСКРЕТНОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЙ

Первый этап фильтрации: аппроксимация и детализация. Для большинства масс-спектрометрических сигналов низкочастотная составляющая анализируемого сигнала является наиболее важной. Это то, что дает возможность идентифицировать сигнал. Высокочастотная составляющая — это шум. В вейвлет-анализе говорят об аппроксимации и деталях. Аппроксимация является высокой шкалой — низкочастотной составляющей сигнала. Детализация — это низкая шкала, высокочастотная составляющая. Рассмотрим основные черты фильтрации с помощью двух частей вейв-

лет-преобразования. На рис. 1 представлено разделение исходного сигнала на части с помощью высокочастотного и низкочастотного фильтров.

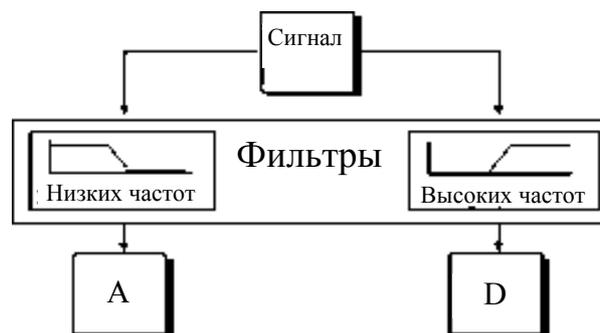


Рис. 1. Разделение исходного сигнала на две части

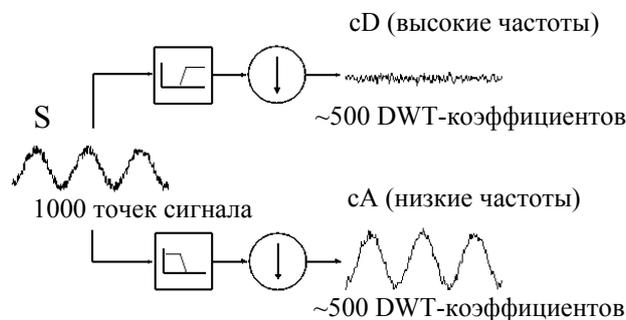


Рис. 2. Получение из 1000 точек исходного сигнала двух частей из 500 точек для низкочастотной и высокочастотной составляющих при дискретном вейвлет-преобразовании (DWT)

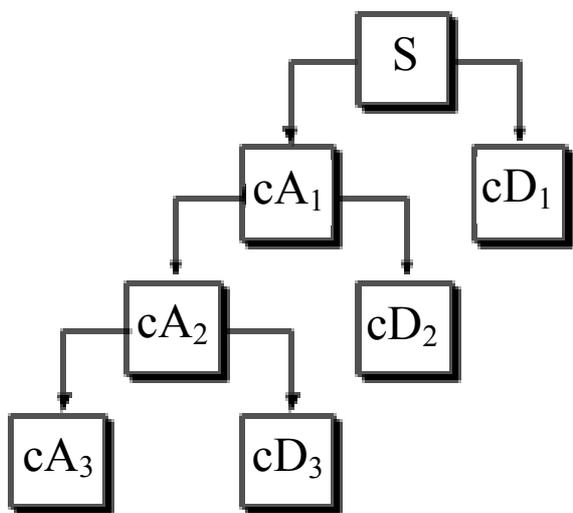


Рис. 3. Вейвлет-декомпозиционное дерево

Пусть исходный сигнал содержит 1000 точек. Такой сигнал проходит через два дополнительных фильтра, и появляются два сигнала. При дискретном вейвлет-преобразовании из 1000 точек в каждой части появляется 500 точек соответственно cA и cD (рис. 2).

**МНОГОУРОВНЕВОЕ
ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ**

Процесс представления исходного сигнала с помощью вейвлет-коэффициентов при разделении на низкочастотную и высокочастотную части называют декомпозицией. Процесс декомпозиции может быть итеративным. Такой процесс называют вейвлет-декомпозиционным деревом (рис. 3).

