

**В.Г. ИВАНОВ**, канд. техн. наук, **М.Г. ЛЮБАРСКИЙ**, д-р физ.-мат. наук,  
**Ю.В. ЛОМОНОСОВ**, канд. техн. наук (г. Харьков)

## ПРИМЕНЕНИЕ ВЕЙВЛЕТ-АНАЛИЗА К СЖАТИЮ ЗВУКОВЫХ СИГНАЛОВ

Розглянуті теорія та практика вейвлет-перетворень при компресії звукових сигналів на основі формування вейвлет-пакетів та оригінальних квадратурних дзеркальних фільтрів. Наведені значення показників якості методу, який використовується. Представлена порівнювальна оцінка ефективності стиску з відомим форматом представлення звукових даних MP3.

The theory and practice Wavelet-transformations are considered at a compression of sound signals on the basis of formation Wavelet-packages and original quadrature mirror filters. Values of parameters of quality of a used method are given. The comparative estimation of efficiency of compression with a known format of representation of sound given (data) MP3 is given.

Одним из наиболее популярных на сегодняшний день направлений в мире персональных компьютеров является мультимедиа. С помощью приложений мультимедиа текстовая, графическая, аудио и видеoinформация объединяются, и пользователь может управлять ходом представления данной информации. При этом объем информации подлежащий обработке или передаче весьма велик, поэтому эффективное использование мультимедийных данных на персональных ЭВМ, возможно, только после их обработки специальными программами кодирования (сжатия).

За годы работы над проблемой компрессии аудиоинформации, а именно о ней пойдет речь в статье, накопилось множество продуктов, методов, алгоритмов, которые пересекаются между собой, зачастую имеют несколько разных названий и функций [1, 2]. Следует также отметить повышенный интерес к исследованию возможности сжатия данных с применением вейвлет-преобразований [5, 6]. Однако использование вейвлет-преобразований при обработке звуковых сигналов в настоящее время исследовано недостаточно. Поэтому представляет практический интерес рассмотреть возможность использования вейвлет-преобразований при обработке речевых и аудио сигналов.

Целью данной работы является анализ использования вейвлет-базисов для решения задачи компактного представления (сжатия) речевых и аудио сигналов.

1. **Вейвлет-базис.** Рассмотрим вещественный сигнал  $x(t)$ , определенный на числовой оси  $t \in R$  с конечной энергией  $\int_{-\infty}^{+\infty} x^2(t) dt$ .

Пространство всех таких сигналов принято обозначать символом  $L^2(R)$ . Это гильбертово пространство со скалярным произведением

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(t)g(t)dt, f, g \in L(R^2)$$

Идея *вейвлет-анализа* или, иначе, *многомасштабного анализа* заключается в разложении сигнала  $x(t)$  по очень специальному ортонормированному базису в  $L^2(R)$  [5, 6]. Этот базис порождается всего одной функцией  $\psi(t)$  и состоит из ее масштабных преобразований по «времени»  $t$ , кратных числу 2, и целочисленных сдвигов:

$$2^{-j/2} \psi(2^j t - i), i, j \in Z$$

(множитель  $2^{-j/2}$  нужен для нормировки).

Функция  $\psi(t)$  называется *порождающим вейвлетом*, а построенный по ней базис – *вейвлет-базисом*. Основной особенностью вейвлета является то, что он хорошо локализован по времени, то есть быстро убывает на бесконечности или имеет ограниченный носитель. Это свойство и объясняет название: вейвлет – «маленькая волна». Для практических целей используются вейвлеты с ограниченным носителем.

Если разложить сигнал  $x(t)$  по вейвлет базису:

$$x(t) = \sum_{i,j} c_{ij} 2^{-j/2} \psi(2^j t - i)$$

то индекс  $j$  указывает на временной масштаб каждого слагаемого, а индекс  $i$  – на его положение на временной оси. Слагаемые этой суммы удобно сгруппировать по масштабу:

$$x(t) = \sum_j y_j(t) \tag{1}$$

$$y_j(t) = \sum_i c_{ij} 2^{-j/2} \psi(2^j t - i) \tag{2}$$

Тогда сигнал  $x(t)$  будет представлен в виде суммы сигналов  $y_j(t)$ , каждый из которых имеет свой «характерный» временной масштаб.

