

С.А. Бабичев, Н.И. Бабенко, А.А. Дидык, В.И. Литвиненко,
А.А. Фефелов, С.В. Шкурдода

ФИЛЬТРАЦИЯ ХРОМАТОГРАМ С ПОМОЩЬЮ ВЕЙВЛЕТ – АНАЛИЗА С ИСПОЛЬЗОВАНИЕМ КРИТЕРИЯ ЭНТРОПИИ

В статье представлены теоретические разработки по использованию вейвлет-анализа для выделения шумовой составляющей хромато-масс-спектрометрического сигнала на этапе его предварительной обработки. Разработана методика, позволяющая оптимизировать задачу выбора параметров вейвлет-преобразования при построении цифровых фильтров очистки сигналов от шума. В качестве основных критериев оценки уровня шумовой составляющей при выборе структуры вейвлет-разложения предложено использовать критерии энтропии: “log energy” и “shannon”. Результаты, полученные в процессе проведения исследований, позволяют повысить разрешающую способность приборов при решении задач классификации и кластеризации объектов за счет уменьшения разброса параметров, характеризующих заданный объект.

Ключевые слова: вейвлет-анализ, хромато-масс-спектр, фильтрация, критерий энтропии.

1. Введение

В настоящее время вопросам обработки сигналов различной природы посвящено множество теоретических и практических исследований в различных областях науки и техники [1-4]. Современное состояние теории обработки нестационарных сигналов характеризуется дальнейшим развитием и совершенствованием статистических методов, а также бурным проникновением в неё новых физических идей и широким использованием классических и нестандартных разделов математики [5,6]. В последнее десятилетие для обработки нестационарных сигналов начали применять новый математический инструмент – вейвлет-преобразование [7-12]. Как показала практика, традиционный аппарат представления произвольных функций и сигналов в виде рядов Фурье оказывается малоэффективным для функций и сигналов с локальными особенностями. Это связано с тем, что базисная функция рядов Фурье – синусоида определена от $-\infty$ до $+\infty$ и по своей природе является гладкой и строго периодической функцией. Такая функция на практике в условиях ограничения числа членов ряда или спектра разложения принципиально не способна описывать произвольные сигналы и функ-

© Бабичев С.А., Бабенко Н.И., Дидык А.А.,
Литвиненко В.И., Фефелов А.А., Шкурдода С.В., 2010

ции.

По сравнению с Фурье-анализом, вейвлет-анализ сложных нестационарных сигналов обладает рядом неоспоримых преимуществ. Разнообразие набора типа функций, представляющих вейвлеты, существенно расширяет круг решаемых с их помощью задач. Вейвлет-анализ имеет большие возможности сжатия и фильтрации данных [13,14]. Помимо возможностей сжатия и фильтрации данных, анализ в базисе вейвлет-функций [15, 16] позволяет решать задачи идентификации, моделирования, аппроксимации стационарных и нестационарных процессов, исследовать вопросы наличия разрывов в производных, осуществлять поиск точек склеивания данных, отыскивать признаки фрактальности информации.

Масс хроматограмма представляет собой один из методов анализа веществ представленных как масс-спектрометрические данные в виде хроматограмм, где ось X представляет время, а ось Y представляет интенсивность сигнала. Набор данных содержит информацию о массах, однако это не графическое представление, а визуализация изменения интенсивности сигнала в течение времени. Наиболее распространенное использование этого представления данных являются случаи, когда масс-спектрометрия используется сочетается с той или иной формой хроматографии, такими, как жидкосная спектрометрия хромато-масс-спектрометрия или газовая масс-хроматография. В этом случае, ось X представляет время удерживания, аналогично как в любой другой хроматограмме. Ось Y представляет собой интенсивность сигнала или относительной интенсивности сигнала. Существует множество различных метрик для представления этой интенсивности, все зависит от того, какая конкретно информация исследуется каждым масс-спектре

Вопросам предобработки масс-спектрометрического анализа с использованием вейвлетов посвящены работы [17,18,19]. Авторы проводят исследования по анализу сигналов различной природы с использованием вейвлет-функций. Применение вейвлет анализа к масс-спектрометрическому анализу позволяет существенно повысить точность определения состава вещества и находящихся в нём примесей. Сложность обработки масс-спектрометрического сигнала заключается в том, что ряд параметров масс-спектра отличается от параметров сигнала

лов изучаемых в других областях их обработки. К таким областям относятся: аппаратные функции, которыми описываются формы пиков, возникающих в масс-спектрах, а также параметры шумов, присутствующие в масс-спектрометрических сигналах, которые обусловлены не только шумами, возникающими в электронных трактах, но и шумами, связанными с физическими процессами, происходящими в масс-спектрах.

Однако анализ литературных источников в данной предметной области позволяет сделать вывод, что задача определения оптимального уровня вейвлет-декомпозиции сигнала с учетом выбора типа вейвлета и значения порогового коэффициента обработки детализирующих коэффициентов, решение которой основано на количественных критериях, в настоящее время однозначного решения не имеет.

II. Постановка задачи и методы

Целью работы является разработка оптимальной системы фильтрации хромато-масс-спектрометрического сигнала, основанная на использовании вейвлет-функций и критерия энтропии. Исследования включают в себя:

- Определение оптимального уровня вейвлет-декомпозиции сигнала;
- Нахождение оптимального уровня порогового критерия обработки детализирующих коэффициентов;
- Выбор типа используемого вейвлета.