ВЕЙВЛЕТ-РЕКОНСТРУКЦИЯ

Выше был рассмотрен процесс использования дискретного вейвлет-преобразования для анализа или декомпозиции сигнала. Другая составляющая обработки сигнала состоит в том, как полученные при декомпозиции составляющие сигнала могут быть обратно соединены вместе для получения исходного сигнала без потери информации. Этот процесс называется реконструкцией или синтезом. Комплекс математических операций, который выполняет процесс синтеза, называется обратным дискретным вейвлет-преобразованием (рис. 4).

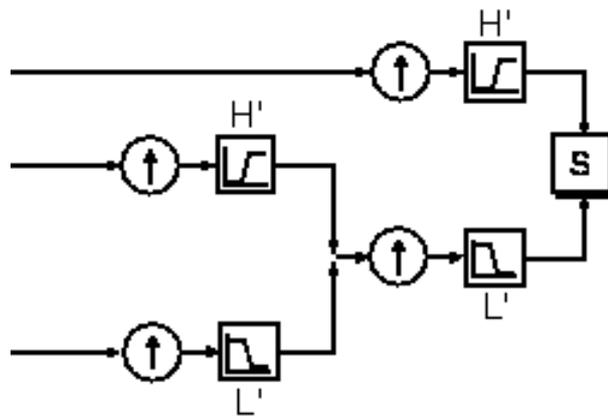


Рис. 4. Реконструкция сигнала из вейвлет коэффициентов после декомпозиции.

⬆ — означает увеличение количества точек за счет добавления в выборку дополнительных нулей; H' и L' — соответственно реконструкционные фильтры верхних и нижних частот

Вейвлет-анализ (декомпозиция) производит фильтрацию и уменьшает количество точек в каждой из отфильтрованных частей (downsampling). Процесс вейвлет-реконструкции (синтеза) состоит из процедуры увеличения количества точек (upsampling) и фильтрации. Синтез исходного сигнала осуществляется из вейвлет-коэффициентов, полученных по результатам анализа. Upsampling — это процесс увеличения количества точек сигнала путем введения нулей между отсчетами.

РЕКОНСТРУКЦИОННЫЕ ФИЛЬТРЫ

Правильный выбор реконструкционного фильтра является решающим моментом в достижении наименьшей разности между исходным и синтезируемым сигналами [4]. Низкочастотный и высокочастотные декомпозиционные фильтры (L и H), а также связанные с ними реконструкционные фильтры (L' и H') образуют систему, которая называется квадратурными зеркальными фильтрами (рис. 5).

**ВЫБОР ОПТИМАЛЬНОГО КОЛИЧЕСТВА
УРОВНЕЙ ДЕКОМПОЗИЦИИ**

Выбор оптимального количества уровней декомпозиции основан на свойствах данного ортогонального вейвлета, с помощью которого было произведено разложение исходного сигнала.

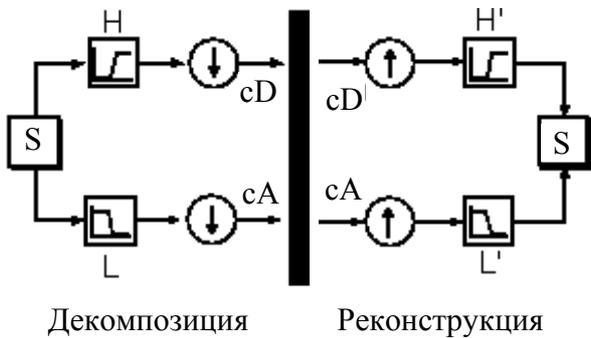


Рис. 5. Квадратурные зеркальные фильтры

Исходный сигнал длиной $N = 2L$ может быть α -"растянут" различными путями (где α — количество бинарных ярусов полного бинарного дерева глубиной L). В результате $\alpha \leq 2^{N/2}$ [6]. Это число может быть очень большим, и поэтому конечное количество вычислений может быть неуправляемым. В связи с этим важным является установить критерий эффективности алгоритма при установлении оптимального количества уровней декомпозиции. Обычно ищется минимум функционала. Для решения данной задачи хорошо подходят классические критерии, основанные на минимуме энтропии. Энтропия является общей концепцией во многих областях знаний, главным образом в обработке сигналов [7]. В соответствии с описанными в [5] свойствами энтропии для сигналов, представляемых в виде ортогональных разложений, справедливо следующее выражение для энтропии сигнала S :

$$E(S) = \sum_i E(s_i), \quad (1)$$

где s_i — коэффициенты разложения сигнала S в ортонормированном базисе.