Желательно, чтобы вейвлет обладал как можно большим количеством нулевых моментов:

$$M_n = \int_{-\infty}^{+\infty} t^n \psi(t) dt = 0, \quad n = 0, 1, \dots \quad (3)$$

и был гладкой функцией. В этом случае он лучше аппроксимирует гладкие сигналы, и значимых слагаемых в сумме (1) становится меньше.

Применение многомасштабного анализа в целях сжатия сигнала полностью аналогично использованию в тех же целях рядов Фурье для периодических функций. Нужно заменить только понятие частоты понятиями масштаба и временной локализации. Можно ожидать, что некоторые масштабы исходного сигнала  $x(t)$  на определенных промежутках времени будут отсутствовать или будут малы. Тогда их можно либо полностью отбросить, либо записать более короткими словами. Кроме того, можно ожидать, что масштабные составляющие  $y_j(t)$  окажутся более однородными по времени, чем исходный сигнал. В этом случае к ним с большим успехом, чем к исходному сигналу, можно применять стандартные алгоритмы сжатия.

Отметим сразу, что приложение описанной схемы к практике в вычислительном смысле не удобно из-за того, что исходный сигнал, как правило, задан дискретным образом. Поэтому вейвлеты, хотя и играют важную теоретическую роль, в вычислительных алгоритмах в явном виде не используются.

Ниже нам понадобятся следующие понятия.

Пусть  $\{x_i\}$  — дискретный сигнал, и  $\{h_s\}$  — некоторая последовательность. Операцию *свертки*, преобразующую сигнал  $\{x_i\}$  в сигнал  $\{y_i\}$  по формуле

$$y_i = \sum_s h_s x_{i-s},$$

называют применением *фильтра*  $\{h_s\}$  к сигналу  $\{x_i\}$ .

Дискретным *преобразованием Фурье* сигнала называется функция

$$x(\omega) = \sum_s x_s e^{-i\omega k}.$$

В терминах преобразований Фурье применение фильтра записывается так:

$$y(\omega) = h(\omega) x(\omega),$$

что собственно и объясняет название «частотный фильтр».

Сопряженным к фильтру  $h$  называется фильтр  $h^*$ , такой что  $h_s^* = h_{-s}$ .

**1.1 Многомасштабный анализ и алгоритм Малла.** Будем задавать временной масштаб с помощью системы подпространств  $\{V_j, j \in Z\}$  пространства  $L^2(R)$  таких, что выполнены следующие условия:

1.  $V_j \subset V_{j+1}$ ,
2.  $v(t) \in V_j \Leftrightarrow v(2t) \in V_{j+1}$ ,
3.  $\bigcup V_j = L^2(R), \bigcap V_j = \emptyset$ ,
4. существует функция  $\varphi(t) \in V_0$ , такая, что система ее целочисленных сдвигов образует ортонормированный базис в подпространстве  $V_0$ .

Свойство 2 означает, что подпространства  $V_j$  отличаются друг от друга только временным масштабом входящих в них функций и вложены друг в друга так, что минимальный временной масштаб функций из  $V_j$  вдвое больше, чем из  $V_{j+1}$ . Минимальный масштаб функций из подпространства  $V_0$  по свойству 4 задается функцией  $\varphi(t)$ . Поэтому такая система подпространств называется *многомасштабным анализом (МА)*, а функция  $\varphi(t)$  – *масштабной (или скейлинг-) функцией*. Совокупность ее целочисленных сдвигов называется *скейлинг-базисом*.

Унифицированное устройство подпространств МА позволяет рассмотреть только два из них, например,  $V_0$  и  $V_{-1}$ . Любая другая подряд идущая пара подпространств отличается только масштабом.

Заметим, что благодаря свойствам 2 и 4 система функций  $\{2^{-1/2} \varphi(t/2 - i), i \in Z\}$  является скейлинг-базисом подпространства  $V_{-1}$ , а порождающая этот базис функция  $2^{-1/2} \varphi(t/2)$  может быть выражена через скейлинг-базис пространства  $V_0$ :

$$\varphi(t/2) = \sqrt{2} \sum_i h_i \varphi(t - i). \quad (4)$$

Последовательность коэффициентов разложения  $h = \{h_i\}$ , которую в дальнейшем будем трактовать как фильтр, играет основную роль.

