

## О НОВОМ МЕТОДЕ СЖАТИЯ МУЗЫКАЛЬНЫХ ФАЙЛОВ

**И.И. Чижов**

*Белгородский государственный  
университет*

*e-mail: [chizhov@bsu.edu.ru](mailto:chizhov@bsu.edu.ru)*

Основным препятствием на пути перехода на новые более совершенные чем MP3 форматы сжатия аудиоданных с потерями, является невозможность конвертации в них накопленной фонотеки MP3-файлов. В статье предложен метод решения данной проблемы.

Ключевые слова: сжатие информации, звуковые файлы, компрессия, степень сжатия, форматы сжатия аудиоданных.

---

В настоящее время существует 2 пути для уменьшения объема музыкального файла: сжатие без потерь (по другой классификации – обратимое искажение) и сжатие с потерями (по другой классификации – необратимое искажение). Стоит отметить, что в обоих случаях речь идет о передаче битового потока, т.е. создаваемый файл имеет бинарную структуру. Кратко охарактеризуем оба подхода:

- Сжатие без потерь – полная идентичность исходного трека аудио CD и конечного файла. Для уменьшения объема используются классические методы сжатия данных, такие как метод Хаффмана или метод арифметического кодирования. Стоит отметить, что степень сжатия при использовании данного подхода крайне мала, и сильно зависит от содержимого файла, что не позволяет передавать по каналам Интернет файлы, сжатые без потерь. Примерный объем 1 минуты записи – 30-40 Мб.

- Сжатие с потерями – удаление из исходного музыкального сигнала некоторых составляющих, которые с точки зрения физиологии, не воспринимаются человеческим ухом (психоакустический принцип) и сужение кодируемой полосы частот. На сегодняшний день, большинство из известных форматов сжатия звука (MP3, WMA, OGG и т.д.) базируются на данном подходе. При сжатии с потерями уменьшение объема достигается за счет последовательного применения к исходному сигналу ряда процедур: переход из временной области в частотную; применение психоакустического принципа; квантование по уровню; сжатие данных классическими методами (как правило – методом Хаффмана). В зависимости от необходимого пользователю качества звучания, объем минуты записи может составлять от 2 до 20 МБ. Качество звучания (или другими словами – ошибка в представлении данных) определяется параметром, называемым битрейт. Кратко напомним, что битрейтом называют количество бит необходимое для кодирования секунды записи мультимедиа потока. Единицей измерения данной величины является "килобит в секунду", т.е. кбит/с. Типичными битрейтами являются 32 кбит/с и 64 кбит/с - для мобильных устройств, а так же 128 кбит/с, 256 кбит/с, 320 кбит/с. – для стационарных устройств. Не сложно заключить, что с повышением битрейта качество воспроизведения аудиозаписи возрастает.

Как уже отмечалось выше, сжатие с потерями использует большинство известных на сегодняшний день форматов сжатия музыкальных файлов с потерями. В настоящее время, наибольшее распространение получили следующие форматы:

MP3 – это сокращение от MPEG-1 Layer III. Формат разработан Институтом Фраунгофера (Fraunhofer IIS) и фирмой Thomson. На сегодняшний день он является самым распространенным стандартом сжатия аудио с потерями. Благодаря высокой компактности сжатых файлов при достаточно хорошем качестве этот формат вполне подходит для обмена аудиоинформацией через Интернет и создания большой фонотеки на персональном компьютере. На сегодняшний день морально устарел, на смену ему приходит формат MP3Pro. Кодек MP3Pro анонсирован в июле 2001 года компанией Coding Technologies вместе с Tomson Multimedia и институтом Fraunhofer. Формат MP3Pro является развитием MP3. В MP3Pro использована новая технология – SBR (Spectral Band Replication). Эта технология предназначена для передачи верхнего час-



тотного диапазона. Идея данной технологии следующая: кодируется более узкий диапазон частот чем обычно, а верхние частоты восстанавливаются декодером на основе информации о более низких частотных составляющих. Таким образом, технология SBR применяется фактически не столько на стадии сжатия, сколько на стадии декодирования. Качество звучания MP3Pro можно назвать субъективно очень хорошим даже на битрейте 64 Кбит/с, другими словами, субъективно несложные композиции при таком битрейте воспринимаются не хуже чем MP3 128 Кбит/с. Однако необходимо учитывать тот факт, что подобное звучание достигается искусственным путем, и что слышимый сигнал представляет собой уже не столько оригинал, сколько копию низких и средних частот оригинала с искусственно добавленными некими верхними частотами, зачастую совершенно не совпадающими с исходными. Кроме того, стоит отметить, что файл, сжатый по алгоритму MP3Pro занимает наибольший объем из рассматриваемых.

