

В.Г. ИВАНОВ, канд. техн. наук,
Ю.В. ЛОМОНОСОВ, канд. техн. наук (г. Харьков)

ОБОБЩЕНИЕ JPEG-ФОРМАТА ДЛЯ КОДИРОВАНИЯ МУЛЬТИМЕДИЙНЫХ ДАННЫХ

Запропоновано та досліджено методику узагальнення JPEG-формату для кодування мультимедійних даних на основі розширеного сімейства ортогональних перетворень, що дозволяє домогтися універсальності і зберегти прийнятну якість стиску. Приводяться порівняльні результати комп'ютерного моделювання мовних і звукових сигналів з використанням JPEG-формату і MP3.

The technique of generalization of a JPEG-format for coding the multimedia data is offered and investigated on the basis of the expanded family of the orthogonal transformations, allowing to achieve universality and to keep acceptable quality of compression. Comparative results of computer modelling of speech and sound signals with use of a JPEG-format and MP3 are resulted.

Постановка проблемы. Профессиональная и социально-культурная среда современного человека становится электронной, а парадигмой этого процесса является огромный объем цифровых данных различной физической природы, который создается, хранится и циркулирует в этой среде. Сегодня в мире происходит формирование новой отрасли, называемой пакетной телефонией, которая основана на передаче голоса, данных и видео в сетях с пакетной коммутацией [1 – 3]. Так одна минута цифровой аудиозаписи с качеством звучания лазерного диска занимает порядка 10,8 Мб информации и для прослушивания в реальном времени такого файла нужно иметь скорость передачи до 1,5 Мбит/сек. Трудно надеяться, что отечественные телефонные каналы обеспечат трафик более чем 16 – 32 Кбит/сек без применения специальных средств сжатия данных.

Анализ исследований по этой проблеме показывает, что самым популярным на сегодняшний день аудиоформатом является MPEG-1, который использует три метода сжатия, называемые слоями (layer), а наиболее эффективным является его третий слой, который широко известен по аббревиатуре MP3 [4, 5].

Высокая степень компактности MP3 достигается с помощью дополнительного квантования по установленной схеме, позволяющей минимизировать потери качества на основе учета особенностей человеческого слуха, в том числе эффекта маскирования слабого сигнала одного диапазона частот более мощным сигналом соседнего диапазона, или мощным сигналом предыдущего фрейма, вызывающего временное понижение чувствительности уха к сигналу текущего фрейма.

Подобная техника называется адаптивным кодированием и позволяет экономить на наименее значимых с точки зрения восприятия человеком

деталей звучания. Структурная схема, реализующая данный формат, представлена на рис. 1. Однако этот формат не предназначен для работы с изображениями, что снижает его функциональные возможности и универсальность.

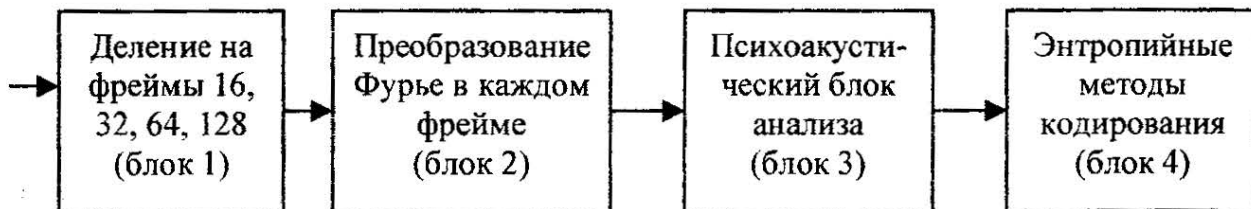


Рис.1. Схема формирования формата MP3

Целью работы является проведение исследований, направленных на обобщение и расширение возможностей известного JPEG-формата компрессии изображений для обработки аудиоданных, что позволит наметить пути создания универсальных форматов данных различной физической природы. Исследовать влияние основных ортогональных преобразований: Фурье, дискретного косинусного преобразования и преобразования Хаара, на качество и степень сжатия аудиосигналов различной структуры и интенсивности.