В [5] приведен список из четырех типов энтропии для коэффициентов разложения сигнала по ортонормированному базису.

1 тип. Ненормированная энтропия по Шеннону:

$$E1(s_i) = -s_i^2 \log(s_i^2) \quad (2)$$

и поэтому

$$E1(S) = \sum_i s_i^2 \log(s_i^2) \quad (3)$$

при соглашении для $0 \rightarrow \log(0) = 0$.

2 тип. Концентрация по норме p при $p < 1$ (в данной работе $p = 0.5$):

$$E2(s_i) = |s_i|^p \quad (4)$$

и поэтому

$$E2(S) = \sum_i |s_i|^p = \|s\|_p^p. \quad (5)$$

3 тип. Логарифм "энергии" энтропии:

$$E3(s_i) = \log(s_i^2) \quad (6)$$

и поэтому

$$E3(S) = \sum_i \log(s_i^2) \quad (7)$$

при соглашении для $0 \rightarrow \log(0) = 0$.

4 тип. Пороговая энтропия:

$$E4(s_i) = 1, \text{ если } |s_i| > \varepsilon ;$$

$$E4(S) = K,$$

где K — количество раз превышения коэффициентами вейвлет-разложения порогового уровня ε .

Произведем оценку количества уровней декомпозиции для масс-спектра, представленного на рис. 6. Количество отсчетов этого сигнала $N = 6000$. В работе [2] описан алгоритм фильтрации того же сигнала, который представлен на рис. 6. Тот алгоритм основан на фильтре нижних частот Чебышева, поэтому возможно сравнение методов.

Результаты вычисления энтропии для различных уровней декомпозиции сведены в таблицу. В качестве материнских вейвлет-функций использовались функции Добеши [3]. Приведем примеры различных уровней вейвлет-декомпозиции для исходного сигнала (рис. 6).

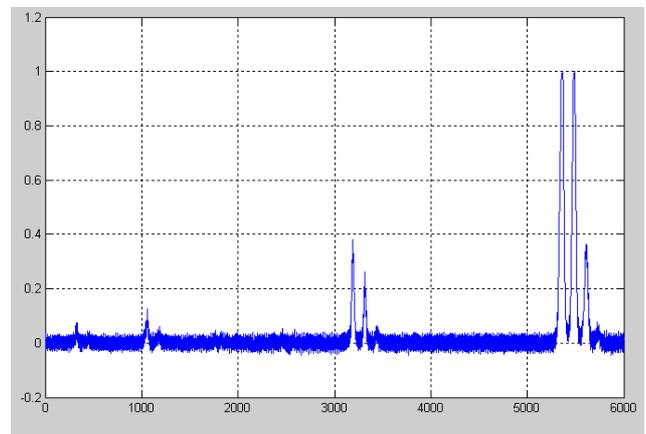


Рис. 6. Исходный сигнал в нормализованном виде

Значения различных типов энтропии при декомпозиции масс-спектра (рис. 6)

Уровень декомпозиции	Энтропия типа 1 (Шеннон)	Энтропия типа 2 (концентрация по норме, $p = 0.5$)	Энтропия типа 3 (логарифм "энергии")
Нулевой (исходный сигнал)	73.12	92	-4.8e4
Первый	36.63	47	-2.38e4
Второй	16.85	23	-1.27e4
Третий	7.14	12	-7.99e3
Четвертый	3.58	7	-4.06e3
Пятый	1.64	4	-2.06e3
Шестой	1.50	4	-1.04e3
Седьмой	1.05	2	-559

