

УДК 004.627

СОВРЕМЕННЫЕ ТЕХНОЛОГИИ СЖАТИЯ АУДИОСИГНАЛОВ

Е.В. Пиуновский, А.А. Тропченко

Рассмотрены возможности и преимущества сжатия аудиоданных с применением вейвлет-преобразований. Выявлены проблемы существующих методов компрессии звука, проанализированы способы внедрения вейвлетов (wavelets) в алгоритмы сжатия, предложена схема сжатия аудиоданных на основе адаптивных ортогональных преобразований.

Ключевые слова: сжатие аудио, кодек, кодирование, вейвлет-преобразование, MPEG.

Введение

В последние десять лет можно наблюдать взрывной рост в использовании сети Интернет и мобильных телефонов, а сближение двух этих технологий открывает широкий диапазон новых возможностей на уже процветающем рынке мультимедиа. Эти возможности побуждают к проведению исследований, которые могут и должны выявить недостатки существующих методов обработки цифровых данных и показать пути их (методов) оптимизации для удовлетворения современных нужд рынка. В данном контексте очень большой интерес представляют вопросы сжатия звуковой информации, которые в настоящее время являются весьма актуальными, о чем говорят большое количество практических исследований и постоянная разработка и опубликование стандартов в этой области [1]. Но, несмотря на подобное внимание, распространенные алгоритмы сжатия аудиоданных все еще содержат ряд существенных недостатков [2], которые и описаны в данной статье. Очевидно, что для их устранения необходимо внедрение новых технологий в процесс компрессии звука. Одной из таких перспективных технологий являются вейвлет-преобразования (ВП).

По мнению исследователей, можно без преувеличения сказать, что вейвлеты (wavelets) произвели революцию в области теории и практики обработки нестационарных сигналов [3]. Вейвлет-функции уже достаточно долго используются в различных областях не только науки, но и техники, что связано с большими преимуществами, которые обещает их применение. Но некоторые возможности их приложения все еще вызывают вопросы о целесообразности и споры в научных сообществах. Одним из таких приложений является сжатие аудиоданных с использованием ВП. В данной статье рассматривается возможность применения ВП для обработки сигналов с целью их сжатия и выявляются достоинства такого подхода путем сравнения основных научных работ в данной области [4, 5] и анализа предлагаемых в них подходов к использованию ВП в процессе кодирования. В итоге, на основании проведенных исследований, предлагается структурная схема аудиокодека, включающего определенный вид ВП.

Сжатие аудиоданных: обобщенная схема и существующие проблемы

Сжатие аудиосигналов (АС) может быть осуществлено как без потерь (архивация), так и с потерями, причем в последнем случае кодирование можно подразделить на три вида [1–3]:

1. кодирование непосредственно реализации аудиосигнала (Wave Form Codec);
2. измерение, кодирование и передача на приемную сторону параметров АС, по которым уже на приемной стороне производится синтез этого (искусственного) АС. Такие системы называют вокодерными (Source Codec);
3. гибридные способы кодирования, т.е. сочетание первого и второго способов кодирования.

Кодеры формы волны аппроксимируют изменение сигнала во времени. Они требуют наибольших скоростей передачи, но имеют наилучшие показатели качества воспроизведенной речи.

В вокодерных системах моделируется процесс речеобразования человека. В коде-ре для аудиосигнала вычисляются определенные параметры, передаваемые к декодеру, в котором они применяются для восстановления формы исходного сигнала. Вокодерные преобразования отличаются наименьшими требованиями к полосе пропускания.

Один из способов повышения эффективности использования полосы пропускания состоит в применении гибридных методов, основанных на принципах линейного предсказания. В табл. 1 приведены данные по соответствию качества речи, MOS, задержек передачи и типов каналов, удовлетворяющих предъявленным требованиям [3].

