ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

Нижегородский государственный университет им. Н.И. Лобачевского

**Факультет вычислительной математики и кибернетики**

**Кафедра математической логики и высшей алгебры**

**РЕФЕРАТ**

для сдачи кандидатского экзамена по английскому языку

(Перевод)

Выполнил:

Аспирант 2-го года обучения

Мартынов Игорь Михайлович

Специальность 01.01.09 – Дискретная математика и математическая кибернетика

Н.Новгород

2015

май

Реферат выполнен по журналу:

Digital Technical Journal Vol. 5 No. 2, Spring 1993

Davis Yen Pan – Digital Audio Compression

# Сжатие цифрового звука

Дэвис Йен Пан (Davis Yen Pan)

Аннотация

По сравнению с большинством типов цифровых данных, за исключением цифрового видео, объём данных, связанный с несжатым цифровым звуком, существенен. Сжатие цифрового звука позволяет более эффективно хранить и передавать аудио данные. Множество видов техник сжатия звука предлагают широкий диапазон сложности кодера и декодера, качества сжатого звука и коэффициентов сжатия. µ-преобразование и ADPCM-кодер – простые подходы с низкой сложностью, низким коэффициентом сжатия и алгоритмами среднего качества звука. Стандарт MPEG – алгоритм высокой сложности, высоким коэффициентом сжатия и высоким качеством звука. Эти техники применимы к звуковым сигналам общего вида, и не приспособлены специальным образом для сжатия речи.

Введение

Сжатие цифрового звука позволяет рационально хранить и передавать звуковые данные. Разнообразные методы сжатия звука предлагают различные уровни сложности, качества сжатого звука и величины сжатия.

Эта статья – исследование методов, используемых для сжатия цифровых звуковых сигналов. Её цель – предоставить полезную информацию для читателей любого уровня опыта в обработке цифрового звука. Статья начинается с обзора начального процесса оцифровки звука. Два следующих раздела представляют детальное описание двух относительно простых подходов к сжатию звука: µ-закон и адаптивной дифференциальной импульсно-кодовой модуляции (adaptive differential pulse code modulation, ADPCM). В следующем разделе статьи рассматривается третий, гораздо более изощрённый, алгоритм сжатия звука от Motion Picture Experts Group (MPEG). Вопросы, охваченные в этом разделе, довольно сложны, и рассчитаны на читателя, знакомого с цифровой обработкой сигналов. Статья завершается обсуждением программных реализаций реального времени.

Цифровые аудио данные

Цифровое представление звуковых данных предлагает много преимуществ: высокую шумоустойчивость, стабильность и воспроизводимость. Звук в цифровом виде также позволяет эффективно реализовывать многие процедуры обработки звука (например, смешивание, фильтрацию и стабилизацию) с помощью цифровой ЭВМ.

Преобразование из аналоговой в цифровую область начинается с замеров входного звука через регулярные, дискретные промежутки времени и отображения замеренных значений на конечное число равномерно распределённых уровней. Цифровые звуковые данные состоят из последовательности двоичных значений, представляющих число уровней квантизации для каждого звукового сэмпла. Способ представления каждого сэмпла с помощью независимого машинного слова называется импульсно-кодовой модуляцией (pulse code modulation, PCM). Схема 1 демонстрирует процесс оцифровки звука.

Схема 1. Процесс оцифровки звука

Согласно теории Найквиста (*прим: имеется в виду теорема Котельникова; в англоязычной литературе – теорема Найквиста-Шеннона*), сэмплированный сигнал может верно представлять сигналы частотой вплоть до половины частоты сэпмлирования. Типичные частоты сэмплирования лежат в диапазоне от 8 килогерц (кГц) до 48 кГц. Частота в 8 кГц покрывает диапазон частот вплоть до 4 кГц, и поэтому покрывает большую часть частот, производимых человеческим голосом. Частота в 48 кГц покрывает диапазон частот вплоть до 24 кГц, что более чем достаточно для покрытия всего диапазона слышимых частот, который, для людей, обычно простирается только до 20 кГц. На практике, диапазон частот несколько меньше, чем половина частоты сэмплирования из-за практических ограничений системы.