Обозначим через  $W_{-1}$  ортогональное дополнение к подпространству  $V_{-1}$  в  $V_0$ . Замечательным фактом является то, что в подпространстве  $W_{-1}$  существует ортонормированный базис  $\{2^{-1/2}\psi(t/2-i), i \in Z\}$ , порождаемый функцией

$$\psi(t/2) = \sqrt{2} \sum_i g_i \varphi(t-i), \quad (5)$$

где фильтр  $g = \{g_i\}$  связан с фильтром  $h$  соотношением:

$$g_i = (-1)^i h_{1-i}. \quad (6)$$

Это и есть вейвлет, связанный со скейлинг функцией  $\varphi(t)$ .

Пусть  $x_0(t)$  – заданный сигнал. Из первого свойства 3 МА следует, что сигнал при некоторой заданной точности можно считать принадлежащим одному из подпространств МА. Обозначим это подпространство через  $V_0$ , то есть примем за единицу минимальный масштаб, содержащийся в этом сигнале.

Ортогональная проекция  $x_{-1}(t)$  сигнала на подпространство  $V_{-1}$  представляет собой ту его часть, которая имеет минимальный масштаб 2. Оставшаяся часть, то есть ортогональная проекция на подпространство  $W_{-1}$ , в точности имеет масштаб 1. Она совпадает с функцией  $y_{-1}(t)$  из совокупности (2). Обе эти проекции легко найти. Пусть  $\{a_i\}$  – известные коэффициенты разложения сигнала  $x_0(t)$  по скейлинг-базису пространства  $V_0$ , а  $\{b_i\}$  и  $\{c_i\}$  – коэффициенты разложения проекций  $x_{-1}(t)$  и  $y_{-1}(t)$  соответственно в скейлинг-базисе пространства  $V_{-1}$  и в базисе подпространства  $W_{-1}$ , порождаемом вейвлетом  $2^{-1/2}\psi(t/2)$ . Тогда, используя связи (4) и (5) между базисами, легко найти:

$$b_i = \sum_s h_s^* a_{2i-s} \quad \text{и} \quad c_i = \sum_s g_s^* a_{2i-s} \quad (7)$$

Другими словами, коэффициенты разложения проекций получаются применением сопряженных фильтров  $h^*$  и  $g^*$  с последующим «прореживанием» – отбрасыванием членов с нечетными индексами.

Аналогично, если известны проекции  $x_{-1}(t) \in V_{-1}$  и  $y_{-1}(t) \in W_{-1}$  исходного сигнала, то его можно восстановить применением фильтров  $h$  и  $g$ :

$$a_i = \sum_s h_s b_{i-s}^1 + \sum_s g_s c_{i-s}^1, \quad (8)$$

к последовательностям  $\{b_s^1\}$  и  $\{c_s^1\}$ , полученным из коэффициентов разложения вставкой нулей на нечетные места. Например,  $b_{2i}^1 = b_i$  и  $b_{2i+1}^1 = 0$ .

К проекции  $x_{-1}(t)$  можно снова применить ту же процедуру, что и к исходному сигналу, то есть разложить на две составляющие из подпространств  $V_{-2}$  и  $W_{-2}$ , и таким образом выделить из сигнала компоненту  $y_{-2}(t)$  масштаба 2. После бесконечного повторения этой процедуры в силу второго свойства 3 МА, гарантирующего, что  $x_{-j}(t)$  исчезает при  $j \rightarrow \infty$ , получается разложение (1).

Указанный алгоритм называют *быстрым вейвлет-преобразованием* или *алгоритмом Малла*, а фильтры  $h$  и  $g$ , связанные соотношением (6), – *квадратурными зеркальными фильтрами*.

Конечно, для реального сигнала  $x_0(t)$ , заданного с определенной точностью и ограниченного по времени, этот процесс является конечным. Но и в этом случае его не обязательно доводить до конца. Можно ограничиться «глубиной погружения»  $N$ , то есть вычислить  $y_{-1}(t), y_{-2}(t), \dots, y_{-N}(t)$  и  $x_{-N}(t)$ . После этого, произведя необходимые действия по сжатию получившихся компонент, запомнить или передать их. Для восстановления исходного сигнала нужно рекуррентным образом, начиная с  $j = N$ , по функциям  $x_{-j}(t)$  и  $y_{-j}(t)$  восстановить  $x_{-j+1}(t)$  с помощью формулы (8). И так до того момента, когда будет найден исходный сигнал  $x_0(t)$ .

