

УДК 519.62



О СПОСОБЕ КОДИРОВАНИЯ СИГНАЛОВ РАЗЛИЧНОЙ ПРИРОДЫ

С.Н. Герасин¹, О.В. Калиниченко², В.П. Лоцман³, А.С. Трошило⁴

¹ ХНУРЭ, г. Харьков, Украина, sgerasin@mail.ru

^{2, 3, 4} ХНУРЭ, г. Харьков, Украина

На основе представления нового подхода кодирования звуковой информации проведен сравнительный анализ его с широко известными методами кодирования звуковых данных. Приведены описания применяемых при этом методик сравнения, а также итоговые результаты испытаний и перспективы дальнейших исследований.

КОДИРОВАНИЕ, СЖАТИЕ АУДИОДАНЫХ, МЕТРИКИ, ТЕСТИРОВАНИЕ, СОНОГРАММА

Введение

Не так давно в университете проводились работы по разработке и тестированию эффективного способа кодирования информации. Основной задачей была разработка таких алгоритмов кодирования/декодирования сигналов различной природы, применение которых, например, в сфере звукозаписи позволило бы добиться большего на порядок коэффициента сжатия аудио данных по сравнению с ныне используемыми звуковыми форматами [1-4].

В представленном на тестирование способе многокомпонентного кодирования, как следует из документации, для получения высоких характеристик использовался ряд методологических предположений. Основная предпосылка заключалась в том, что сигнал любой природы представляет собой информационный объект, который нельзя разложить на составляющие элементы. Следующая методологическая предпосылка состояла в том, что не все цифровые значения сигнала имеют одинаковую информационную значимость. Поэтому часть цифровых значений сигнала должна фиксироваться в виде значимых компонентов. Другая часть может быть восстановлена при воспроизведении сигнала по информационно значимым компонентам. Алгоритм нахождения информационно значимых компонентов сигнала, а также восстановления всего множества значений составляет основную суть способа.

Способ многокомпонентного кодирования (CD-способ) применительно к различным информационным сигналам показал широкие возможности его применения в различных областях науки и техники (системах связи). Его можно использовать в системах речевой обработки сигналов, управления и хранения данных, а также в сжатии при передаче по каналам связи.

Теоретические результаты получили подтверждение в проведенной серии испытаний (тестов), которые опирались на методы компьютерного моделирования [5, 6]. Кроме этого были получены данные успешной работы способа с речевыми сигналами, звуковыми данными CD качества, не уступающие при воспроизведению оригиналам [7].

Для проведения тестовых испытаний работа на подготовленных заранее тестовых сигналах различного происхождения, специалисты университета проанализировали и отработали на практике общепринятые и хорошо известные тестовые методики проверки. Для сравнения на испытание были отобраны широко применяемые алгоритмы сжатия данных [8,9,10].

Основной целью было сравнение качества работы кодеков при кодировании музыки и речи, а также выявление особенностей кодирования, отражающихся на качестве звучания [11]. С таких же позиций рассматривался универсальный программный кодек, отражающий (моделирующий) новый способ многокомпонентного кодирования. Задачи проведения более критических анализов, с детальными оценками реализации способа, перед исследователями не ставились по той причине, что способ проходит этап патентования.

В ходе исследования (испытаний) применялись различные методики сравнения: от оценки качества при прослушивании до использования автоматических методов сравнения [1]. Конечно, в большинстве случаев чаще доверяют метрикам, которые основаны на оценках прослушивания именно музыкальными экспертами. Однако применение автоматических методов сравнения позволяет существенно ускорить темпы и качество разработки новых кодеков, в особенности для предотвращения появления в коде ошибок, влияющих на воспроизведение.

В рамках тестирования были использованы следующие кодеки:

Речевые кодеки:

- 1) GSM 6.10 (встроенный кодек windows 98 SE);
- 2) CELP (встроенный кодек windows 98 SE);
- 3) TrueSpeech (встроенный кодек windows 98 SE).

Универсальные кодеки:

- 1) MP3 кодек Lame 3.93 MMX (RazorLame V 1.1.5.1342);
- 2) MP3 кодек из пакета Blaze Media Convert 1.4 (BMC);
- 3) MP3 Pro Fraunhofer (CoolEdit 2.0 mp3pro);
- 4) Advanced Audio Coding (AAC) (производитель неизвестен);

5) Windows Media Audio (WMA) кодек из пакета Blaze Media Convert 1.4;

6) Yamaha Sound VQ Format (VQF);

7) Coder/decoder (CD) – программная модель способа многокомпонентного кодирования/декодирования.