Число уровней квантизации обычно является степенью двойки, чтобы полностью использовать фиксированное число бит на звуковой сэмпл для представления квантизованных уровней. С равномерным распределением шагов квантизации, каждый дополнительный бит может увеличить отношение сигнал-шум, или динамический диапазон, квантизованной амплитуды примерно на 6 децибел (6 дБ). Типичное число бит на сэмпл, используемое для цифрового звука, лежит в диапазоне от 8 до 16. Таким образом, возможности динамического диапазона лежит от 48 до 96 дБ соответственно. Если взглянуть на эти диапазоны со стороны, если 0 дБ соответствует уровню звукового давления самого слабого слышимого шума, то 25 дБ – минимальный уровень шума в обычной звукозаписывающей студии, 25 дБ – уровень шума внутри тихого дома, а 120 дБ – самый громкий звук до появления ощущения неудобства. С точки зрения восприятия звука, 1 дБ – минимальное слышимое изменение звукового давления в идеальных условиях, и удвоение звукового давления соответствует одному шагу восприятия громкости.

По сравнению с большинством типов цифровых данных (включая цифровое видео), объём данных, связанный с несжатым цифровым звуком, существенен. Например, звуковые данные на компакт-диске (2 звуковых канала, сэмплированные с частотой 44.1 кГц с 16 битами на сэмпл) требует потока данных примерно в 1.4 мегабит в секунду. Ясна необходимость какой-то формы сжатия, чтобы эффективно хранить и передавать эти данные.

Многие виды техник сжатия звука отличаются компромиссом между сложностью кодера и декодера, качеством сжатого звука и величиной сжатия. Техники, представленные в следующих разделах этой статьи описывают полный диапазон, от µ-закона – алгоритма низкой сложности, низкой степени сжатия и средним качеством, до MPEG – алгоритма высокой сложности, высокой степени сжатия и высокого качества звука. Эти техники применимы к звуковым сигналам общего вида, и не заточены специально для речевых сигналов. Эта статья не описывает алгоритмы сжатия звука, разработанные специально для речевых сигналов. Эти алгоритмы обычно основаны на моделировании речевого аппарата, и не работают хорошо для неречевых звуковых сигналов. Федеральные стандарты 1015 LPC (linear predictive coding, кодирование с линейным предиктором) и 1016 CELP (coded excited linear prediction, линейное предсказание с мультикодовым управлением) попадают в эту категорию.

Сжатие звука через µ-преобразование

µ-закон – базовый метод сжатия звука, определённый рекомендацией G.711 Международного Консультативного Комитета по Телефонии и Телеграфии (Comité Consultatif International Téléphonique et Télégraphique, CCITT). Преобразование в сущности является логарифмическим по характеру, и позволяет с помощью 8 бит на сэмпл покрывать динамический диапазон, эквивалентный 14 битам линейно квантизованных значений. Это преобразование предоставляет коэффициент сжатия, равный (числу бит на исходный сэмпл)/8 к 1. В отличие от линейной квантизации, логарифмической распределение шагов представляет низкоамплитудные звуковые сэмплы с большей точностью, чем высокоамплитудные значения. Таким образом, отношение сигнал-шум преобразованного выхода более равномерно на диапазоне амплитуд входного сигнала. µ-закон есть:

где m = 255, и x – значение входного сигнала, нормализованного, чтобы иметь максимальное значение 1. Рекомендация G.711 CCITT также определяет подобный A-закон. µ-закон имеен широкое применение в Северной Америке и Японии для сэмплированного на 8 кГц голосового цифрового телефонного сервиса ISDN (Integrated Services Digital Network), а A-закон применяется в других местах для ISDN-телефонии.