Алгоритм (кодек) Windows Media Audio (WMA) фирмы Microsoft. Алгоритм WMA, также как и все рассмотренные ранее, позволяет потоковое воспроизведение (stream playback). Качество WMA (если говорить о WMA 7.0 и 8.0) при скорости потока 64 Кбит/с практически не уступает качеству MPEG-1 Layer III 96-128 Кбит/с, а при 96 Кбит/с может превосходить MPEG-2 AAC 128 Кбит/с. Стоит отметить, что речь идет о субъективно воспринимаемом качестве звучания рядом экспертов. Для хранения потока в формате WMA используется универсальный потоковый файловый формат .ASF (Advanced Audio Streaming), пришедший на замену .WAV. Вообще говоря, .ASF – это универсальный формат для хранения аудио и видео информации, сжатой с помощью различных кодеков. .ASF имеет также и свою несколько сокращенную разновидность .WMA. Файлы .WMA предназначены исключительно для хранения аудио данных. Говоря о WMA как о кодеке, следует сказать, что в последнее время он становится все более популярным, так как Microsoft встроила его в Windows'2000 и Windows XP, превратив его таким образом в стандарт. Стоит отметить, что данный кодек бесплатен, однако он некорректно поддерживается некоторыми популярными программными плеерами, а так же аппаратными устройствами (такими как автомагнитолы). По сравнению с первым алгоритмом, созданный с его помощью файл занимает меньший объем.

Ogg Vorbis, разработан в июне 2000. Этот формат является частью проекта Ogg Squish по созданию полностью открытой системы мультимедиа. Иными словами, и сам проект, и Ogg Vorbis в частности являются открытыми и свободными для распространения, а также разработки на его основе нового программного обеспечения. В FAQ от разработчика (группа Xiphophorus) написано, что Ogg Vorbis является алгоритмом, принципиально схожим с известными MPEG-1 Layer II, MPEG-2 AAC, VQF и проч., однако использует оригинальный математический алгоритм и собственную психоакустическую модель, что освобождает его от необходимости выплачивать лицензионные сборы и производить иные выплаты сторонним фирмам-изготовителям аудио форматов. Алгоритм Ogg Vorbis рассчитан на сжатие данных на всех возможных битрейтах без ограничений, то есть от 8 Kbps до 512 Kbps, а также на кодирование с переменным битрейтом (VBR). Алгоритм предусматривает хранение внутри файлов подробных комментариев об исполнителе и названии композиции, а также графической информации. В алгоритме предусматривается также возможность кодирования нескольких каналов аудио (более двух, теоретически до 255), возможность редактирования содержимого файлов, а также так называемый "масштабируемый битрейт" – возможность изменения битрейта потока без необходимости декодирования. Поддерживается потоковое воспроизведение (streaming). Для хранения данных используется собственный универсальный формат bitstream Ogg Squish, рассчитанный на хранение любой информации мультимедиа системы Ogg Squish (идея реализации универсального формата bitstream Ogg Squish аналогична идее ASF от Microsoft). Объем файла созданного по данному алгоритму минимален из известных на сегодняшний день, однако стоит сказать о слабой распространенности данного формата в связи с отсутствием возможности преобразования уже имеющихся записей в формате MP3 в него (точнее, подобная возможность есть, однако это серьезно скажется на качестве звука) и отсутствием поддержки со стороны популярных плееров.



Рассматривая достоинства и недостатки каждого из представленных форматов нельзя не сказать о том, что все они базируются на сходных идеях, различаясь лишь в конкретных алгоритмах и в психоакустической модели.

Стоит отметить, что в настоящее время у пользователей накоплена весьма значительная фонотека MP3 файлов, можно даже говорить о переводе большинства музыкальных композиций в этот формат, а многих из них – многократно. Конвертация MP3 в любой сходный формат, не минуемо ведет к потере качества звучания, вследствие использования различных психоакустических моделей. Таким образом, основной задачей в сжатии музыкальных файлов с потерями на сегодняшний день является создание алгоритма, для которого возрастание погрешности представления аудиоданных при преобразовании в него из любого исходного формата будет не выше 1-2% и степень сжатия после преобразование будет превосходить MP3, т.к. ни один известный формат не способен осуществить подобную конвертацию.

Пути достижения данной цели видятся в отказе от использования психоакустического принципа уменьшения битовых представлений исходных аудиоданных и в отказе от искусственного сужения частотной полосы музыкального сигнала. Данные подходы взаимосвязаны. Рассмотрим их более подробно.