Отметим, что благодаря прореживанию, встроенному в формулу (7), коэффициентов  $\{b_i\}$  и  $\{c_i\}$  в сумме столько же, сколько и коэффициентов  $\{a_i\}$ . Поэтому представление сигнала  $x_0(t)$  в виде его масштабных

компонент, сколько бы их ни было, содержит в сумме столько же числовых значений, сколько и исходное представление.

Существенно, что в рабочих формулах явно не участвуют ни скейлинг-функция  $\varphi(t)$ , ни вейвлет  $\psi(t)$ . Только на начальном этапе требуется разложить сигнал  $x_0(t)$  по скейлинг-базису. Однако, если сигнал задан в дискретном виде  $x = \{x_i\}$ , то можно обойтись и без скейлинг-функции. Просто считать, что заданный дискретный сигнал и есть разложение некоторого сигнала  $x_0(t)$  в скейлинг базисе.

Однако не надо думать, что формулы (7) и (8) справедливы при любом выборе фильтра  $h$ . Этот фильтр имеет важное характеристическое свойство, которое можно получить, применив первую из формул (7) к сигналу  $2^{-1/2}\varphi(t/2)$ , заданному равенством (5):

$$\sum_s h_s^* h_{2i-s} = \delta_i,$$

где  $\delta_0 = 1$  единственный отличный от нуля член последовательности  $\{\delta_i\}$ .

Оказывается, что если фильтр  $h$  удовлетворяет этому свойству и некоторым другим менее существенным требованиям, то по нему можно восстановить всю схему МА. Например, скейлинг-функция  $\varphi(t)$  находится как решение функционального уравнения (4). Поэтому при построении МА исходят именно из этого условия, которое в более удобном виде записывается в терминах преобразований Фурье:

$$|h(\omega)|^2 + |h(\omega + \pi)|^2 = 2. \quad (9)$$

Семейство решений этого уравнения было найдено Ингрид Добеши. При длине фильтра  $2n$  они дают вейвлет с  $n$  нулевыми моментами (3). Класс таких решений довольно узок, и они неудобны с практической точки зрения. Например, они недостаточно гладкие и за исключением вейвлета Хаара ( $n = 1$ ) не могут быть симметричными.

**1.2 Биортогональный многомасштабный анализ и вейвлет-пакеты.** Можно указать простую модернизацию конструкции МА, для которой класс допустимых фильтров заметно шире, они могут быть симметричными, достаточно гладкими и более просто вычисляются. Это – биортогональный МА (БМА). Для него сохраняется алгоритм Малла, но в формулах (7) разложения сигнала на составляющие и восстановления сигнала

используются разные пары фильтров. Пара  $\{\tilde{h}, \tilde{g}\}$  при разложении и пара  $\{h, g\}$  при восстановлении. (Иначе говоря, в формулах (7) над  $h$  и  $g$  нужно поставить тильду, а формулу (8) оставить без изменения.) Фильтры в каждой паре связаны соотношением вида (6). Каждой паре соответствует своя скейлинг-функция и вейвлет. Но порожденные ими базисы не являются ортогональными и не ортогональны друг другу. Каждый из них биортогонален соответствующему базису другой пары фильтров.

Характеристическое свойство (9), позволяющее найти фильтры  $h$  и  $\tilde{h}$  в БМА имеет вид

$$\overline{h(\omega)\tilde{h}(\omega)} + \overline{h(\omega+\pi)\tilde{h}(\omega\pi)} = 2.$$

Для этого уравнения имеется семейство решений, зависящее от двух параметров:

$$h_n(\omega) = \sqrt{2} \left( \frac{1+e^{i\omega}}{2} \right)^n, \quad \tilde{h}_{n,m}(\omega) = \sqrt{2} \left( \frac{1+e^{i\omega}}{2} \right)^n P_m \left( \sin^2 \frac{\omega}{2} \right) e^{-im\omega},$$

где  $P_m(x) = \sum_{s=0}^{m-1} C_{m-1+s}^s x^s$  — многочлен Добеши.