Основная задача тестирования кодеков – сравнение качества сжатия нескольких наиболее известных и распространенных аудио-кодеков с размером получаемого потока данных 32kbps (сжатие порядка двадцати раз) из формата PCM 44100 Hz, 16 bit, mono, 705600 kbps. Такой поток теоретически позволяет передавать звук через модем в реальном времени.

1. Используемые метрики тестирования

Автоматическое тестирование проводилось с использованием двух основных метрик: временной и частотной Peak-Signal-to-Noise-Ratio (PSNR). Общая методика тестирования включала цикл компрессии-декомпрессии исходного файла исследуемыми кодеками с последующим сравнением сжатого и исходного файлов программами реализующими метрики.

Временная PSNR-метрика.

Значение временного PSNR для двух файлов, представленных массивами сэмплов ($a[i], i = 1..n; b[j], j = 1..m$), вычислялось по следующей формуле:

$$PSNR = 20 \times \text{Log} \left(\frac{\text{Max_sample_value}}{\sqrt{\frac{\sum_{i=1}^{\min(n,m)} (a[i] - b[i])^2}{\min(n,m)}}}} \right)$$

где Max_sample_value – максимальное значение амплитуды каждого сэмпла, в нашем случае равное $2^{15} = 32768$.

Частотно-временная PSNR-метрика.

Основная часть кодеков перед сжатием переводит сигнал в частотно-временное пространство, поэтому для них более корректно использовать метрику, работающую как с временными составляющими сигнала, так и с частотными.

Основная методика – сравнение отклонения амплитуд сигналов в частотно-временном пространстве. От исследуемых сигналов берется дискретное оконное преобразование Фурье с некоторым шагом по времени STFT (short-time Fourier transform), после чего вычисляются амплитуды по формуле:

$$A_n[j] = \log \left(\sqrt{\text{FFT}[j].\text{Im}^2 + \text{FFT}[j].\text{Re}^2} \right).$$

Каждый такой вектор является спектром сигнала небольшой области вокруг данной временной точки. Объединив все векторы A_n моментальных спектров в столбцы матрицы, можно получить сонограмму данного сигнала

$$\text{Sono}[i][j] = A_i[j].$$

Частотно-временное PSNR для двух сигналов рассчитывалось следующим образом:

$$PSNR = 20 \times \text{Log} \left(\frac{\text{Max_value}}{\sqrt{\frac{\sum_{i=0}^{\min(T_1, T_2)} \sum_{j=f_1}^{f_2} (\text{Sono}_1[i][j] - \text{Sono}_2[i][j])^2}{\min(T_1, T_2)}}}} \right)$$

где Max_value – амплитуда сигнала максимальной возможной мощности, допустимая в данном представлении звукового сигнала; T_1 и T_2 – количество векторов моментальных пиков для первого и второго сигнала соответственно; f_1 и f_2 – параметры, отфильтровывающие из общей спектрограммы частотную полосу для исследования. Преимущество данного метода заключается в том, что он не чувствителен к фазе сигнала, в результате чего подходит для кодеков, сохраняющих общее звучание, но не сохраняющих фазы (форму волны), а также в том, что появляется возможность рассмотрения искажений в отдельных частотных диапазонах.

2. Методы тестирования по сонограмме

Методика тестирования кодеков с потоком данных 32 кбит/с. по сонограмме. Изображения сонограмм исходного файла и конечного (прошедшего цикл компрессии/декомпрессии) сравнивались между собой. Большие отличия в сонограммах свидетельствуют о большем искажении звука в процессе кодирования. Для большей наглядности строилось, так называемое, изображение «разности сонограмм» массива разности значений сонограмм исходного и полученного сигналов. Соответственно, чем больше оказывались получаемые значения по модулю, тем больше отличался исходный файл при сжатии на данных частотах и моментах времени. Если при кодировании файла какие-то частоты исчезают, то на изображении разности сонограмм эта область помечается сиреневыми или красными цветами в зависимости от величины изменения. А если, наоборот, добавляются какие-то новые звуковые составляющие, которых не было в исходном сигнале, либо они имели меньшую амплитуду, то такие области помечаются синими и зелеными цветами.

Для упрощения изложения амплитуду сигнала будем рассматривать в логарифмической шкале.

Если имеются значения амплитуд сонограммы первого сигнала $A_1[t][f]$ и второго $A_2[t][f]$, где t – координата по времени, а f – по частоте, то разница изображений сонограммы получается так:

$$A_r[i][j] = A_1[i][j] - A_2[i][j] \quad i = t_1..t_2 \quad j = [0, 22050] \text{ Hz}.$$

Если $A_r[i][j] < 0$, то данная точка помечается фиолетовым цветом при малом абсолютном значении и красным, если $abs(A_r[i][j])$ велико. Также при $A_r[i][j] > 0$ цвет переходит от белого к синему, а потом к зеленому с увеличением значения $A_r[i][j]$.