В качестве исследуемого сигнала использовалась хроматограмма вещества амфетамина, исходный вид которой представлен на Рис. 1. Анализ хроматограммы позволяет сделать вывод о достаточной зашумленности сигнала, что способствует увеличению погрешности классификации состава исследуемого вещества.

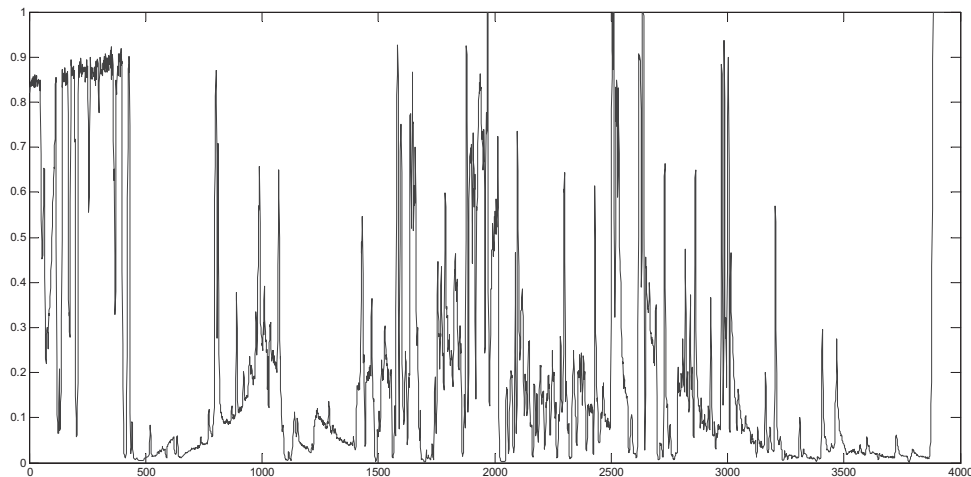


Рисунок 1 – Хроматограмма исследуемого сигнала

III. Вейвлет-предобработка сигнала

Реальный сигнал, излучаемый объектами различной природы, можно представить следующей формулой:

$$s(t) = f(t) + \delta e(t) \quad (1)$$

где $s(t)$ – исследуемый сигнал, $f(t)$ – полезный сигнал, $e(t)$ – шумовая составляющая сигнала, δ – коэффициент, определяющий уровень шумовой составляющей исследуемого сигнала.

Для описания функции $s(t)$ рассмотрим пространство $L^2(\mathbb{R})$ базисных функций $\psi(t)$, определённых на всей действительной оси и обладающих следующими свойствами:

ограниченность, т.е. эти функции должны обладать конечной энергией:

$$\int_{-\infty}^{+\infty} |\psi(t)|^2 dt < \infty \quad (2)$$

локализация, т.е. они должны быть определены на конечном интервале как во временной, так и в частотной областях. Для этого достаточно, чтобы выполнялись при $\varepsilon > 0$ условия:

$$|\psi(t)| \leq C \cdot (1 + |t|)^{-1-\varepsilon} \quad \text{и} \quad |\psi(\omega)| \leq C \cdot (1 + |\omega|)^{-1-\varepsilon}; \quad (3)$$

нулевое среднее, т.е. равенство нулю нулевого момента:

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (4)$$

Перечисленным выше условиям отвечает базис вейвлет-функций:

$$\psi(t) = \frac{1}{\sqrt{|a|}} \psi_0\left(\frac{t-b}{a}\right), \quad (5)$$

где параметр b задаёт положение вейвлета, а параметр a – его масштаб. При прямом непрерывном вейвлет-преобразовании сигнала $s(t)$ параметры a и b могут принимать любые значения в пределах области их определения. Вейвлет-коэффициенты при этом вычисляются по формуле:

$$C(a, b) = \int_{-\infty}^{+\infty} s(t) a^{-1/2} \psi\left(\frac{t-b}{a}\right) dt. \quad (6)$$

Основной недостаток непрерывного вейвлет-преобразования – низкая скорость обработки информации по причине высоких вычислительных затрат на её проведение. Этому недостатка лишено дискретное вейвлет-преобразование, в котором коэффициенты a и b принимают определённые дискретные значения на некотором множестве:

$$a = 2^j \quad \text{и} \quad b = k2^j, \quad (7)$$

где j и k – целые числа, параметр j является параметром масштаба, а k – уровень вейвлет-декомпозиции сигнала. При дискретных значениях a и b вейвлет-функция приобретает вид:

$$\psi_{j,k}(t) = a_0^{-j/2} \psi(a_0^{-j}t - k), \quad (8)$$

а прямое дискретное вейвлет-преобразование сводится к вычислению детализирующих коэффициентов:

$$d_{j,k} = \int_{-\infty}^{+\infty} a_0^{-j/2} \psi(a_0^{-j}t - k) s(t) dt. \quad (9)$$

Обратное дискретное вейвлет-преобразование для непрерывных сигналов задаётся с помощью того же базиса, как и прямое:

$$s(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d_{j,k} a_0^{-j/2} \psi(a_0^{-j}t - k), \quad (10)$$

где C_ψ – нормализующий коэффициент, определяемый по формуле:

$$C_\psi = \int_{-\infty}^{\infty} |\hat{\psi}(\omega)|^2 |\omega^{-1}| d\omega < \infty, \quad (11)$$