Соответствующее этим решениям семейство фильтров симметрично. Разлагающий вейвлет имеет достаточно нулевых моментов, растущих с номером  $n$ , а гладкость восстанавливающего вейвлета увеличивается с ростом параметра  $m$ . При этом длина фильтров соответственно увеличивается. Пример фильтров ( $n=5, m=10$ ), использованных для конкретных вычислений, показан на рис. 1 ( $h, \tilde{h}$ ).

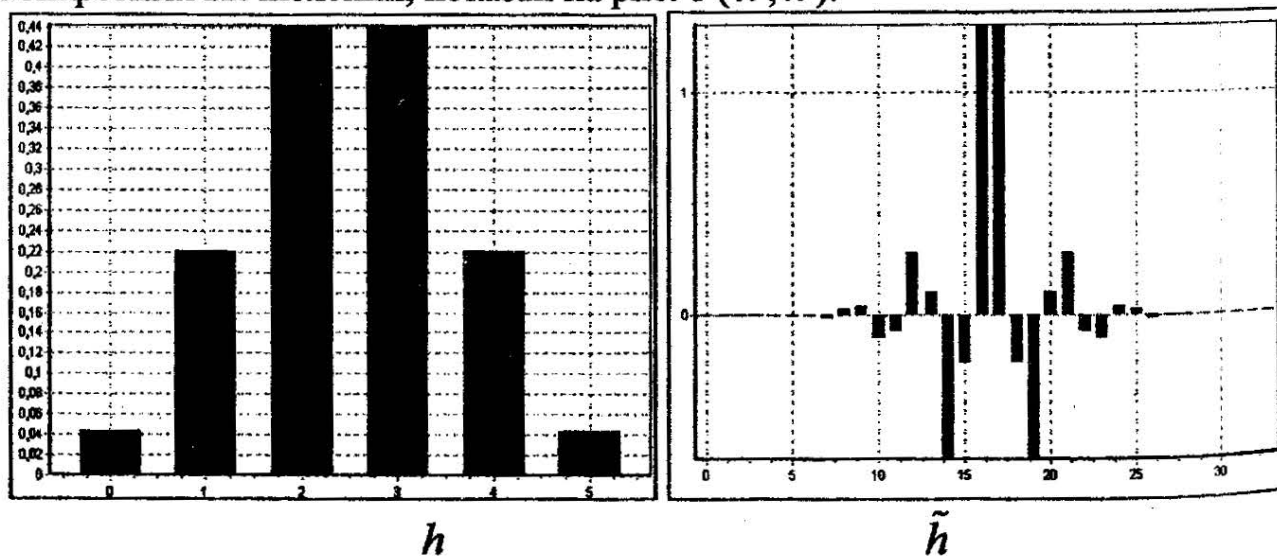


Рис. 1 Фильтры  $h$  и  $\tilde{h}$ .



Эффективность сжатия сигнала увеличивается, если перейти к так называемой схеме *вейвлет-пакетов*. Она отличается от алгоритма Малла следующим. На первом шаге алгоритма Малла (в МА или БМА безразлично) сигнал  $x_0(t)$  разлагается на две составляющие  $x_{-1}(t)$  и  $y_{-1}(t)$ . По алгоритму Малла вторая составляющая  $y_{-1}(t)$  запоминается, а к первой, то есть к  $x_{-1}(t)$ , снова применяется процедура разложения на две составляющие. По алгоритму вейвлет-пакетов процедура разложения