Качество	Лучшее	Хорошее	Среднее	Плохое	Стандарт
MOS	> 4,5	4–4,5	3,5–4	3–3,5	ITU-T P.800, P.830
Задержка, мс	< 150	< 250	< 350	< 450	ETSI TS 101 329
	< 150	< 260	< 400	> 400	ITU-T G.114

Таблица 1. Сравнительные характеристики гибридных методов

На рис. 1 изображены сглаженные зависимости оценок MOS от требований к битовой скорости потока, построенные по усредненным результатам исследований ITU Study Group 15 [2, 4].

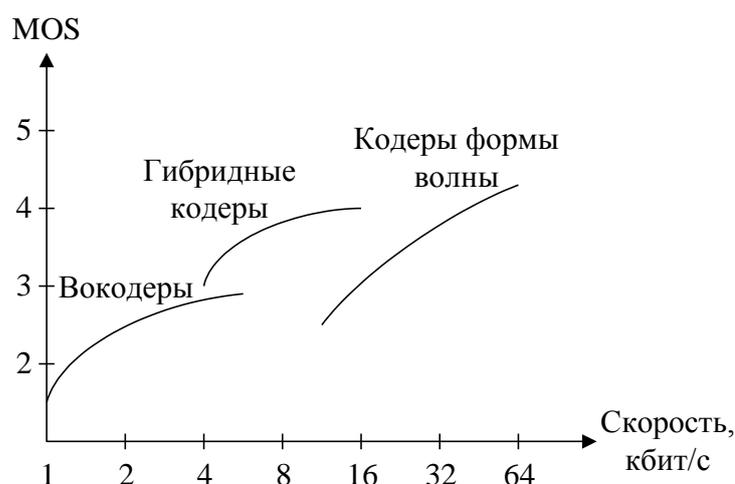


Рис. 1. Зависимость оценок MOS от скорости потока для кодеров формы волны, вокодеров и гибридных кодеров

В таблице 2 приведены сведения о характеристиках основных типов кодеков и используемых в них алгоритмах.

Несмотря на значительное разнообразие алгоритмов компрессии цифровых аудиоданных, структура кодера, реализующего такой алгоритм обработки сигналов, может быть представлена в виде обобщенной схемы, показанной на рис. 2. В блоке частотного и временного анализа исходный звуковой сигнал разделяется на субполосные (подполосные) составляющие и сегментируется по времени. Решение об изменении длины кодируемой выборки может принимать блок психоакустического анализа, в котором учитываются закономерности слухового восприятия звукового сигнала. В этом же блоке по специальной процедуре рассчитывается максимальный уровень искажений («шумов») квантования для каждого субполосного сигнала [1].

Алгоритм компрессии аудиоданных	Метод компрессии	Скорость передачи, кбит/с на канал	Величина компрессии	Области применения
ASPEC	Кодирование с преобразованием	64–192	1:6	ISDN
ATRAC	Субполосное кодирование с преобразованием	256	1:5	MiniDisk
MUSCAM	Субполосное кодирование	128–256	1:4	DAB (Digital Audio Broadcasting)
MPEG-1, Layer 1 и Layer 2	Субполосное кодирование (MUSICAM)	32–448 (Layer 1) 32–384 (Layer 2)	1:4 (Layer 1) 1:6 (Layer 2)	DAB (Layer 2, 128-256 кбит/с), DBS (Direct Broadcast Satellite, Layer 2, 224 кбит/с), DCC (Digital Compact Cassete, Layer 1, 384 кбит/с)
MPEG-1, Layer 3	Субполосное кодирование с преобразованием	32–320	1:9	Internet-вещание
MPEG-2	Субполосное кодирование/ субполосное кодирование с преобразованием	32–384	>1:9	Многоканальное стереофоническое вещание
MPEG-2 AAC	Субполосное кодирование с преобразованием	16–384	1:15	Многоканальное стереофоническое вещание
MPEG-4	Субполосное кодирование с преобразованием/ параметрическое кодирование	2-64	1:20	Мультимедиа приложения
Dolby AC-3	Кодирование с преобразованием	32–384	1:13	Кинематограф, HDTV, спутниковое вещание
Гибридное	Субполосное кодирование с преобразованием/ параметрическое кодирование	32–64	1:(15-20)	Радиовещание, хранение информации

Таблица 2. Характеристики основных типов кодеков и используемых в них алгоритмов

Фактически можно выделить три основных этапа кодирования звука: фильтрация, применение психоакустической модели, квантование и кодирование. Реализациям практически каждого из этих этапов в существующих технологиях присущи определенные недостатки [2].