а $\hat{\psi}$ – фурье-образ вейвлет-функции ψ . Условие конечности константы C_ψ ограничивает класс используемых базисных вейвлет-функций, и при $C_\psi = 1$ окончательная формула для реконструкции сигнала записывается в виде:

$$s(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d_{j,k} a_0^{-j/2} \psi(a_0^{-j} t - k), \quad (12)$$

IV. Критерии оценки уровня зашумленности сигнала.

Для решения задачи оценки уровня шумовой компоненты хроматограммы использовался критерий энтропии. Энтропия – общая концепция оценки информативности во многих областях науки и техники [20, 21]. В области информатизации под энтропией можно понимать меру упорядочивания структурных элементов в системе. Нулевой энтропии соответствует отсутствие информации, в то время как высокой энтропии соответствует максимальное значение интересующей информации. При решении поставленной задачи под информацией понимается уровень шумовой составляющей исследуемого сигнала. В области оценки уровня шумовой составляющей сигнала наибольшее распространение получили следующие критерии энтропии:

Энтропия Шеннона:

$$E1(s) = -\sum_i s_i^2 \log(s_i^2), \quad (13)$$

при условии: $0 \cdot \log(0) = 0$.

Энтропия логарифм энергии:

$$E2(s) = \sum_i \log(s_i^2), \quad (14)$$

при условии: $\log(0) = 0$.

Норма пространства:

$$E3(s) = \sum_i |s_i|^p = \|s\|_p^p. \quad (15)$$

Пороговая энтропия:

$$E_4(s) = 1, \text{ если } |s_i| > \varepsilon; E_4(s) = 0, \text{ если } |s_i| < \varepsilon \quad (16)$$

Для оценки скорости изменения энтропии на каждом уровне вейвлет-декомпозиции предложено использовать критерий относительное изменение энтропии:

$$\frac{dE}{E_i} = \frac{|E_{i+1} - E_i|}{|E_i|} \quad (17)$$

где E_i - энтропия на i - м уровне вейвлет-декомпозиции сигнала, E_{i+1} - энтропия на $(i+1)$ -м уровне.

V. Определение оптимального уровня вейвлет-декомпозиции сигнала.

Оценки оптимального уровня вейвлет-декомпозиции сигнала производилась с использованием ортогонального вейвлета с компактным носителем симплета4 (sym4). Пороговое значение критерия трешолдинга детализирующих коэффициентов было принято равным 3. Исследование проводились в соответствии со следующим алгоритмом:

Декомпозиция сигнала в соответствии с пирамидальным алгоритмом Малла. Исходный сигнал s подаётся на фильтры декомпозиции низких (L) и высоких (H) частот, после чего с помощью операции децимации $\downarrow 2$ получаем коэффициенты аппроксимации (сA) на выходе фильтра низких частот и детализирующие коэффициенты (сD) на выходе фильтра высоких частот.

Трешолдинг детализирующих коэффициентов, в соответствии с которым значения детализирующих коэффициентов, меньшие порогового значения, обнуляются, а значения остальных детализирующих коэффициентов уменьшается на величину порогового значения.

Вейвлет-реконструкция сигнала, основанная на первоначальных аппроксимирующих коэффициентах и модифицированных детализирующих коэффициентах.

Вычисление относительного изменения энтропий логарифм энергии и Шеннона по формуле (17).

Увеличение уровня вейвлет-декомпозиции сигнала и выполнение пунктов 1 – 4 алгоритма.

На Рис. 2 показаны хроматограммы очищенного сигнала и шумовой компоненты на различных уровнях вейвлет-декомпозиции сигнала.

На Рис. 3 представлены графики зависимости относительного изменения энтропий от уровня вейвлет-декомпозиции.

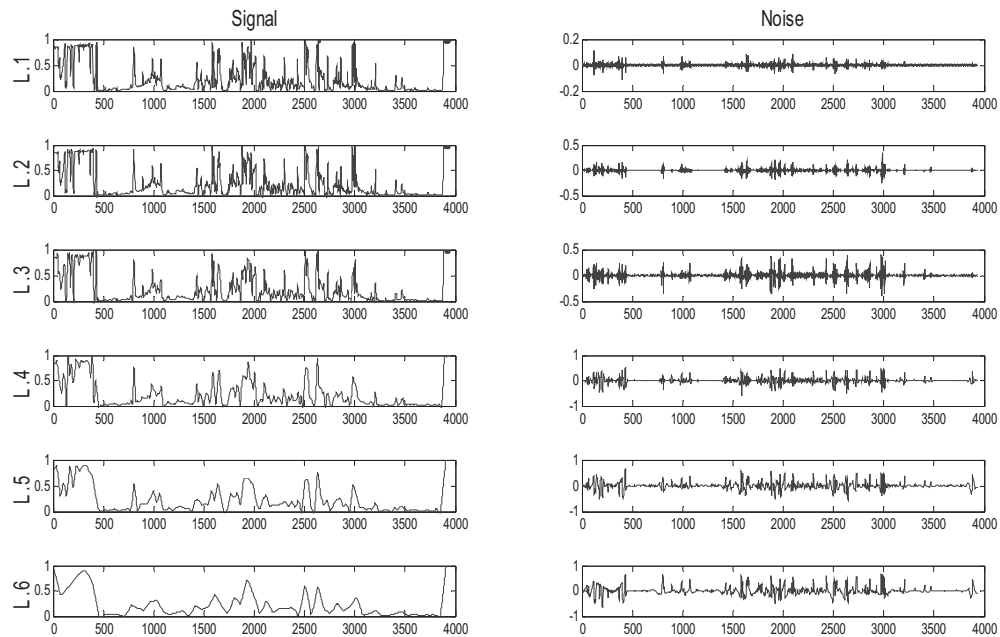


Рисунок 2 – Хроматограммы очищенного сигнала и шумовой составляющей на уровнях вейвлет-декомпозиции от 1(L.1) до 6(L.6)