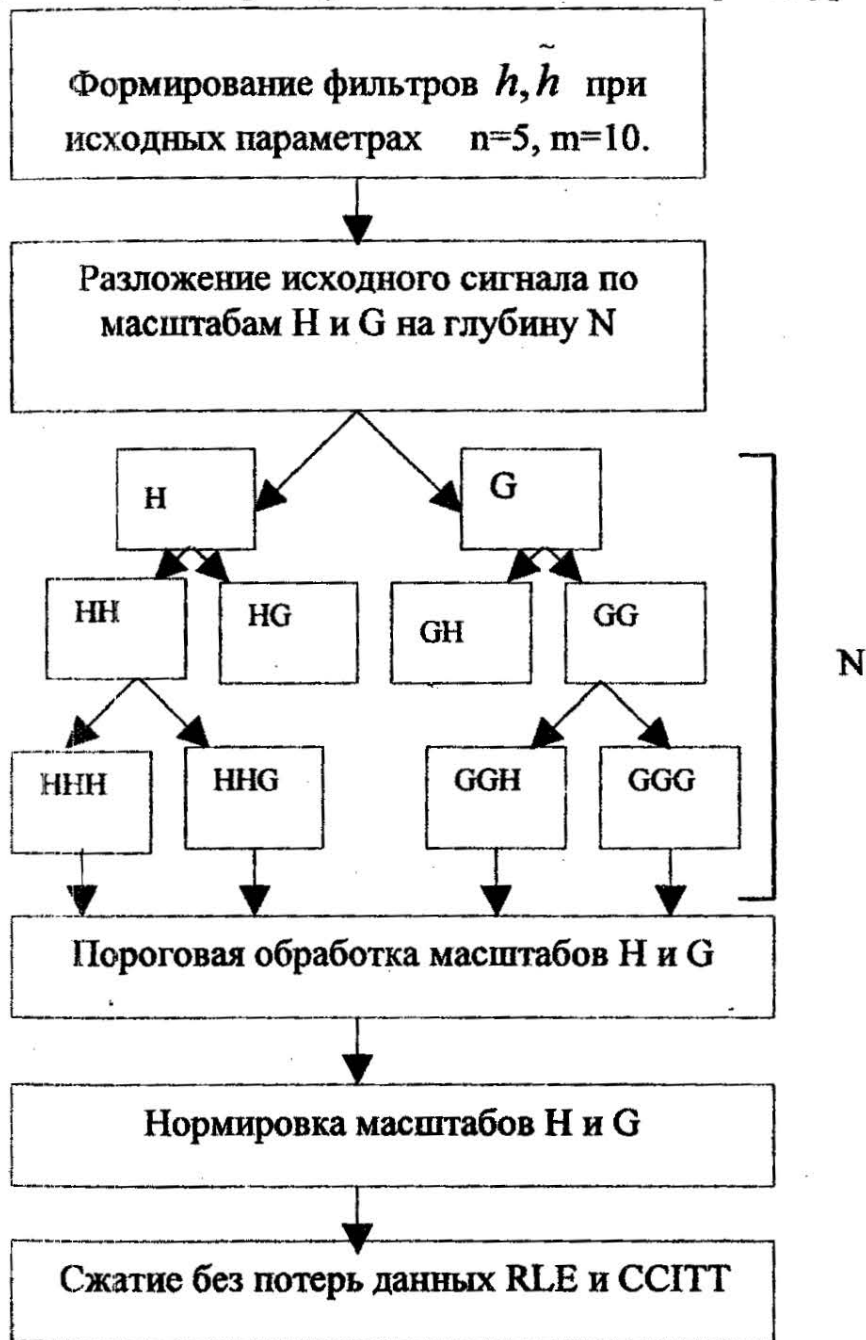


Рис. 2 Схема обработки звуковых сигналов

применяется и ко второй составляющей  $y_{-1}(t)$  и таким образом строится бинарное дерево. Общая схема обработки звуковых сигналов представлена на рис. 2.

**2. Практические результаты.** По предложенной схеме обработки и компрессии звуковых сигналов (рис. 2) формируется древовидная структура разложения, которая позволяет представить исходный звуковой сигнал по соответствующим масштабам  $H_i$  и  $G_j$ . Необходимо отметить, что за счет выбора оптимального дерева для данного класса сигнала можно повысить степень сжатия. Выбор квазиоптимального дерева основан на оценке “информативности” набора коэффициентов каждого масштаба, который определяет уровень “погружения”. Пороговая обработка коэффициентов, полученных после вейвлет-анализа исходного сигнала, подразумевает их ограничение по амплитуде. Соответственно, чем выше пороговый уровень, тем большее количество коэффициентов отбрасывается. Собственно, именно величина порогового уровня наибольшим образом влияет на степень сжатия звукового сигнала, поэтому выбор оптимального значения порогового уровня должен обеспечивать необходимую степень сжатия при допустимых искажениях звукового сигнала после его восстановления (декомпрессии). Пределы среднеквадратического отклонения (СКО), при котором звуковые сигналы не претерпевают изменений, отражающихся на качестве прослушивания после восстановления, не должны превышать 5-6%. Причем для каждой частоты дискретизации ( $F_d$ ) этот предел различен (для  $F_d = 8\text{кГц}$  СКО ( $E$ )  $\cong 5\%$ , для  $F_d = 44\text{кГц}$ ,  $E \leq 1\%$ ) [3]. Большинство алгоритмов сжатия аудиоинформации часто сопровождаются потерями данных. Такие схемы сжатия предусматривают отбрасывание данных с учетом восприятия психоакустической системы человеческого мозга. Этот принцип реализован и успешно применяется в формате MPEG (Motion Picture Experts Group) - Layer-3 (или просто MP3). Степень сжатия звуковой информации, представленной в формате MP3 значительно выше чем в других форматах. Поэтому при определении коэффициента сжатия ( $K$ ) с использованием вейвлет-преобразований целесообразно проводить их сравнительную оценку.