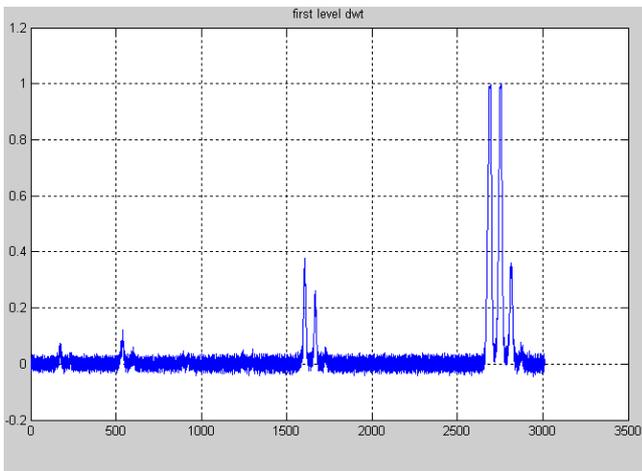


Рис. 7. Первый уровень DWT сигнала в нормализованном виде (образовано $N/2$ точек)

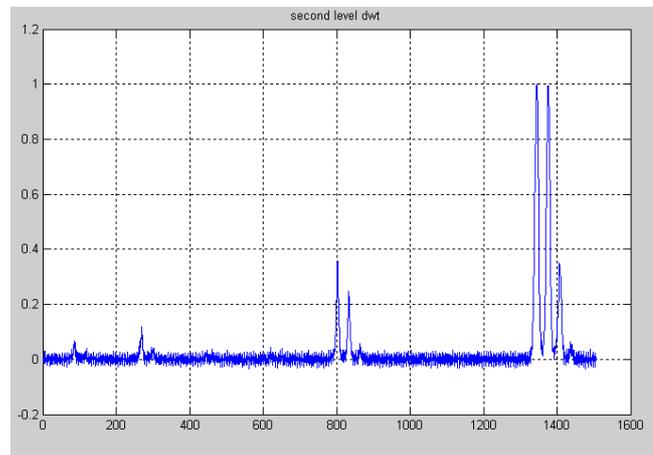


Рис. 8. Второй уровень DWT сигнала в нормализованном виде (образовано $N/4$ точек и повысилась относительная сигнал/шум)

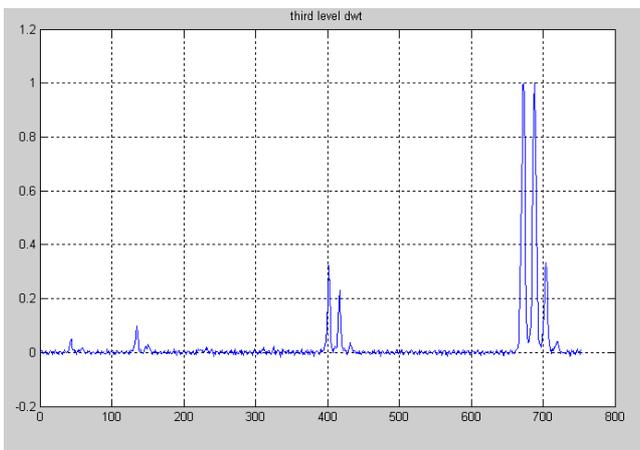


Рис. 9. Третий уровень DWT сигнала в нормализованном виде (образовано $N/8$ точек и повысилась относительная сигнал/шум)

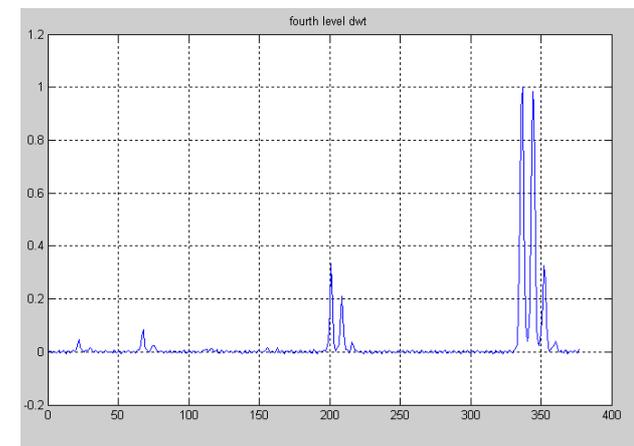


Рис. 10. Четвертый уровень DWT сигнала в нормализованном виде (образовано $N/16$ точек и повысилась относительная сигнал/шум)