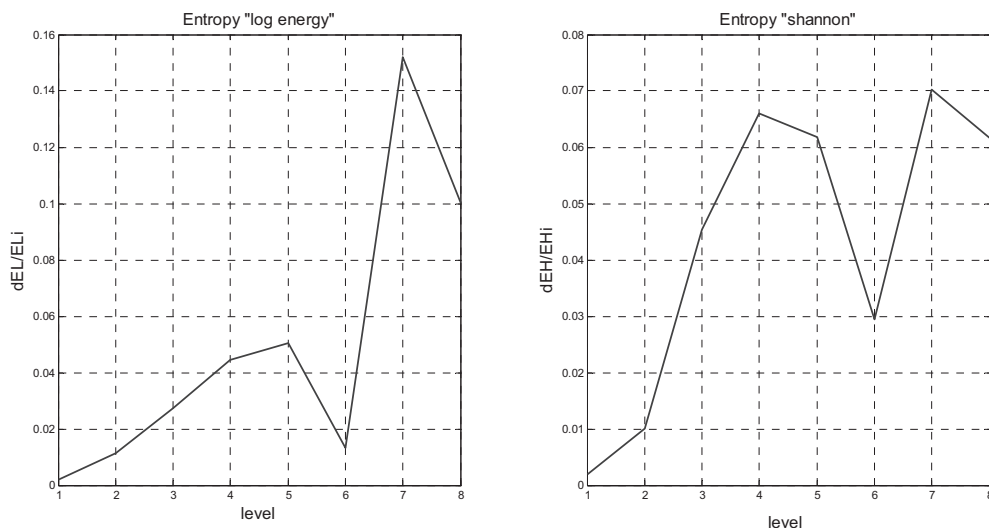


Рисунок 3 – Графики зависимости относительного изменения энтропий от уровня вейвлет-декомпозиции сигнала

Анализ хроматограмм на различных уровнях вейвлет-декомпозиции позволяет сделать вывод, что до четвёртого уровня разложения сигнала происходит поэтапное удаление шумовой составляющей. Это подтверждается графиками зависимости

относительного изменения энтропий от уровня вейвлет-декомпозиции сигнала. До четвертого уровня скорость изменения энтропий плавно возрастает. От четвертого до пятого уровня скорость изменения энтропий резко уменьшается, что свидетельствует об изменении характера информации, удаляемой с сигнала. В нашем случае можно сделать вывод, что после четвертого уровня вейвлет-декомпозиции начинают сглаживаться локальные особенности сигнала, что недопустимо. Данный вывод подтверждается графиками, изображенными на Рис. 2 (L.5 and L.6). Из анализа графиков, изображенных на Рис. 3 также следует, что для определения оптимального уровня вейвлет-декомпозиции сигнала более информативным является критерий энтропия Шеннона.

VI. Определение оптимального значения порогового коэффициента обработки детализирующих коэффициентов

Процедура трешолдинга связана с процессом обработки детализирующих коэффициентов, несущих основную информацию о шумовой составляющей сигнала, а также о его локальных особенностях. От выбора порогового уровня фона зависит качество шумоподавления сигнала, оцениваемое в виде отношения сигнал/шум. Задание малых значения порога сохраняет фон в коэффициентах детализации и поэтому приводит лишь к незначительному увеличению отношения сигнал/шум. При больших значениях порога можно потерять коэффициенты, которые несут существенную информацию.

Вейвлет-декомпозиция сигнала осуществлялась при помощи вейвлета симплета $\text{sym}4$. Значение порогового критерия обработки детализирующих коэффициентов на четвертом уровне вейвлет-декомпозиции сигнала увеличивалось от 1 до 6 с шагом 1. На каждом этапе рассчитывалась скорость изменения энтропий. Графики зависимости относительного изменения энтропий от значения порогового критерия представлены на Рис. 4.

Хромотограммы обработанных сигналов при различных значениях порогового коэффициента представлены на Рис. 5. Анализ графиков, изображенных на Рис. 4 позволяет сделать вывод, что увеличение значения порогового критерия больше 3 не имеет смысла, так как скорость изменения энтропии сигнала равна нулю. Кроме того, из анализа хромотограмм, изображенных на Рис. 5 также

следует, что значение порогового критерия не является определяющим в процессе очистки сигнала от шума.