На обычных сонограммах цвета переходят от белого к синему, потом к зеленому, красному и желтому по мере увеличения значений $A_n[i][j]$, $n \in 1, 2$.

Значения амплитуд $A_n[i][j]$ получаются из дискретного преобразования Фурье от исходного сигнала следующим образом:

$$A_n[i][j] = \log\left(\sqrt{FFT[j].Im^2 + FFT[j].Re^2}\right), i \in [0, T],$$

где i – время. FFT -вектор, полученный в результате обработки сигнала быстрым преобразованием Фурье.

Выводы и перспективы исследований

В итоге проделанной специалистами работы, можно сказать, что исследованные метрики не дают результата, полностью соответствующего человеческой оценке качества звука. Но результаты всех метрик в совокупности немного приближаются к корректному ответу (с точки зрения человека). Методы сравнения, сравнивающие только форму волны, оказываются наименее пригодными для сравнения звука среди всех метрик, описанных в этой статье. Однако по ним можно определить особенности сжатия того или иного кодека, например, по ним легко узнать, сохраняет ли кодек форму волны или нет. Следует отметить, что если кодек не сохраняет форму волны, то это совершенно не значит, что кодек некачественно сжимает файл, так как, например, шум можно сохранить как волну сложной формы, а можно просто записать параметры ее распределения и при декомпрессии восстановить шум с другой формой волны, но аналогичный по звучанию. Получается, что качественное сравнение кодеков с сохранением формы волны и без ее сохранения, по форме сигнала оказывается попыткой сопоставить практически несравнимые предметы, имеющие разные характеристики, как, например, сравнение кошки и собаки на предмет того, кто лучше. В этом случае мы лишь можем сказать о том, что исследуемые объекты имеют разные параметры. Использование кодеком психоакустической модели также вносит серьезные искажения в результаты тестирования, если в метрике не использовать информацию о таких моделях. В то же время, построение спектральных моделей и метрик на основе спектра может сильно помочь при анализе и классификации звука, а также при определении причин появления артефактов.

С точки зрения близости к оригиналу по описанным метрикам предпочтения следует отдать новому CD-способу. Он может определиться как

лучший по общим критериям сравнений.

В ходе испытаний программной модели CD-способа было установлено следующее:

1. *Работа со специальными сигналами.* В представленном спектре стандартных тестовых сигналов от 10 Гц до 22,5 кГц дополнительно вводились определенные функциональные «нагрузки» и тестовые характеристики. Получен максимальный коэффициент сжатия $K=3000$. Средний коэффициент на множестве специальных сигналов составил $K=2000$.

2. *Сжатие речевых сигналов.* На тестовых примерах сжатия речи был получен сравнительно высокий коэффициент сжатия, максимальный $K=246$. Тестировались примеры английской, немецкой, французской и русской связной речи. Средний коэффициент на множестве речевых сигналов составил $K=124$.

3. *Сжатие аудио-сигналов формата wav.* Тестовые примеры аудио-сигналов имели широкий жанрово-музыкальный спектр, при этом получен средний коэффициент $K=40$. Для музыкальных файлов данный способ кодирования может составить серьезную конкуренцию на рынке аудиозаписи. Он также вполне отвечает требованиям по применению в телефонии и системах радиосвязи.

4. Испытания показали, что представленный на исследование способ многокомпонентного кодирования и декодирования составляет серьезную конкуренцию формату mp3 и обеспечивает конкурентный уровень не только на рынке аудио-записей, но и рынке Embedded технологий.

Важным направлением совершенствования способов общения человека и ЭВМ является создание специальной аппаратуры и программ, обеспечивающих эффективный обмен информацией между ЭВМ и ее пользователем с помощью звучащей речи. Речь – это самый легкий и удобный способ общения, наиболее часто используемый человеком. Письмо, набор текста на компьютере – значительно менее мобильные и оперативные способы общения. Речь является самым быстрым способом связи, она эффективно действует на расстоянии, в темноте, она легко совмещается с другими действиями человека. Но эффективное решение проблемы автоматического речевого общения человека и ЭВМ невозможно без решения проблем, связанных с дискретизацией акустических колебаний, их отделения от различных шумов и помех и их кодирования таким образом, чтобы иметь возможность распознавать отдельные фонетические единицы.

Исходя из всего сказанного выше, можно сделать однозначный вывод: изучение свойств слухового восприятия человека является одной из важнейших задач, решение которой приведет к значительному облегчению общения человека и

ЭВМ и расширению сферы применения средств вычислительной техники.