При использовании психоакустической модели из исходного сигнала удаляются различные частотные компоненты, якобы невоспринимаемые человеческим ухом. Подобная уловка получила название частотного и временного маскирования. Однако не стоит забывать, что формат MP3 разрабатывался в начале 90-х годов прошлого века, и задача авторов состояла в том, чтобы повторить качество магнитофонной записи. Таким образом, несложно заметить, что вместо глубокого изучения слухового анализатора человека, авторы банально воспользовались несовершенством воспроизводящей аппаратуры того времени. Подтверждением тому является сужение полосы частот в зависимости от битрейта при кодировании сигналов, что нельзя объяснить ни частотным ни временным маскированием, а лишь узостью полосы пропускания колонок персонального компьютера от 0 до 16 кГц, что соответствует битрейту 128 кбит/с, который являлся на тот момент базовым. Таким образом, психоакустический подход сам по себе является несовершенным, и кроме того, влечет значительные трудности для конвертации из одного формата сжатия аудиоданных с потерями в другой.

Вследствие того, что каждая из фирм-разработчиков использует собственную психоакустическую модель, при конвертации из исходного сигнала удаляется часть частотных компонент, в надежде на то, что оставшиеся смогут «замаскировать» от слушателя их отсутствие, но т.к. при сжатии в исходный формат (как правило MP3) из сигнала уже была удалена часть (как правило совершенно иная) частотных компонент, то качество сигнала резко ухудшается, т.к. маскирование уже невозможно. Таким образом, достаточно очевидно, что подобная конвертация возможна только при условии что формат сжатия аудиоданных в который происходит преобразование не использует психоакустическую модель, т.е. искусственно не удаляет из сигнала частотные компоненты.

Искусственное сужение частотной полосы исходного сигнала в зависимости от битрейта также является препятствием для успешного преобразования файла сжатого по одному алгоритму в другой, т.к. в этом случае верхние частоты исходного сигнала безвозвратно удалены и не могут быть восстановлены. Таким образом, несмотря на то что прямым препятствием к конвертации музыкальных файлов сужение частотной полосы исходного сигнала и не является, но оно не позволяет пользователю получить в свое распоряжение оригинал музыкальной композиции, что является значительным недостатком для формата сжатия аудиоданных с потерями.

На факультете компьютерных наук и телекоммуникаций Белгородского государственного университета разработан формат сжатия аудиоданных с потерями свободный от перечисленных недостатков (свидетельство о регистрации в Отраслевом фонде алгоритмов и программ № регистрации: 4314).

Алгоритм его работы состоит в выполнении следующих последовательных шагов:



### **1. Дискретное косинус-преобразование (ДКП) всего обрабатываемого отрезка;**

Переход из временной области в частотную целесообразен, так как большое количество коэффициентов ДКП, особенно высокочастотных может быть равен нулю. Таким образом, подобный переход уменьшает энтропию исходной последовательности отсчетов музыкального сигнала.

### **2. Адаптивное квантование по уровню коэффициентов косинус-преобразования;**

В исходной звуковой последовательности каждый отсчет кодируется при помощи 16 или 24 бит. С точки зрения восприятия звука человеческим ухом, во многих случаях столь точное представление не является необходимым. Вследствие этого, при сжатии музыкальных данных прибегают к более грубому представлению аудиоданных. Степень подобной "грубости" определяется задаваемой пользователем погрешностью представления исходных данных квантованными значениями.

### **3. Сжатие квантованных значений при помощи арифметического кодера.**

Арифметический метод сжатия данных разрабатывался как развитие идеи кода Хаффмана, что позволяет данному методу показывать более высокую степень сжатия по сравнению с другими методами сжатия данных без потерь. Тот факт, что во всех известных на сегодняшний день алгоритмах сжатия музыкальных файлов используется метод Хаффмана объясняется существованием патентов, ограничивавших до недавнего времени использование арифметического кодера и значительной стандартизованностью (в первую очередь в выборе таблиц Хаффмана) современных алгоритмов сжатия музыкальных файлов с потерями. В ходе исследований статистических свойств дискретизованных значений звукового сигнала была экспериментально установлена возможность аппроксимации на основе функции распределения вероятности Коши. Данная модификация метода сжатия, позволила значительно повысить его эффективность для пакетной передачи данных.

Длина обрабатываемого блока избрана равной 512 отсчетам, исходя из следующих обстоятельств:

- для выполнения быстрого косинус-преобразования (БКП) необходимо, чтобы длина отрезка была равна целой степени двойки;
- музыкальный сигнал нестационарен, поэтому целесообразно использовать достаточно короткие интервалы.

Для проверки эффективности разработанного формата сжатия аудиоданных с потерями был проведен ряд вычислительных экспериментов. В качестве исходных сигналов, были выбраны различные музыкальные композиции (около 650 файлов), отобранные из возможных вариантов по методу котируемой выборки.