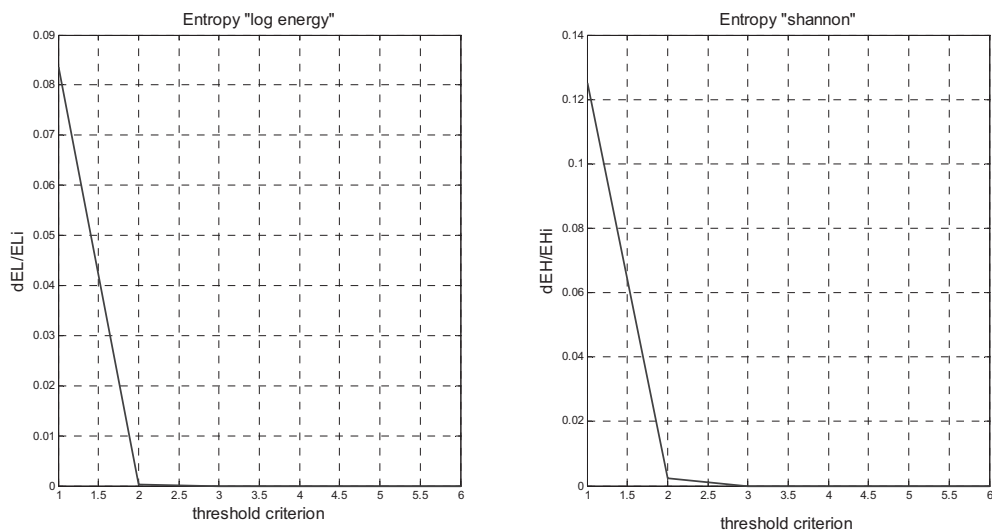


Рисунок 4 – Графики зависимости относительного изменения энтропий от значения критерия пороговой обработки детализирующих коэффициентов

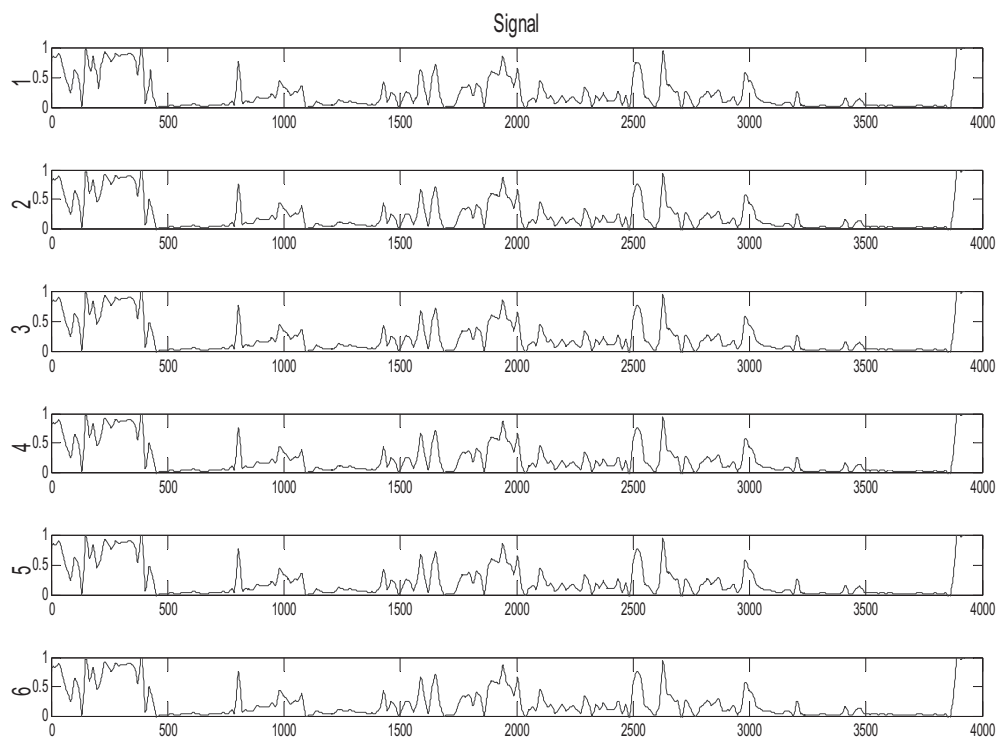


Рисунок 5 – Хромотограммы обработанного сигнала при различных значениях критерия пороговой обработки

VII. Определение оптимального типа используемого вейвлета.

Вейвлет-обработка хромотограммы производилась ортогональными вейвлетами с компактным носителем: Добеши (dbN), симплеты (symN) и койфлеты (coifN). Этот выбор определялся следующими обстоятельствами: вейвлеты симплеты и койфлеты практически симметричны, все типы данных вейвлетов имеют компактный носитель, их моменты обращаются в нуль и при дискретном преобразовании возможно использование быстрых алгоритмов, что сокращает время обработки информации.

При определении оптимального вейвлета использовался четвёртый уровень вейвлет-декомпозиции сигнала и значение порогового коэффициента обработки детализирующих коэффициентов принималось равным трём. В качестве критериев оценки уровня шумовой компоненты сигнала использовались абсолютное значение относительных изменений энтропии Шеннона и логарифм энергии, при этом за начальное значение энтропии принималась энтропия исходного сигнала. На первом этапе для каждого типа вейвлета находилось определялось наилучшее дерево из условия минимума энтропии. На втором этапе сопоставлялись значения критериев для каждого выбранного типа вейвлетов с целью определения оптимального вейвлета для обработки данного типа сигнала. Графики зависимости относительного изменения энтропий от типа вейвлета Добеши приведены на Рис.6.