Компьютерное моделирование JPEG-формата на аудиоданных. В работах [6, 7] авторы предложили и исследовали оригинальную методику сжатия звуковых данных, но она совершенно отлична от JPEG-технологий и потому не является универсальной. Известно также, что MP3 является потоковым форматом [4, 5]. Это означает, что передача данных происходит потоком независимых отдельных блоков данных – фреймов. Для этого исходный сигнал при кодировании разбивается на равные по продолжительности участки, именуемые фреймами и кодируемые отдельно. При декодировании сигнал формируется из последовательности декодированных фреймов.

Предлагаемый алгоритм формирования аудиосигнала в сжатом виде аналогичен тем преобразованиям, которые используются в формате MP3 за исключением основного блока – блока психоакустической обработки (блок 3, рис. 1). Психоакустическая обработка в формате MP3 осуществляется после того, как исходный сигнал поделен на фреймы. В MP3 каждый фрейм имеет даже собственный заголовок, но в нашем случае это не обязательно, так как длины их одинаковы. После разделения сигнала на фреймы каждый фрейм подвергается выбранному преобразованию (преобразование Фурье, дискретное косинусное преобразование – DCT, преобразование Хаара). Конечным этапом обработки является преобразование аудиоданных без потерь. В качестве энтропийных методов кодирования использовались известные алгоритмы группового кодирования (RLE) и алгоритм Хаффмена.

Блок 3, реализующий психоакустическую обработку, в данном случае отсутствует.

Параметры аудиосигналов:

- частота дискретизации – 8, 11, 22, 44 КГц;
- динамический диапазон – 8 бит (0–255);
- количество каналов – моно.

Длина фреймов выбиралась равной 16, 32, 64, 128 отсчетов исходного звукового сигнала. Далее каждый фрейм отдельно подвергался соответствующему преобразованию. После чего производилось перераспределение полученных коэффициентов со всех участков обработки. Отдельно группировались все первые коэффициенты из всех фреймов, за ними следовали в заданной последовательности все вторые и т.д. Такое структурирование коэффициентов позволяет получить наиболее равномерное распределение амплитуд всей совокупности обрабатываемых коэффициентов. Таким образом формируется последовательность значений перед энтропийным кодированием.

Необходимо отметить, что подобная схема обработки используется и в алгоритме JPEG при обработке изображений (рис. 2) [8]. Единственным отличием является то, что изображение разбивается на сегменты (8×8) и в каждом сегменте отдельно производится дискретное косинусное преобразование на плоскости. Однако после этого происходит считывание полученных коэффициентов по методу Z-сканирования, и каждый сегмент так же представлен вектором полученных коэффициентов. Дальнейшее преобразование происходит по одинаковой схеме – сортировка коэффициентов и энтропийное кодирование.

Таким образом, можно говорить об универсализации методов обработки мультимедийных данных различной физической природы (аудиоданные и изображения).