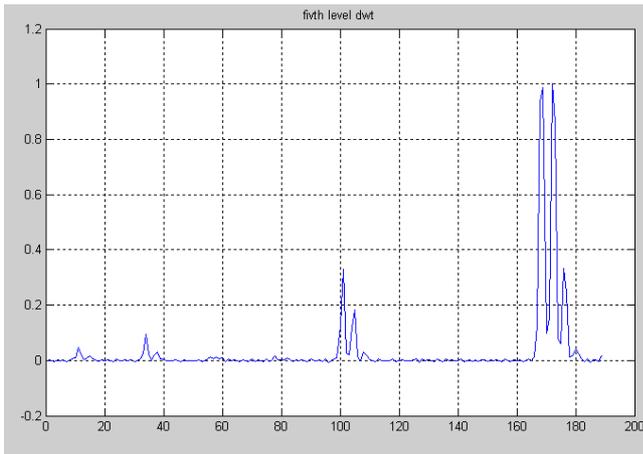


Рис. 11. Пятый уровень DWT сигнала в нормализованном виде (образовано $N/32$ точек и повысилось отношение сигнал/шум, ухудшилось разрешение)

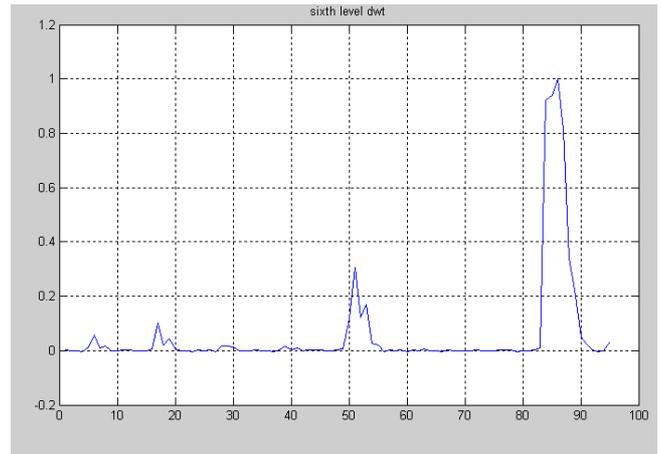


Рис. 12. Шестой уровень DWT сигнала в нормализованном виде (образовано $N/64$ точек, повысилось отношение сигнал/шум, но и сильно ухудшилось разрешение)

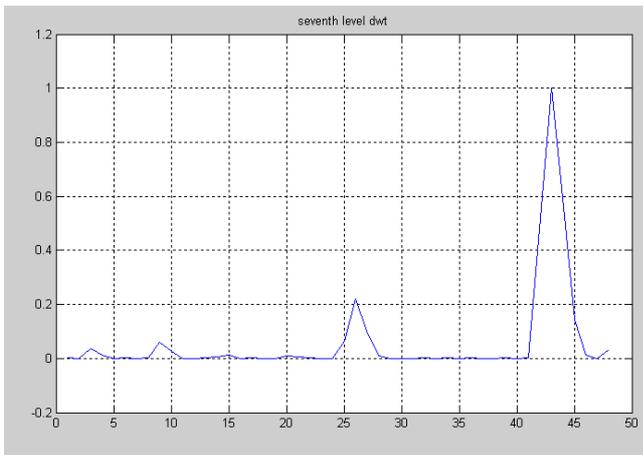


Рис. 13. Седьмой уровень DWT сигнала в нормализованном виде (образовано $N/128$ точек, повысилось отношение сигнал/шум, но и очень сильно ухудшилось разрешение)