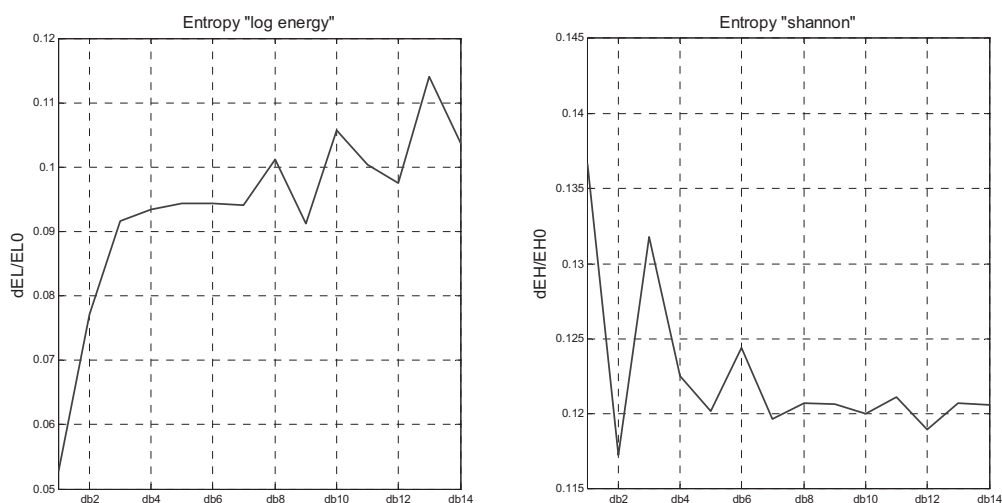


Рисунок 6 – Графики зависимости относительного изменения энтропий от типа вейвлета Добеши

Анализ графиков позволяет сделать вывод, что согласно критерию относительное изменение энтропии логарифм энергии наилучшее качество обработки сигнала достигается при использовании вейвлета Добеши 13, а согласно энтропии Шеннона – Добеши 1. Хромотограммы оригинального сигнала и сигналов, обработанных данными вейвлетами, приведены на Рис.7. Анализ хромотограмм позволяет сделать вывод о неэффективности применения критерия Шеннона для решения данной задачи, так как максимальное значение критерия соответствует искаженному сигналу, что недопустимо. Таким образом, выбор оптимального типа вейвлета будем производить на основании критерия энтропии логарифм энергии.

На Рис. 8 представлены графики относительного изменения энтропии логарифм энергии для разных типов вейвлетов симплета и койфлета. Анализ графиков позволяет сделать вывод, что наилучшее качество обработки сигнала получается при использовании симплета7 или койфлета4.

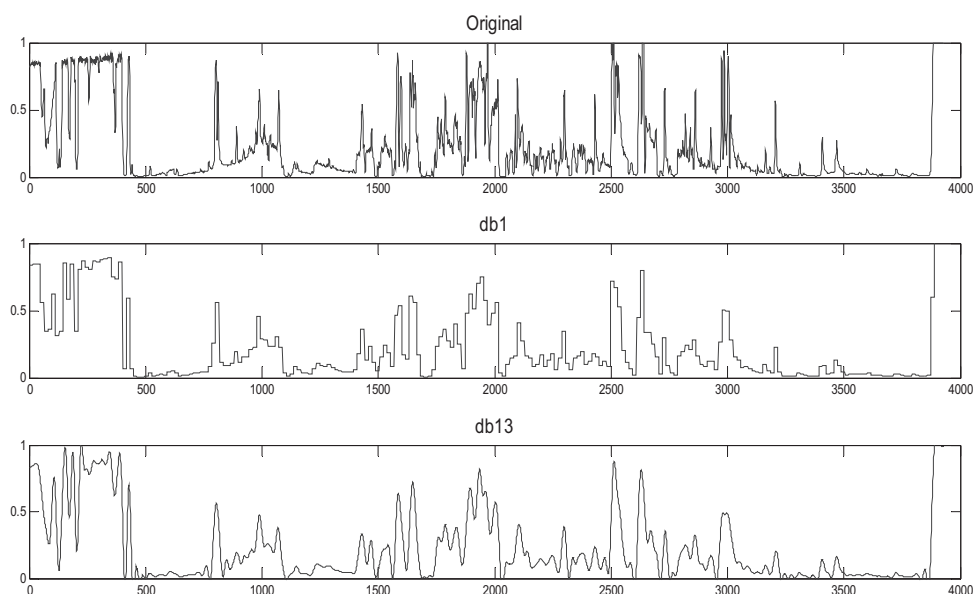


Рисунок 7 – Хромотограммы оригинального сигнала и обработанного вейвлетом db1 и db13