Адаптивная дифференциальная импульсно-кодовая модуляция (Adaptive Differential Pulse Code Modulation, ADPCM)

Схема 2 демонстрирует упрощённую блок-схему кодера адаптивной дифференциальной импульсно-кодовой модуляции (ADPCM). Для ясности, схема опускает детали, такие как форматирование битового потока, возможное использование сторонней информации, и адаптивных блоков. ADPCM-кодер пользуется тем фактом, что соседние звуковые сэмплы обычно похожи друг на друга. Вместо представления каждого звукового сэмпла независимо, как в PCM, ADPCM-кодер вычисляет разницу между каждым звуковым сэмплом и его предсказанным значением, и выводит PCM-значение этой разницы. Заметьте, что ADPCM-кодер (Схема 2a) использует большую часть компонент ADPCM-декодера (Схема 2b) для вычисления предсказанных значений.

Схема 2. Сжатие и разжатие в ADPCM

Выход квантизатора – обычно только (знаковое) представление числа уровней квантизации. Реквантизатор восстанавливает значения квантизованного сэмпла, умножая число уровней квантизации на размер шага квантизации, и, возможно, добавляя отступ в половину шага. В зависимости от реализации квантизатора, этот отступ может быть необходим для центрирования восстановленного значения между порогами квантизации.

ADPCM-кодер может адаптироваться к характеристикам звукового сигнала, изменяя размер шага квантизатора, предсказателя, или их обоих. Метод вычисления предсказанного значения и способ, которым предсказатель и квантизатор адаптируются к звуковому сигналу, различаются среди разных систем кодирования ADPCM.

Некоторые ADPCM-системы требуют, чтобы кодер предоставлял стороннюю информацию вместе с разностными PCM-значениями. Эта сторонняя информация может служить двум целям. Во-первых, в некоторых ADPCM-схемах декодеру необходима дополнительная информация, чтобы установить размер шага предиктора или квантизатора, или их обоих. Во-вторых, эти данные могут предоставлять декодеру избыточную контекстную информацию, что позволяет восстанавливаться после ошибок в битовом потоке или предоставляет прямой доступ к закодированному потоку битов.

Следующий раздел описывает алгоритм ADPCM, предложенный Ассоциацией интерактивного мультимедиа (The Interactive Multimedia Association, IMA). Этот алгоритм обеспечивает коэффициент сжатия, равный (числу бит на исходный сэмпл)/4 к 1. Среди других ADPCM-схем сжатия звука – рекомендация G.721 CCITT (поток сжатых данных 32 килобита в секунду), рекомендация G.723 (поток сжатых данных 24 килобит в секунду) и алгоритм сжатия интерактивного аудио на компакт-диске.

*Алгоритм ADPCM IMA.* IMA – это консорциум производителей компьютерного оборудования и программного обеспечения, сотрудничающих с целью разработать стандарт де-факто для компьютерных мультимедиа-данных. Цель IMA для их предложения по сжатию звуковых данных была выбрать публично доступный алгоритм сжатия звука, способный обеспечить хорошее качество сжатого звука с хорошей производительностью сжатия данных. Кроме того, алгоритм должен был быть достаточно простым для программного расжатия, в реальном времени, звукового сигнала 44.1 кГц на 20-мегагерцевом (МГц) компьютере 386-класса. Выбранный алгоритм ADPCM не только отвечает этим целям, но и достаточно прост для программного сжатия в реальном времени на том же компьютере.

Простота предложения IMA ADPCM обусловлена грубостью его предиктора. Предсказываемое значение звукового сэмпла – просто декодированное значение следующего прямо перед ним сэмпла. Таким образом, блок предиктора на Схеме 2 – всего лишь элемент временной задержки, чей выход – это вход, отложенный на один интервал сэмплирования. Поскольку предиктор не адаптивен, сторонняя информация не нужна для восстановления предиктора.