В качестве исходных звуковых сигналов использовались файлы (WAV – формата), содержащие речевые сообщения, а также музыкальные мелодии различного темпа звучания.

При формировании звуковых файлов выбирались следующие параметры записи:

- размерность звуковых файлов (\*.wav) формата 0,5 – 3,5 МБ;
- частота дискретизации сигнала 8, 11, 22, 44 кГц;
- динамический диапазон амплитуд звукового сигнала 0-255 или 8 бит;

Сравнительный анализ проводился по среднему значению коэффициента сжатия, полученного на основе обработки различных звуковых фрагментов.

На рисунке 3 представлено семейство кривых, характеризующих зависимость коэффициента сжатия речевого сигнала от величины уровня порогового ограничения, для частот дискретизации 8, 11, 22 и 44 кГц. Количество масштабов определяется числом уровней погружения  $N$  (рис. 2). Степень сжатия речевых сигналов представленных в формате MP3 для выбранных частот дискретизации соответственно будет равна: 8 кГц = 3,7; 11кГц = 4,2; 22кГц = 5,4; 44кГц = 8,6. Таким образом, при уровне порогового ограничения коэффициентов, равного 20, степень сжатия речевого сигнала (К) после вейвлет обработки превосходит показатели формата MP3 (11 и 8.6 раза).

