

УДК 004.627

В.А. Строганов, В.Н. Бондарев, доцент, канд. техн. наук

Севастопольский национальный технический университет

Студгородок, г. Севастополь, Украина, 99053

E-mail: vstroganov@mail.ru

ПЕРЦЕПТИВНОЕ СЖАТИЕ АУДИОСИГНАЛОВ С ИСПОЛЬЗОВАНИЕМ ВЕЙВЛЕТ-АНАЛИЗА

Представлены результаты разработки и экспериментального исследования характеристик метода сжатия аудиосигналов на основе вейвлет-анализа и психоакустической модели слухового восприятия.

Методы сжатия аудиосигналов основываются на учёте психоакустических особенностей восприятия звука, поскольку конечным приёмником аудиосигналов является, как правило, слуховая система человека.

В настоящее время широко используется ряд методов сжатия аудиосигналов [1], однако достигаемая ими степень сжатия и качество декодированного сигнала часто являются недостаточными, поэтому представляет интерес разработка новых, более эффективных методов.

Целью работы является разработка и исследование характеристик метода сжатия аудиосигналов на основе вейвлет-анализа и психоакустического моделирования.

В качестве исходного рассматривается сигнал с частотой дискретизации 44,1 кГц. Для кодирования каждого отсчёта используется 16 бит. Схема кодера, реализующего предложенный метод сжатия аудиосигналов, изображена на рисунке 1.

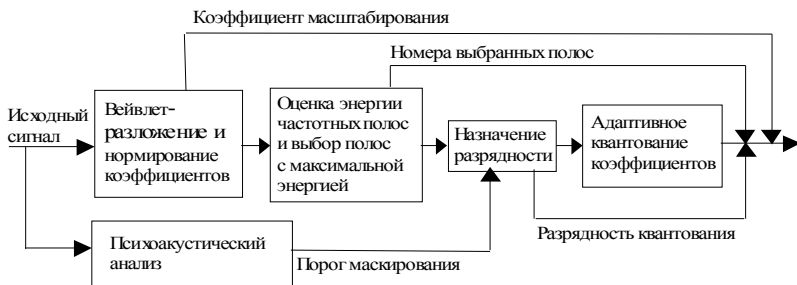


Рисунок 1 – Схема кодера аудиосигналов

Входной сигнал разбивается на блоки фиксированной длины. Выполняется разложение каждого блока по вейвлет-пакетам. При этом используются вейвлет-функции Добеши различных порядков.

делить реальный уровень звукового давления, создаваемого сигналом, для его сравнения с абсолютным порогом слышимости. Поэтому абсолютный порог слышимости нормируется так, чтобы его минимальное значение равнялось оценке спектральной плотности мощности сигнала с амплитудой, представляемой 1 битом (т.е. минимального сигнала, представимого с помощью данной разрядности квантования по уровню).

После вычисления абсолютного порога слышимости определяют пороги маскирования, возникающие вследствие наличия маскирующих составляющих в спектре сигнала. Предполагается, что в каждой критической полосе присутствует одна тональная маскирующая составляющая. Влияние шумовых маскирующих составляющих не учитывается. Тональная маскирующая составляющая определяется как максимальное значение СПМ сигнала в пределах одной критической полосы. Для анализируемого блока сигнала определяют 25 маскирующих составляющих, по одной в каждой критической полосе. Индивидуальный порог маскирования, создаваемый на частоте f_1 тональной маскирующей составляющей с частотой f_{TM} , определяется выражением

$$T(f_1, f_{TM}) = P_{TM}(f_{TM}) - 0,275z(f_{TM}) + SF(f_1, f_{TM}) - 6,025,$$

где $z(f_{TM})$ – частота в барках, соответствующая частоте f_{TM} Гц.

Функция $SF(f_1, f_{TM})$, описывающая распространение маскирующего эффекта, аппроксимируется выражением

$$SF(f_1, f_{TM}) = \begin{cases} 17\Delta_z - 0,4P_{TM}(f_{TM}) + 11, & -3 \leq \Delta_z < -1, \\ (0,4P_{TM}(f_{TM}) + 6)\Delta_z, & -1 \leq \Delta_z < 0, \\ -17\Delta_z, & 0 \leq \Delta_z < 1, \\ (0,15P_{TM}(f_{TM}) - 17)\Delta_z - 0,15P_{TM}(f_{TM}), & 1 \leq \Delta_z < 9, \end{cases}$$

где $\Delta_z = z(f_1) - z(f_{TM})$ – расстояние в барках между маскирующей и маскируемой составляющими. При вычислении порогов маскирования считается, что действие эффекта маскирования распространяется на 3,5 барк до маскирующей составляющей и на 8,5 барк после неё. Суммарный порог на частоте f определяется выражением

$$T_{sum}(f) = \max\{ATH(f), \max_k(T_k(f))\},$$

где $T_k(f)$ – индивидуальный порог маскирования, создаваемый на частоте f k -й маскирующей составляющей.