Схема 3 демонстрирует блок-схему процесса квантизации, используемого алгоритмом IMA. Квантизатор выдаёт четыре бита, представляющих знаковую величину числа уровней квантизации для каждого входного сэмпла.

Адаптация к звуковому сигналу происходит только в блоке квантизации. Квантизатор подстраивает размер шага в зависимости от текущего размера шага и вывода квантизатора с предыдущего входа. Это подстройка может быть выполнена как два последовательных поиска в таблице. Три бита, представляющие число уровней квантизации, служат индексом для первого поиска, чей выход – смещение для второго поиска. Это смещение прибавляется к сохранённому значению индекса, и ограниченный по диапазону результат используется для второго поиска в таблице. Суммарное значение индекса сохраняется для использования на следующей итерации подстройки размера шага. Результат второго поиска в таблице – это новый размер шага квантизации. Заметьте, что имея начальное значение индекса для второго поиска в таблице, данные для подстройки полностью выводятся из выхода квантизатора; сторонняя информация не нужна для подстройки квантизатора. Схема 4 изображает блок-схему процесса подстройки размера шага, а таблицы 1 и 2 демонстрируют содержимое таблиц для поиска.

Схема 3. Квантизация IMA ADPCM

Таблица 1. Первый поиск в таблице для адаптации квантизатора IMA ADPCM

|  |  |
| --- | --- |
| **Трёхбитовая квантизованная величина** | **Смещение индекса** |
| 000 | -1 |
| 001 | -1 |
| 010 | -1 |
| 011 | -1 |
| 100 | 2 |
| 101 | 4 |
| 110 | 6 |
| 111 | 8 |

Таблица 2. Второй поиск в таблице для подстройки квантизатора IMA ADPCM

*IMA ADPCM: Восстановление от ошибок.* Как удачный побочный эффект архитектуры этой ADPCM-схемы, ошибки декодера, вызванные изолированными ошибками в кодовых словах или правками, склейками или случайным доступом к сжатому потоку бит, обычно не имеют катастрофического эффекта на выход декодера. Обычно это неверно для схем сжатия, использующих предсказание. Поскольку предсказание опирается на корректное декодирование предыдущего звукового сэмпла, ошибки в декодере склонны распространяться. Следующий раздел объясняет, почему распространение ошибок обычно ограниченно, и не гибельно для алгоритма IMA. Декодер восстанавливает звуковой сэмпл, , прибавляя предыдущий декодированный сэмпл, , к результату знаковой величины произведения кодового слова, , и размера шага квантизации, плюс смещение в половину размера шага:

где одной второй, плюс подходящее численное представление .

Анализ второго поиска в таблице размера шага открывает, что каждая следующая запись примерно в 1.1 раз больше предыдущей. Поскольку диапазон, ограничивающий индекс второй таблицы, не имеет места, значение примерно равно произведению предыдущего значения, , и функции от кодового слова, :

Два уравнения выше можно преобразовать, чтобы выразить декодированный сэмпл, , как функцию от размера шага и декодированного значения сэмпла в момент времени , и набор кодовых слов в диапазоне времени от до .

Заметьте, что элементы суммы – функция только от кодовых слов времени и далее. Ошибка в кодовом слове, , или случайная запись в поток битов во время , может привести к ошибке в разжатом выводе, , и размере шага квантизации, . Уравнение выше показывает, что ошибка в равняется константному отступу в будущих значениях . Это смещение неслышимо, если разжатое значение не выходит за рамки допустимого диапазона и не обрезается. Такое отсечение приводит к кратковременному слышимому искажению, но также служит для частичной или полной корректировки смещения. Уравнения выше также показывает, что ошибка в равняется нежелательному приросту или ослаблению будущих значений декодированного результата . Форма волны выходного сигнала не меняется, если индекс для второй таблицы поиска размера шага не выходит за рамки диапазона. Ограничение по диапазону приводит к частичной или полной коррекции значения размера шага.