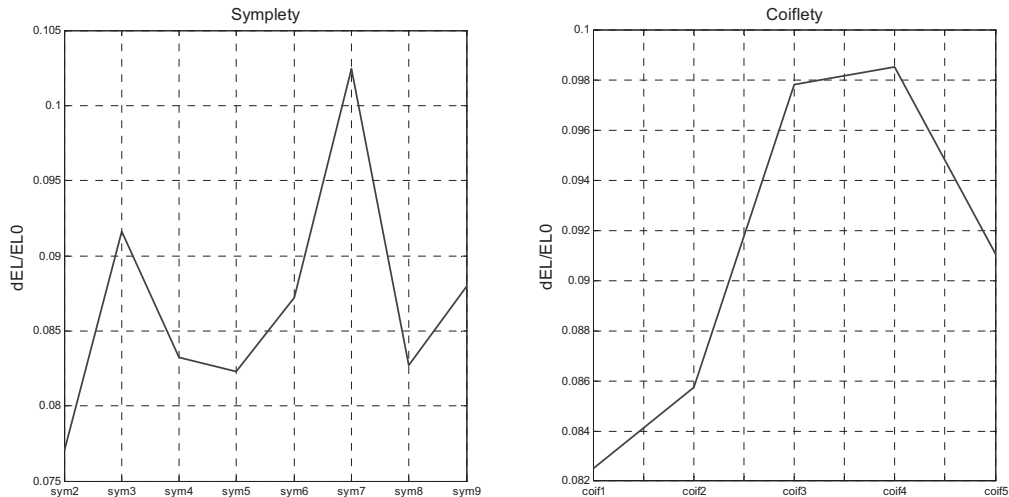


Рисунок 8 – Графики значений относительного изменения энтропии логарифм энергии в зависимости типа симплета и койфлета

На Рис. 9 представлены графики относительного изменения энтропии логарифм энергии при обработке сигнала вейвлетом db13, симплетом sym7 и койфлетом coif4. Максимальное изменение энтропии наблюдается при обработке сигнала вейвлетом Добеши 13, однако следует отметить, что симплет 7 также даёт высокое качество очистки сигнала от шумовой составляющей. На Рис. 10 показаны хромотограммы оригинального сигнала и сигнала, обработанного при помощи вейвлета Добеши 13 и симплета 7. Анализ хромотограмм подтверждает высокое качество очистки сигнала от шума.

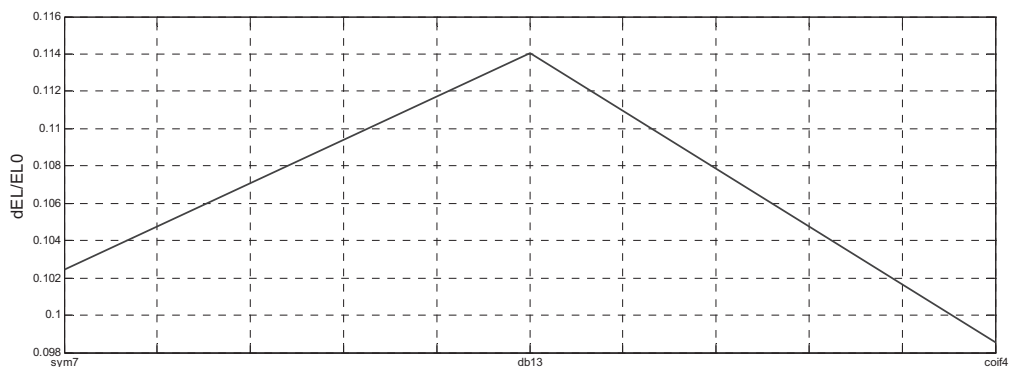


Рисунок 9 – Графики зависимости относительного изменения энтропий от типа используемого вейвлета

VIII. Заключение

В работе представлены исследования по созданию системы очистки сигнала от шума при помощи вейвлет-анализа и критериев энтропии. Показано, что для выбора оптимального уровня вейвлет-

декомпозиции сигнала эффективными являются критерии относительные изменения энтропий Шеннона и логарифм энергии, при этом за начальную энтропию принимается энтропия на n -ом уровне вейвлет-декомпозиции. Проведенные исследования так же показали, что для решения поставленной задачи более предпочтительной является энтропия Шеннона.

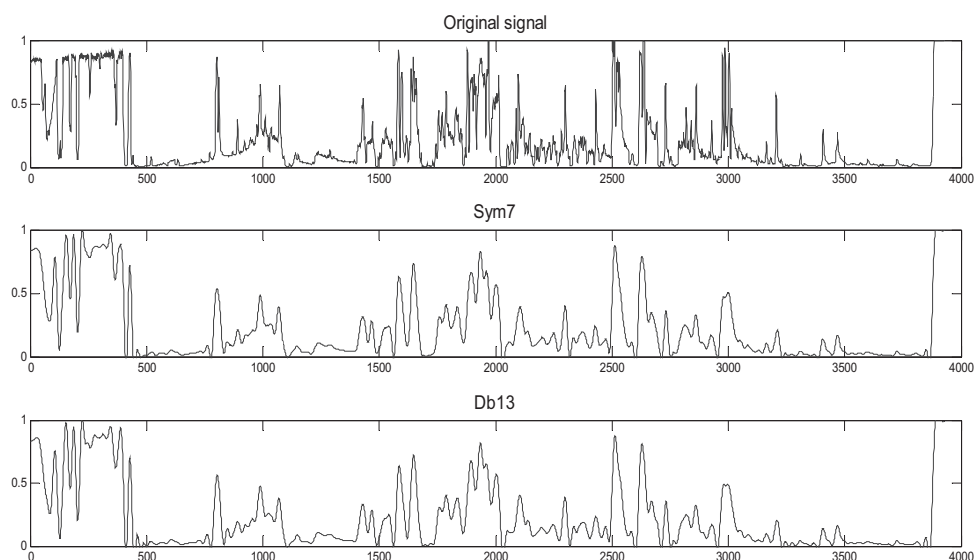


Рисунок 10 – Хроматограммы оригинального сигнала и сигнала, обработанного симплетом `sym7` и вейвлетом Добеши `db13`

При оценке степени влияния коэффициента пороговой обработки детализирующих коэффициентов также использовались критерии относительное изменение энтропий, при этом было показано, что выбор порогового значения коэффициента не является определяющим в системе обработки сигнала, и значение критерия больше трёх нецелесообразно.