Так, в подавляющем большинстве алгоритмов во время фильтрации отсчеты сигнала делятся на субполосы (в алгоритме самого популярного на сегодняшний день формата MP3 (MPEG-1, Layer 3) их 32) равной ширины, что упрощает фильтры, но сильно контрастирует с особенностями слухового восприятия, которое зависит от частоты звука [2]. Кроме того, смежные фильтры должны идеально пропускать разные диапазоны частот. На практике они имеют существенное частотное перекрытие. Звук, состоящий из одного чистого тона, может попасть в два фильтра и породить сигналы (которые потом будут квантоваться) в две подполосы вместо одной [1].

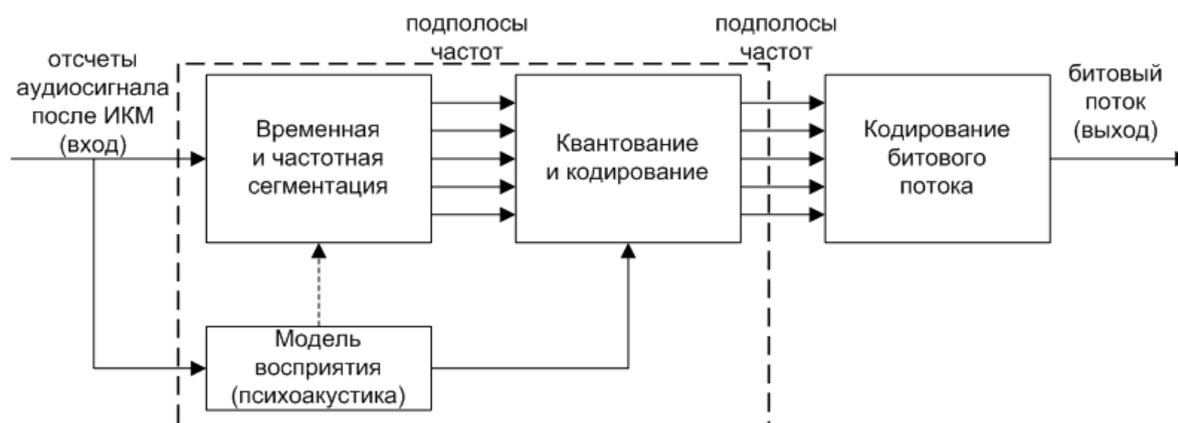


Рис. 2. Обобщенная структурная схема аудиокодера

Кроме того, традиционно используемое в алгоритмах модифицированное дискретное косинусное преобразование (МДКП) обеспечивает лучшее частотное разрешение, но одновременно, в силу принципа неопределенности, оно имеет более плохое временное разрешение. На практике это выражается в том, что квантование коэффициентов МДКП порождает ошибки, которые распределены во времени, и, кроме того, образует искажения, которые проявляются в виде предшествующего эха («пред-эха») [1].

Применяемая в настоящее время в распространенных алгоритмах психоакустическая модель, называемая NMR (Noise-to-Mask Ratio, «соотношение шум–маска») также не лишена недостатков [1]: в ней при расчете глобального (суммарного) порога маскировки учитываются лишь абсолютный порог слышимости и явление маскировки в частотной области, но игнорируются другие важные эффекты (такие, как маскировка во временной области и явление демаскировки сигналов).

Вейвлеты в кодировании звука

Перечисленные в предыдущем разделе недостатки говорят о необходимости применения новых технологий в области сжатия аудиоданных. Одной из перспективных технологий являются вейвлеты. В пользу возможности применения ВП в области кодирования звука говорят следующие факты:

- многократные заявления разработчиков кодека Vorbis, известного как Ogg Vorbis (Xiph.Org Foundation), о скором включении дискретного вейвлет-преобразования (ДВП) в процесс сжатия в данном формате;
- попытки применения вейвлетов для кодирования аудиосигналов (для разнообразных применений) фирмой Adapted Wave Technologies;
- поддержка Институтом инженеров по электротехнике и радиоэлектронике (IEEE, Institute of Electrical and Electronics Engineers) исследований по разработке адаптивных вейвлет-кодеров для звуковых данных [5];
- наличие нескольких успешных разработок в области сжатия аудиоданных с помощью ВП на низких скоростях передачи битов (bit rates, битрейтах) [4, 5].