В области изучения свойств слухового восприятия человека имеется множество задач, ждущих своего решения. Одной из них является выяснение принципов и математическое описание закономерностей, на основе которых человек воспринимает непрерывные колебания воздуха и превращает их в дискретную форму.

Данная работа посвящена изучению свойств и механизмов восприятия слуховым анализатором человека дискретизированных акустических колебаний и разработке системы кодирования звука, которая, опираясь на соответствующие научные исследования, позволяла бы принимать во внимание и кодировать тот минимальный объем информации, который необходим слуховому анализатору человека для восприятия кода звука как полноценной картины исходного звукового сигнала.

При вводе в ЭВМ осциллограмма звука кодируется с помощью АЦП-преобразователей. При этом диаграмма звукового давления заменяется решетчатой функцией. Численные значения отдельных линий решетчатой функции запоминаются вычислительной машиной. Такой способ кодирования звука получил повсеместное применение. Между тем, исследование механизма слухового анализатора человека показывает, что природа использует совершенно иной способ кодирования звука, а именно - частотно-импульсный. Если с помощью микроэлектронной техники определить, какая информация передается по отдельным волокнам слухового нерва, то обнаружится, что по ним следуют одиночные электрохимические импульсы, причем тем реже, чем ниже громкость сигналов, и наоборот. Каждый импульс возникает в момент накопления некоторой фиксированной порции площади под кривой звукового давления, отсчитываемой от момента возникновения предыдущего импульса. Исследования показывают, что частотно-импульсный метод кодирования звука выгодно отличается от применяемого повсеместно в настоящее время АЦП как в смысле повышенной помехоустойчивости, так и в смысле экономности при хранении в памяти ЭВМ. Установлено, что частотно-импульсное кодирование звука обеспечивает, как минимум, семикратную экономию памяти при

хранении кода звука по сравнению с АЦП-кодированием. Значительная экономия достигается также за счет уменьшения трудоемкости обработки звуковых кодов.

Широкое практическое использование дискретизации речевых сигналов выдвигает для научной проработки ряд актуальных вопросов:

— можно ли строго обосновать правомерность дискретизации звука;

— является ли использование аналого-цифрового преобразователя стандартного типа наилучшим способом дискретизации звукового сигнала;

— с какой точностью и частотой следует проводить дискретизацию и на чем основывать выбор численных значений параметров дискретизации?

Анализ литературных данных показывает, что ни на один из этих вопросов пока нет исчерпывающих ответов.

Очевидно, что ответ на все эти вопросы может быть получен лишь на пути психофизического изучения свойств слухового восприятия.

Список литературы: 1. *Ватолин Д., Ратушняк А., Смирнов М.* Методы сжатия данных. — М.: Диалог-МИФИ, 2003. — 382 с. 2. *Ковалгин Ю.А.* Алгоритмы компрессии цифровых аудиоданных // Системы и средства связи, телевидения и радиовещания. — 2000. — № 3. — С. 17-29. 3. *Ковалгин Ю.А.* Компрессия цифрового звука: психоакустические основы и алгоритмы // Звукорежиссер. — 2000. — № 6. — С. 56-63. 4. *Ковалгин Ю.А., Вологдин Э.И.* Цифровое кодирование звуковых сигналов. — Санкт-Петербург, КОРОНА, принт. -2004. 5. *Jurgen Herre, Michael Schug.* Analysis of Decompressed Audio - The "Inverse Decoder" Presented at the 109th Convention 2000 September 22-25 Los Angeles, California, USA. 6. *M. Kowalski and B. Torr.esani* A STUDY OF BERNOULLI AND STRUCTURED RANDOM WAVEFORM MODELS FOR AUDIO SIGNALS // LAMP, CMI, 39 rue Joliot-Curie, 13453 Marseille cedex 13, France 2005. 7. Audio Test and Demonstration CD SOUND CHECK 2 // <http://www.interstudio.co.uk/schck2a.htm> 8. *Church S.* ISDN and ISO/MPEG Layer III Audio Coding: Powerful New Tools for Broadcast Audio Production / AES an audio engineering society preprint/ The 95th Convention/ — New-York, 1993. — October. — P. 7-10. 9. Digital Audio Compression Standard (AC-3) // Doc.A/52.-1995-12-20. 10. *Davis Pan,* "A Tutorial on MPEG/Audio Compression", IEEE Multimedia, Vol. 2, No. 2, pp. 60-74, 1995. 11. Structured Audio: Creation, Transmission, and Rendering of Parametric Sound Representations (1998) *Barry Vercoe, William Gardner, Eric Scheirer* // <http://citeseer.ist.psu.edu/context/342384/442890>.

Поступила в редколлегию 16.05.2008