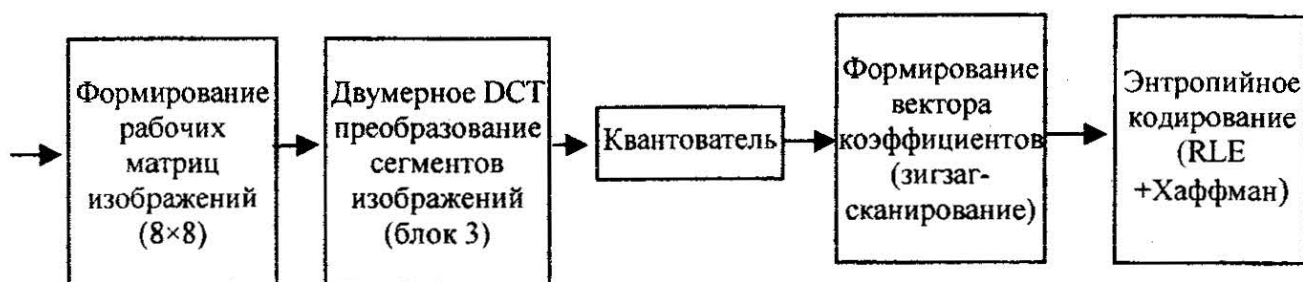


Рис. 2. Схема кодирования изображений в терминах JPEG-технологий

На рис. 3, 4 представлены графики зависимости коэффициента сжатия ($K_{сж}$) от частоты дискретизации звукового сигнала, длины фреймов и типа преобразования. На рис. 3 представлена обработка слов “пять, четыре, три, два, один”, на рис. 4 – музыкальная композиция “Мажор”.

Для каждой частоты дискретизации отдельно указаны значения для алгоритма MP3, в правой части каждого графика на рис. 3, 4.

Блок 3, реализующий психоакустическую обработку, в данном случае отсутствует.

Параметры аудиосигналов:

- частота дискретизации – 8, 11, 22, 44 КГц;
- динамический диапазон – 8 бит (0–255);
- количество каналов – моно.

Длина фреймов выбиралась равной 16, 32, 64, 128 отсчетов исходного звукового сигнала. Далее каждый фрейм отдельно подвергался соответствующему преобразованию. После чего производилось перераспределение полученных коэффициентов со всех участков обработки. Отдельно группировались все первые коэффициенты из всех фреймов, за ними следовали в заданной последовательности все вторые и т.д. Такое структурирование коэффициентов позволяет получить наиболее равномерное распределение амплитуд всей совокупности обрабатываемых коэффициентов. Таким образом формируется последовательность значений перед энтропийным кодированием.

Необходимо отметить, что подобная схема обработки используется и в алгоритме JPEG при обработке изображений (рис. 2) [8]. Единственным отличием является то, что изображение разбивается на сегменты (8×8) и в каждом сегменте отдельно производится дискретное косинусное преобразование на плоскости. Однако после этого происходит считывание полученных коэффициентов по методу Z-сканирования, и каждый сегмент так же представлен вектором полученных коэффициентов. Дальнейшее преобразование происходит по одинаковой схеме – сортировка коэффициентов и энтропийное кодирование.

Таким образом, можно говорить об универсализации методов обработки мультимедийных данных различной физической природы (аудиоданные и изображения).



Рис. 2. Схема кодирования изображений в терминах JPEG-технологий

На рис. 3, 4 представлены графики зависимости коэффициента сжатия ($K_{сж}$) от частоты дискретизации звукового сигнала, длины фреймов и типа преобразования. На рис. 3 представлена обработка слов “пять, четыре, три, два, один”, на рис. 4 – музыкальная композиция “Мажор”.

Для каждой частоты дискретизации отдельно указаны значения для алгоритма MP3, в правой части каждого графика на рис. 3, 4.