Два последних факта представляют особый интерес, так как в них представлены реальные результаты попыток внедрения вейвлетов в процесс кодирования аудиоданных, вейвлет-технологии, используемые в каждой из этих работ, различны, и оба этих исследования впоследствии послужили основанием для ряда научных статей в области нововведений в цифровой обработке звука. Результаты анализа исследований [4, 5] представлены в табл. 3.

Используемые преобразования и технологии	Прозрачное сжатие аудио на низких битрейтах с использованием адаптированных вейвлетов [4]	Сжатие аудио высокого качества с использованием адаптивной декомпозиции вейвлет-пакетов и психоакустической модели [5]
Вейвлет-технология (замена традиционному МДКП)	ДВП	Пакетное ДВП (ПДВП)
Базисная вейвлет-функция	Не указана (предлагается адаптивный подбор функции)	Биортогональный вейвлет (5-го порядка)
Психоакустическая модель	Упрощенная	Модифицированная психоакустическая модель 2 стандарта MPEG ISO/IEC 11172–3
Дополнительные технологии	Устранение статистической избыточности с использованием динамического словаря	Алгоритм распределения бит на основе нуль-дерева
Способ проверки качества кодирования	Субъективное тестирование	Субъективное тестирование
Битрейт, на котором достигается прозрачное кодирование	48–66 кбит/с (с применением дополнительных технологий); 64–70 кбит/с	45 кбит/с

Таблица 3. Сравнительные характеристики технологий сжатия звука на базе ВП

Комментируя результаты сравнения, представленные в таблице, необходимо уточнить, что под «прозрачным» кодированием понимается такое сжатие, при котором слушатель (участник «субъективного тестирования») не может отличить сжатый звуковой сигнал от оригинала (их качество кажется ему идентичным). Здесь сразу становится очевидной сложность процесса оценки качества сжатия аудиоинформации, ее субъективность. Но процесс тестирования представлен в рассмотренных работах весьма основательно, и для сравнения следует заметить, что разработчики формата MP3 заявляли о достижимости прозрачного кодирования при использовании соответствующего алгоритма лишь на битрейте, равном 128 кбит/с (и даже это заявление подвергается частой критике).

Как видно из таблицы, рассмотренные работы схожи по общему подходу к модификации процесса сжатия звука, но различаются реализацией отдельных его шагов. Отталкиваясь от того, что понятие «вейвлет», более всего нас интересующее, выделено даже в названиях обеих статей, можно отметить следующие общие недостатки, присутствующие данным исследованиям (именно при акцентировании внимания на ВП):

- использование дополнительных технологий (что мешает точному определению вклада вейвлетов в улучшение результатов сжатия);
- недостаточно глубокий подход к выбору ВП, используемого при кодировании.

В то же время можно заметить, что при подобной расстановке приоритетов и выделении представленных выше недостатков работа [5] является все же более предпочтительной: дополнительные технологии, применяемые в ней, заимствованы (и проверены другими исследователями), в ней осуществлен переход от ДВП к более сложным и эффективным ПДВП. Скорее всего, именно с этим и связан достигнутый в ней более высокий результат (низкий битрейт).

Применение адаптивных ортогональных преобразований для сжатия аудиоданных

Основываясь на результатах сравнения, приведенных в предыдущем разделе, и учитывая возможность применения адаптивных ортогональных преобразований [3], основанных на ВП, можно предложить модифицированную схему сжатия аудиоданных с использованием данных преобразований. В такой схеме кодирования звука основное

внимание уделено оптимизации этапа фильтрации – применению банков фильтров на основе адаптивных преобразований.