Для выбора типа вейвлета за начальную энтропию принималась энтропия исходного сигнала, а оптимальный уровень обработки соответствовал наибольшему отклонению энтропии от исходного значения при предварительно определённом оптимальном уровне вейвлет-декомпозиции сигнала и значении порогового коэффициента обработки детализирующих коэффициентов.

По результатам исследований можно сделать вывод, что оптимальная степень фильтрации хромато-масс-спектрометрического сигнала достигается при использовании вейвлета Добеши 13 или симплета 7 при четвёртом уровне вейвлет-декомпозиции сигнала и значении

порогового критерия обработки детализирующих коэффициентов, равного 3. При этом использовалась мягкая система очистки сигнала.

В перспективе авторы планируют использование проведенных исследований на предварительном этапе решения задач классификации признаков объектов сложной природы с целью повышения разрешающей способности диагностических систем.

ЛИТЕРАТУРА

1. L. Hualou, L. Qiu-Hua, J. Chen, Application of the empirical mode decomposition to the analysis of esophageal manometric data in gastroesophageal reflux disease, *IEEE Trans. Biol. Eng.* 52 (10) (2005) 620–623.
2. E. Delechelle, J. Lemoine, O. Niang, Empirical mode decomposition: an analytical approach for sifting process, *IEEE Signal Process. Lett.* 12 (11) (2005) 764–767.
3. Y. Kopsinis, S. McLaughlin, Investigation and performance enhancement of the empirical mode decomposition method based on a heuristic search optimization approach, *IEEE Trans. Signal Process.* 56 (1) (2008) 1–13.
4. Justusson B.I. Median Filtering. *Digital signal processing. Math. Inst. Stockholm.* 1981. 238 p.
5. Hiroshi Ushikava. *Ultrafast All-Optical Signal Processing Devices.* National Institute of Advanced Industrial Science and Technology. Japan. 2008. 245 p.
6. John Minkoff. *Signal Processing Fundamentals and Applications for Communications and Sensing Systems.* Artech House, INS. 2002. 305 p.
7. N. Bi, Q. Sun, D. Hunag, Z. Yang, J. Huang, Robust image watermarking based on multiband wavelets and empirical mode decomposition, *IEEE Trans. Image Process.* 16 (8) (2007) 1956–1966.
8. Dwight F. Mix, Kraig J. Olejniczak. *Element sof Wavelets for Engineers and Scientists.* Wiley-Interscience. 2003. 236 p.
9. Daubechies I. The wavelet transform, time-frequency localization and signal analysis // *IEEE Trans. Inform. Theory.* – Vol. 36. – 1990. – P. 961 – 1005.
10. Daubechies I. *Ten Lectures on Wavwlwts.,* SIAM Publ., Philadelphia, 1992.

11. C. Chui. An introduction to wavelets, Academic Press, New York, 1992.
12. C. Chui (Ed.). Wavelets: A Tutorial in Theory and Applications II, Academic Press, New York, 1992.
13. Coifman R.R., Meyer Y., Wickerhauser M.V. Wavelet Analysis and Signal Processing // Wavelets and Their Applications / Ruskai et al. (ed.). – Boston Jones and Bartlett, 1992., P. 153 – 178.
14. Daubechies I. Orthonormal basis of compactly supported wavelets // Comm. Pure Appl. Math., vol. 46, 1988. P. 909 – 996.
15. Donoho D. L. De-noising by soft-thresholding. IEEE Trans. On Inf. Theory, vol. 41, 3, 1995. P. 613 – 627.
16. Donoho D. L., Johnstone. Ideal spatial adaptation by wavelet shrinkage. Biometrika, vol. 81, 1995. P. 425 – 455.
17. Kevin R., Jeffrey S., Keith A., Henry M. Improved Peak Detection and Quantification of Mass Spectrometry Data Acquired from Surface-Enhanced Laser Desorption and Ionization by Denoising Spectra with the Undecimated Discrete Wavelet Transform.
18. <http://www.mdanderson.org/education-and-research/departments-programs-and-labs/departments-and-divisions/division-of-quantitative-sciences/research/biostats-utmdabtr-001-04.pdf>
19. Samsul Ariffin Abdul Karim, Mohd Tahir Ismail. Compression Of Chemical Signal Using Wavelet Transform. European Journal of Scientific Research, ISSN 1450-216X Vol.36 No.4 (2009), pp.513-520.
20. Marina Vannucci, Naijun Sha, Philip J. Brown. NIR and mass spectra classification: Bayesian methods for wavelet-based feature selection. Chemometrics and Intelligent Laboratory Systems 77, 2005, P. 139– 148
21. Coifman R.R., Wickerhauser M.V. Entropy-based algorithms for best basis selection. IEEE Trans. On Inf. Theoriz, vol. 38, №2, 1992. P. 713-718.
22. Mallat S. A wavelet tour of signal processing. Academic Press, 1998. 452 p.