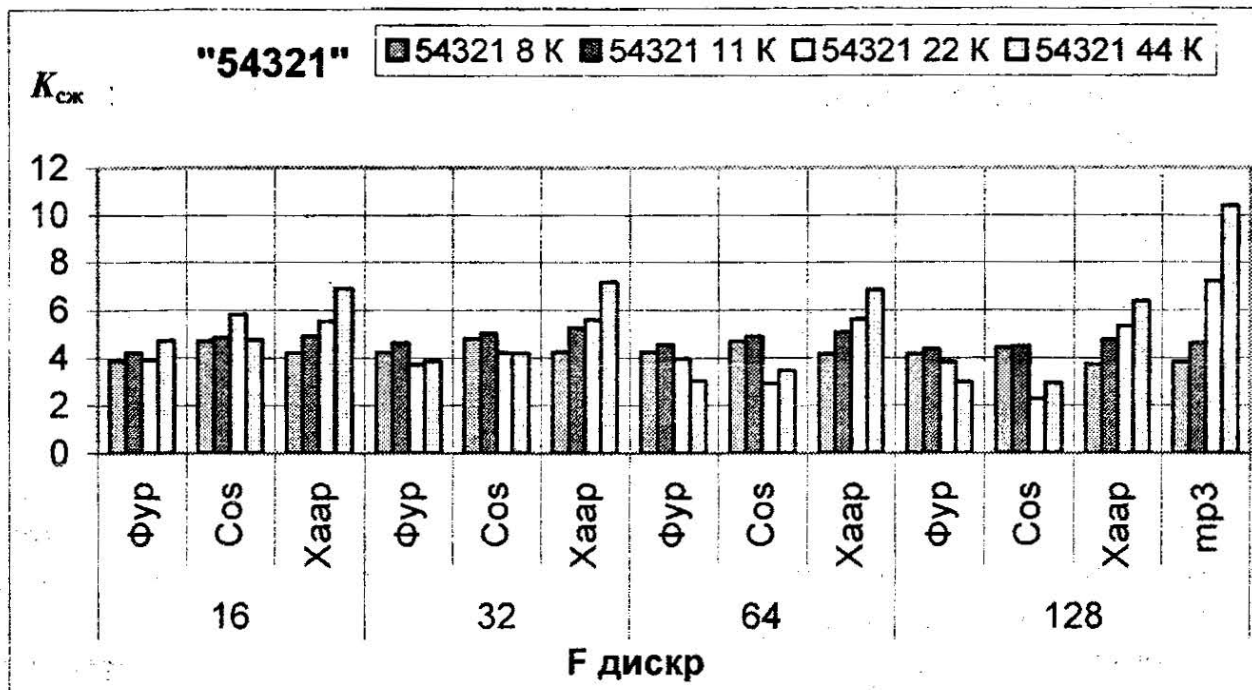


Рис. 3. Зависимость $K_{сж}$ от частоты дискретизации, длины фрейма и анализируемого преобразования последовательности слов "пять, четыре, три, два, один"