Следует отметить, что банки фильтров и вейвлет-преобразования являются весьма близкими понятиями. Фактически идея ДВП возникла на основе методов субполосного кодирования и может быть представлена в виде низкочастотных (НЧ) и высокочастотных (ВЧ) фильтров, применяемых к исходному сигналу. Но для более гибкого подхода к декомпозиции сигналов можно перейти к так называемым пакетам вейвлетов, где дальнейшему воздействию фильтров подвергается не только НЧ-компонента [3]. Настоящее же преимущество для фильтрации дают алгоритмы выбора последующего пути каскадирования сигнала. Существует три таких алгоритма: одиночное дерево (ПДВП), двойное дерево, частотно-временное дерево. Все они позволяют производить адаптивную фильтрацию; разницу же между ними удобно показать на результатах разбиения частотно-временной плоскости с их помощью [3], как представлено на рис. 3.

Из рис. 3 видно, что адаптивный подход к ВП позволяет перейти от сигнально-независимых преобразований к адаптированным, позволяющим реализовывать произвольное разбиение частотно-временной плоскости [3], что дает огромное преимущество при кодировании реальных сигналов. Так, алгоритм одиночного дерева (вейвлет-пакеты) позволяет осуществить адаптацию в частотной области, алгоритм двойного дерева – сначала во временной, потом в частотной, а алгоритм частотно-временного дерева – одновременно в обеих областях [3]. Но, конечно, вместе с увеличением гибкости, возрастает и трудоемкость данных алгоритмов, что говорит о возможности использования каждого из них в зависимости от требований к процессу и результатам сжатия.

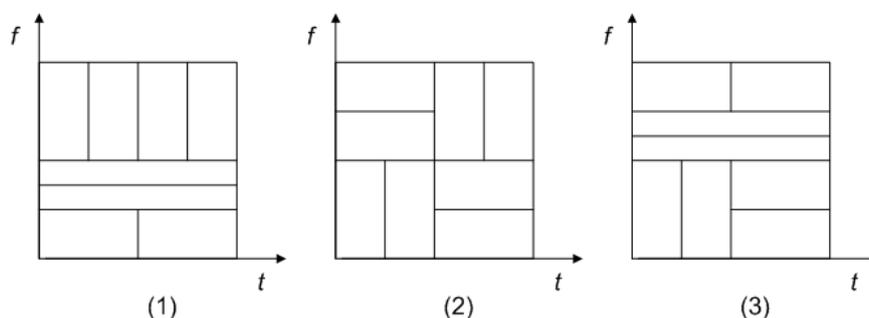


Рис. 3. Примеры разбиения, достигаемые алгоритмами:

1 – одиночного дерева; 2 – двойного дерева; 3 – частотно-временного дерева

Очевидно, что рассмотренные методы могут значительно улучшить процедуру временной и частотной сегментации звукового сигнала. Предлагаемый способ включения данных технологий в процесс кодирования аудиоданных показан на рис. 4.

К кодеру (схема которого расположена сверху) поступают данные о допустимой сложности вычислений (заранее сформированные и преобразованные к унифицированному виду), которые могут быть вычислены на основании условий, в которых выполняется сжатие, и требований, предъявляемым к его результатам. В зависимости от поступивших данных происходит выбор банка фильтров соответствующей сложности, основанного на одном из трех алгоритмов, рассмотренных выше. Кроме того, внутри кодера выходные данные блока психоакустического анализа также передаются к блоку фильтрации и служат функцией стоимости [3], на основе которой осуществляется преобразование (определяется наилучший путь по дереву). В таком случае осуществляется адаптивная фильтрация, при которой не только подбирается оптимальный базис разложения в пространственной и частотной областях, но также учитываются внешние ограничения на время и ресурсоемкость преобразования.