Эксперимент состоял из нескольких этапов:

1. Сжатие исходного сигнала по технологии MP3 (кодек LAME 2008 года) с указанным в таблице битрейтом;

Файлы сжимались с наиболее типичными на сегодняшний день битрейтами: 64 кбит/с, 128 кбит/с и 256 кбит/с;

2. Восстановление исходного сигнала из .mp3 файла;
3. Определение относительной среднеквадратической погрешности представления исходных данных восстановленными значениями;
4. Квантование спектральных коэффициентов исходного сигнала с относительной погрешностью равной вычисленной с последующим арифметическим сжатием и записью в файл разработанного формата;
5. Расчет степени сжатия для каждого из форматов;

Степень сжатия представляет собой отношение исходного объема файла к объему файла после сжатия.

6. Вычисление отношения объема MP3 к объему файла разработанного формата.



Результаты данных экспериментов показывают, что при сжатии всех рассмотренных файлов, на любых битрейтах степень сжатия предлагаемого алгоритма в 1,3-1,5 раза выше, чем у формата MP3 при равной погрешности представления исходных данных.

Если же говорить о возможности перекодирования в новый формат файлов сжатых по технологии MP3, то погрешность после конвертации не превышает 1%, что обусловлено применением доказанных выше путей решения данной задачи.

Резюмируя все вышеизложенное, можно говорить о том, что разработанный на факультете компьютерных наук и телекоммуникаций новый формат сжатия аудиоданных с потерями лишен недостатков присущих современным форматам сжатия подобного рода и обладает рядом существенных конкурентных преимуществ, что наглядно продемонстрировали вычислительные эксперименты.

#### Литература

1. Артюшенко, В. М. Цифровое сжатие видеoinформации и звука [Текст] : учеб. пособие / В. М. Артюшенко, О. И. Шелухин, М. Ю. Афонин ; под ред. В. М. Артюшенко. – М. : Дашков и К, 2003. – 426 с.
2. Баранов, Л. А. Квантование по уровню и временная дискретизация в цифровых системах управления [Текст] / Л. А. Баранов. – М. : Энергоатомиздат, 1990. – 304 с.
3. Витерби, Э. Д. Принципы цифровой связи и кодирования [Текст] / А. Д. Витерби, Дж. К. Омура ; пер. с англ. и под ред. К. Ш. Зигангирова. – М. : Радио и связь, 1982. – 536 с. : ил. – (Статистическая теория связи ; вып. 18).
4. Вологдин, Э. И. Слух и восприятие звука [Текст] : учеб. пособие / Э. И. Вологдин. – СПб. : СТ «Факультет ДВО», 2004. – 52 с.
5. Голд, Б. Цифровая обработка сигналов [Текст] : пер. с англ. / Б. Голд, Ч. Рейдер. – М. : Сов. радио, 1973. – 376 с.
6. Грудинин, А. С. Кодирование сигналов звукового вещания в базисе дискретного косинусного преобразования [Текст] / А. С. Грудинин, А. М. Синильников // Техника средств связи. – 1986. – Вып. 3. – С. 3-10. – (Сер. ТРПА).
7. Жилияков, Е. Г. О субполосном кодировании сигнала [Текст] / Е. Г. Жилияков, И. Г. Попов, И. И. Чижов // Вестник НТУ (ХПИ) : сб. науч. тр. – Харьков, 2004. – № 46. – С. 10-20. – (Тем. вып. «Информатика и моделирование»).
8. Жилияков, Е. Г. Оптимальный синтез квантователя по уровню [Текст] / Е. Г. Жилияков, И. Г. Попов, И. И. Чижов // Вестник НТУ (ХПИ) : сб. науч. тр. – Харьков, 2004. – № 46. – С. 101-106. – (Тем. вып. «Информатика и моделирование»).
9. Ковалгин, Ю. А. Цифровое кодирование звуковых сигналов : учеб. пособие [Текст] / Ю. А. Ковалгин, Э. И. Вологдин. – СПб. : КОРОНА-принт, 2004. – 240 с.

### ABOUT THE NEW METHOD OF SOUND COMPRESSION

**I.I. CHIZHOV**

*Belgorod state university  
e-mail: [chizhov@bsu.edu.ru](mailto:chizhov@bsu.edu.ru)*

The main obstacle on the way of the transition to the new more perfect than MP3 formats of the sound loss compression, is impossibility of convert the existing MP3-files in them. In the article the method of the given problems decision is offered.

Key words: the compression of the information, sound files, compression, compression rate, formats of the sound compression.