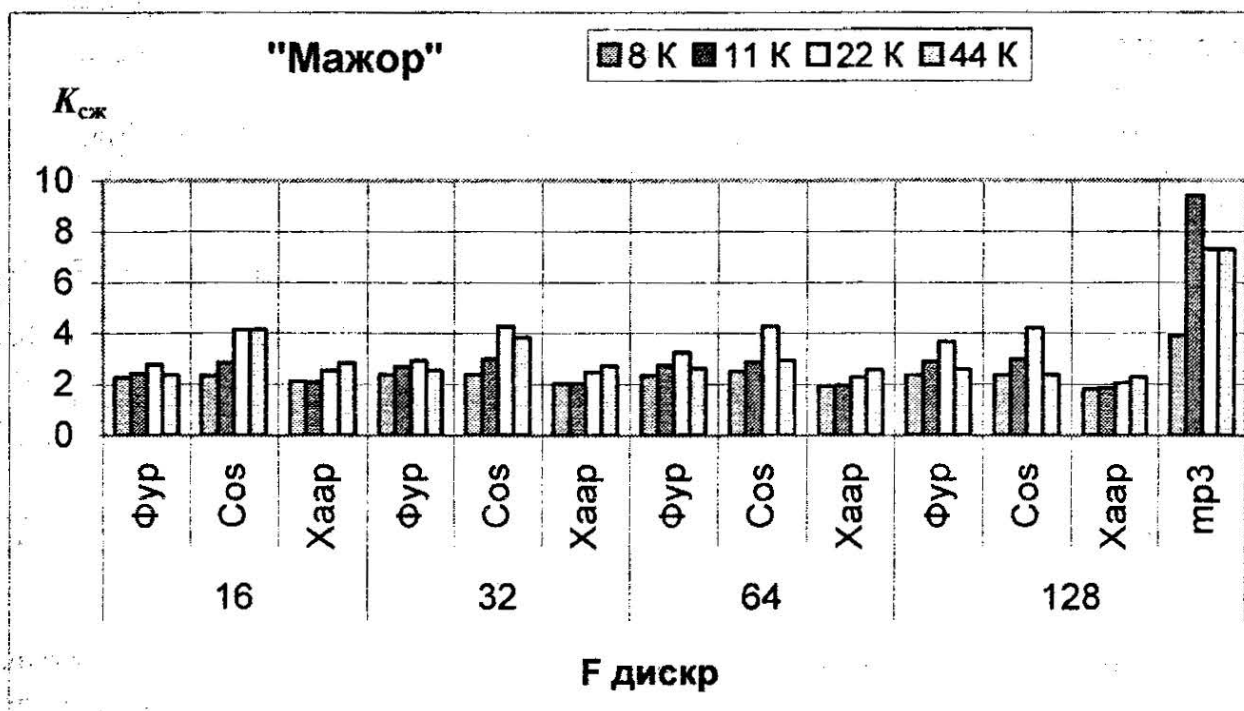


Рис. 4. Зависимость $K_{сж}$ от частоты дискретизации, длины фрейма и анализируемого преобразования музыкальной композиции "Мажор"

Выводы. Из анализа приведенных графиков можно сказать следующее:

– на речевых сообщениях, произносимых одним диктором при высоких частотах первичной дискретизации, наблюдаются преимущества ортогонального базиса Хаара по сравнению с другими базисами при любой длине фрейма обрабатываемого сигнала;

– для речевого сообщения (рис. 3), на низкой частоте дискретизации в 8 КГц все линейные преобразования не уступают по эффективности сжатия формату МРЗ при любой длине фрейма;

– для звукового сигнала это свойство утрачивается и лидером становится дискретное косинусное преобразование при любых исходных параметрах эксперимента;

– при высоких частотах первичной дискретизации формат МРЗ имеет явные преимущества, как на речи, так и на звуке;

– при изменении длины фрейма речевого сигнала в сторону увеличения, эффективность сжатия в базисе Фурье падает, что можно объяснить большим проявлением свойств нестационарности речевых данных.

Список литературы: 1. Головных А. Цифровая среда обитания // СНІР. Компьютеры и коммуникации. – К.: Издат. дом "Софт Пресс". – 2003. – № 1. – С. 68 – 70. 2. Букашкин С.А., Лисицын Г.Ф., Миронов В.Г. Применение цифровых процессоров обработки сигналов – магистральный путь развития современных средств телекоммуникаций / Тр. 3-й Международной конференции "Цифровая обработка сигналов и ее применение". – Часть 3. – М.: Институт проблем управления Российской академии наук. – 2000. – С. 3 – 4. 3. Винцюк Т. Мовленнєві інформаційні технології в Україні – на шляху до європейського співробітництва Схід-Захід / Праці Сьомої Всеукраїнської міжнародної конференції "Оброблення сигналів і зображень та розпізнавання образів" (УкрОБРАЗ' 2004). – К.: Кібернетичний центр Національної академії наук України, 2004. – С. 9 – 17. 4. Китцель Тим. Руководство программиста по работе со звуком. – М.: ДМК Пресс, 2000. – 432 с. 5. Сэломон Д. Сжатие данных, изображений и звука. – М.: Техносфера, 2004. – 368 с. 6. Иванов В.Г., Ломоносов Ю.В. Многоэтапный алгоритм сжатия мультимедийных данных // Научно-технический журнал ХГТУРЭ "Радиоэлектроника и информатика". – 2000. – № 4 (13). – С. 87 – 89. 7. Иванов В.Г., Ломоносов Ю.В. Алгоритм сжатия данных на основе вычислений точек перегиба в структуре сигнала // Вестник ХГПУ. Тематический выпуск "Системный анализ, управление и информационные технологии". – 2000. – № 94. – С. 25 – 29. 8. Иванов В.Г., Ломоносов Ю.В., Любарский М.Г. Фурье и вейвлет анализ изображений в плоскости JPEG-технологий // Проблемы управления и информатики. – 2004. – № 5. – С. 111 – 124.

Поступила в редакцию 07.04.2006