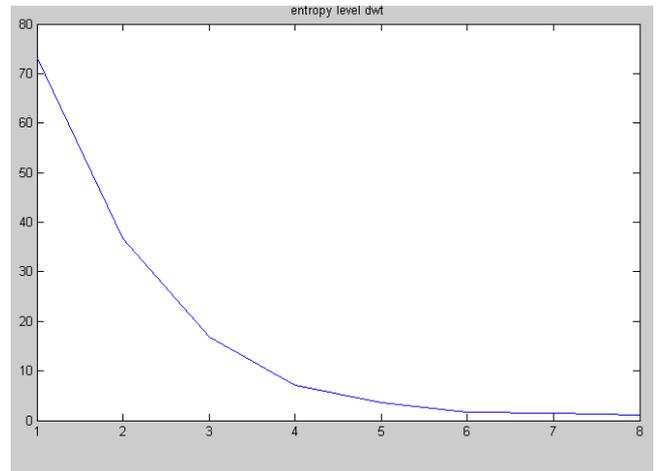


Рис. 14. Уменьшение энтропии по Шеннону в зависимости от уровня вейвлет-декомпозиции

Иначе говоря, представим вейвлет-декомпозиционное дерево. На рис. 7–13 представлены аппроксимирующие части различных уровней вейвлет-декомпозиции.

Как видно из рис. 10 и 11, наиболее приемлемыми уровнями вейвлет-декомпозиции являются уровни 4 и 5, т. к. они дают максимальное отношение сигнал/шум и не искажают разрешающую способность. Этот же вывод можно сделать из характера уменьшения энтропии, представленного в таблице и на рис. 14. Исходя из этих данных, видно, что после 5-го уровня энтропия падает значительно медленнее, чем в промежутке от 1-го до 5-го уровня.

Сравнение результатов данной работы и рабо-

ты [2] дает основание сделать заключение, представленное ниже.

ЗАКЛЮЧЕНИЕ

На основе минимизации уровня энтропии предложен алгоритм выбора оптимального количества уровней декомпозиции ортогонального вейвлет-преобразования. Оптимальный уровень декомпозиции дает максимальное отношение сигнал/шум, и в обрабатываемом сигнале не искажается разрешающая способность. Сравнение качества фильтрации масс-спектрометрических сигналов с помощью фильтров Чебышева и с помощью вейвлет-декомпозиции показывает, что

оба метода фильтрации дают примерно равные результаты в смысле отношения сигнал/шум. Преимуществом вейвлет-декомпозиции при установлении с помощью минимума энтропии уровня декомпозиции является минимизация количества вычислений по сравнению с количеством вычислений, требуемых для реализации фильтров Чебышева применительно ко всему массиву данных.

СПИСОК ЛИТЕРАТУРЫ

1. Макс Ж. Теория и техника обработки сигналов при физических измерениях. Т. 1. М.: Мир, 1983. 280 с.
2. Заруцкий И.В., Манойлов В.В. Аппаратно-программный комплекс транспортабельного хромато-масс-спектрометра ТХМС // Научное приборостроение. 2003. Т. 13, № 4. С. 47–54.
3. Добеши И. Десять лекций по вейвлетам (пер. с англ.). Ижевск: НИЦ "Регулярная и хаотическая динамика", 2001. 464 с.
4. Strang, G., Nguyen T. Wavelets and filter banks. Wellesley-Cambridge Press., 1996. 240 p.
5. Coifman R.R., Wickerhauser M.V. Entropy-based algorithms for best basis selection // IEEE Trans. on Inform. Theory. 1992. V. 38, N 2. P. 713–718.
6. Mallat S. A wavelet tour of signal processing. Academic Press, 2000. 637 p.
7. Dohodo D. // IEEE Trans. on Inform. Theory. 1995. V 41. P. 613–627.

Институт аналитического приборостроения РАН, Санкт-Петербург

Материал поступил в редакцию 18.10.2006.

PRELIMINARY ELIMINATION OF NOISE FROM MASS SPECTRA BY WAVELET FILTERS

I. V. Zarutsky, V. V. Manoilov

Institute for Analytical Instrumentation RAS, Saint-Petersburg

The paper analyzes applicability of the direct and inverse transformations to the problem of removing noises from mass spectra. To find the low-frequency filter band that ensure optimal results while the resolution remains constant, a multilevel conversion of initial signal is used. As the optimality criterion, an entropy level was taken, which quickly decreases on the early stages of approximation and detailing and decreases much slower on the later stages. The paper presents the cases of filtering an actual mass spectrum by the method suggested.