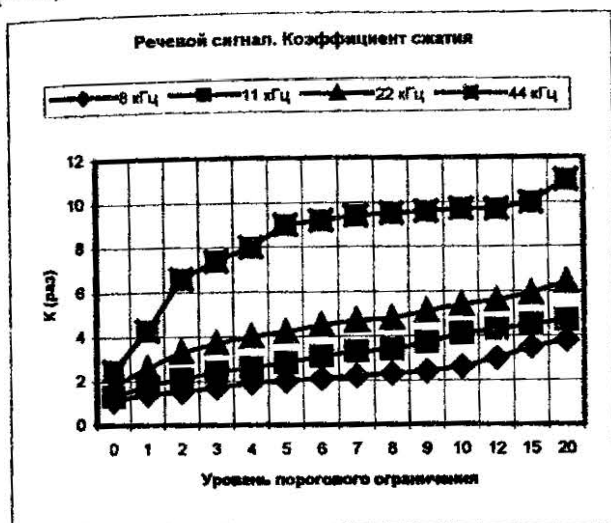


Рис.3

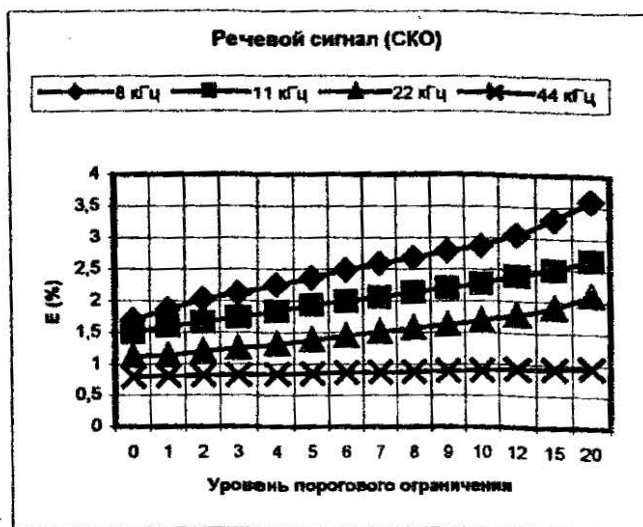


Рис. 4



Рис. 5

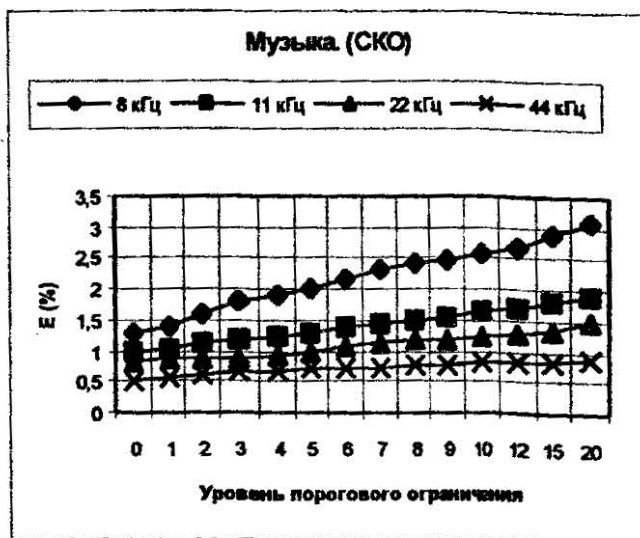


Рис.6

Среднеквадратическое отклонение (СКО) исходного и восстановленного речевых сигналов, при использовании пороговой обработки масштабов, полученных в результате вейвлет-анализа не превышает 3,5% (для частоты дискретизации 8 кГц), что обеспечивает высокое качество речи при воспроизведении. Семейство кривых характеризующих зависимость СКО от уровня порогового ограничения представлено на рис. 4.

На рисунках 5 и 6 показаны аналогичные зависимости для сигналов содержащих музыкальные композиции различного темпа звучания и частоты дискретизации. Используемые музыкальные композиции, в формате MP3, имеют степень сжатия для соответствующих частот дискретизации: 8 кГц = 3,9; 11кГц = 4,5; 22кГц = 5,8; 44кГц = 9,2. Эти значения соизмеримы с соответствующими показателями, полученными с помощью вейвлет-анализа и пороговой обработкой масштабов. При этом, все значения СКО (E) находятся в диапазоне 0,5 – 3 % для всех частот дискретизации, рис. 6.

В перспективе для получения более высоких степеней сжатия звуковых сигналов целесообразно использовать алгоритм сжатия данных на основе вычисления точек перегиба в структуре сигнала [4]. Это позволило бы уменьшить объем тех масштабов, которые содержат низкочастотные коэффициенты представляющую более “грубую” структуру сигнала. Дальнейшее повышение эффективности сжатия, после вейвлет-анализа, может быть достигнуто за счет использования алгоритмов сжатия без потерь. Вместо алгоритма группового кодирования (RLE) и модифицированного алгоритма Хаффмена (ССИТТ) можно применить алгоритм сжатия Лемпеля-Зива-Уэлча (LZW), который так же повысит (как минимум в 2 раза) степень сжатия звуковой информации после вейвлет-анализа.

Список литературы: 1. *Б. Вакулок*. Как и зачем сжимают звук // Компьютерное обозрение. – К.: Изд-во ООО “ИТС”, 1998 г., № 34, с. 20-26. 2. *Тим Кинтцель*. Руководство программиста по работе со звуком: Пер. с англ. – М.: ДМК Пресс, 2000. – 432 с., ил. 3. *Иванов В.Г., Ломоносов Ю.В.* Многоэтапный алгоритм сжатия мультимедийных данных // Научно-технический журнал ХГТУРЭ. Радиоэлектроника и информатика 2000 г. - № 4 (13). - с. 87-89. 4. *Иванов В.Г., Ломоносов Ю.В.* Алгоритм сжатия данных на основе вычислений точек перегиба в структуре сигнала // Вестник ХГПИУ. Системный анализ, управление и информационные технологии 2000 г. - № 94. - с. 25-29. 5. *Воробьев В.И., Грибунин В.Г.* Теория и практика вейвлет-преобразования. СПб.:Изд-во ВУС, 1999.-208 с. 6. I. Daubechies. Ten Lectures on Wavelets. SIAM, 1992.

*Поступила в редколлегию 19.04.03*