На рисунке 3 показан график СПМ аудиосигнала и суммарный порог маскирования, полученный в результате психоакустического анализа.

Спектральные составляющие сигнала, лежащие ниже суммарного порога слышимости, не воспринимаются на слух. Поэтому можно понизить разрядность квантования вейвлет-коэффициентов таким образом, чтобы уровень шума квантования остался ниже суммарного порога. Для удобства вычислений суммарный порог слышимости $T_{sum}(f)$ и СПМ сигнала $P(f)$ нормируются следующим образом:

$$T_N(f) = \frac{T_{sum}(f) + \left| \min_k P(k) \right|}{\max \left(T_{sum}(f) + \left| \min_k P(k) \right| \right)} \cdot 2^{16},$$

$$P_N(f) = \frac{P(f) + \left| \min_k P(k) \right|}{\max \left(T_{sum}(f) + \left| \min_k P(k) \right| \right)} \cdot 2^{16}.$$

Тогда разрядность квантования n_i для вейвлет-коэффициентов, соответствующих i -й критической полосе с граничными частотами f_1 и f_2 , определяется выражением

$$n_i = \log_2 \frac{\max_{f_1 \leq f \leq f_2} P_N(f)}{\min_{f_1 \leq f \leq f_2} T_N(f)}.$$

На приёмную сторону передаются номера используемых частотных полос, вейвлет-коэффициенты, соответствующие этим полосам, разрядность квантования коэффициентов и нормировочный коэффициент.

В декодере, схема которого показана на рисунке 4, восстанавливается разрядность вейвлет-коэффициентов, выполняется их домножение на нормировочный коэффициент и вычисляется обратное вейвлет-преобразование.



Рисунок 4 – Схема декодера аудиосигналов



Рисунок 3 – СПМ аудиосигнала и суммарный порог маскирования

Для экспериментального определения эффективности разработанного метода сжатия использовался набор из четырёх тестовых сигналов, представляющих собой короткие фрагменты музыкальных произведений различных жанров: 1) «organ» – запись классического органа; 2) «piano» – фортепианное соло; 3) «rock band» – запись рок-группы; 4) «orchestra» – запись симфонического оркестра.

Для сравнения качества исходного и восстановленного сигналов применялся так называемый «слепой тест»: эксперту предлагается прослушать два фрагмента, один из которых подвергался сжатию. Считается, что качество восстановленного сигнала субъективно не отличается от качества исходного, если эксперт не в состоянии правильно идентифицировать исходный сигнал.

В ходе тестирования ставилась цель определить максимальную степень сжатия для каждого из тестовых сигналов без потери субъективного качества. Кроме того, анализировалась эффективность использования вейвлет-функций различных порядков в зависимости от спектральной и временной структуры сигнала.

Таблица 1 – Результаты тестирования

Тестовый сигнал	Степень сжатия					Макс. сжатие	MPEG-Layer3
	Db2	Db4	Db6	Db8	Db10		
«organ»	–	2,93	4,74	5,67	5,52	5,67	14,69
«piano»	–	2,05	4,66	5,71	5,56	5,71	7,34
«rock band»	2,92	2,87	2,83	2,78	–	2,92	2,75
«orchestra»	2	1,98	1,96	1,94	–	2	2,2

Результаты экспериментов приведены в таблице 1. Сокращение DbN , где N – число, использовано для обозначения вейвлет-функций Добеши порядка N . Для сравнения приведены коэффициенты сжатия, полученные для тех же тестовых сигналов с помощью метода сжатия аудиосигналов MPEG-Layer3. Следует отметить, что в методе MPEG-Layer3 используется более сложная психоакустическая модель.

Проведённые эксперименты показывают, что предложенный в данной работе метод позволяет получить степень сжатия, аналогичную результатам применения существующих методов, для сигналов, имеющих сложную частотно-временную структуру, т.е. для таких сигналов, при обработке которых проявляются достоинства вейвлет-анализа. Совершенствование метода возможно за счёт применения вейвлет-функций и схем разложения, адаптируемых к анализируемому звуковому сигналу, а также за счёт совершенствования применяемой психоакустической модели, что и будет представлять перспективы дальнейших исследований в данном направлении.

Библиографический список

1. Painter T. Perceptual coding of digital audio / T. Painter, A. Spanias // Proceedings of the IEEE. — 2000. — V. 88. — № 4. — P. 451–513.

Поступила в редакцию 24.10.2003 г.