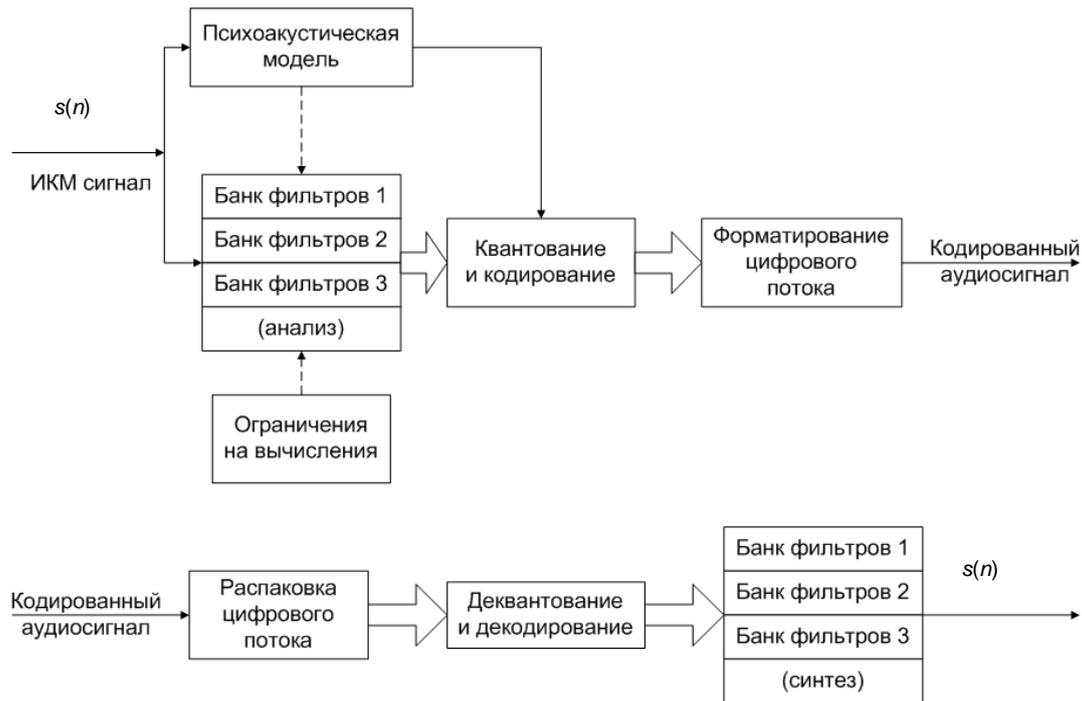


Рис. 4. Предлагаемая схема адаптивного аудиокодека

Заключение

Современные методы сжатия аудиоданных не лишены недостатков и поэтому могут быть улучшены. В статье были показаны основные проблемы распространенных алгоритмов кодирования звука. Также был проведен анализ двух успешных исследований, в которых была доказана возможность улучшения существующих технологий звуковой компрессии с помощью вейвлетов, являющихся одной из самых перспективных технологий в области цифровой обработки сигналов. В результате была предложена схема аудиокодека на основе адаптивных ортогональных преобразований, резко повышающих гибкость процесса кодирования.

Литература

1. Ковалгин Ю.А, Вологдин Э.И. Цифровое кодирование звуковых сигналов. – СПб: КОРОНА-принт, 2004. – 240 с.
2. Сэломон Д. Сжатие данных, изображений и звука. – М.: Техносфера, 2004. – 368 с.
3. Воробьев В.И., Грибунин В.Г. Теория и практика вейвлет-преобразования. – СПб: Изд-во ВУС, 1999. – 208 с.
4. Sinha D., Tewfik A. Low Bit Rate Transparent Audio Compression Using Adapted Wavelets // IEEE Trans. ASSP. – December 1993. – V. 41. – № 12. –
5. Srinivasan P., Jamieson L. High Quality Audio Compression Using an Adaptive Wavelet Packet Decomposition and Psychoacoustic Modeling // IEEE Transactions on Signal Processing. – April 1998. – V. 46. – № 4.

Пиуновский Евгений Витальевич – Санкт-Петербургский государственный университет информационных технологий, механики и оптики, аспирант, evgpiun@gmail.com

Тропченко Андрей Александрович – Санкт-Петербургский государственный университет информационных технологий, механики и оптики, кандидат технических наук, доцент, zaika_98rus@mail.ru