

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ**

**ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО**

**ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ**

**«Донской государственный технический университет»**

**(ДГТУ)**

**Доклад на тему**

**«Сжатие звука»**

(теория информации и кодирования)

Выполнил: студент ВМО21  
Оганесьянц К.П.

Проверил: Ляхницкая О.В.

Ростов-на-Дону

2021

**Введение**

Звук является простой волной, а оцифрованный звук — цифровое представление этой волны. Это достигается запоминанием уровня аналогового сигнала множество раз в течение одной секунды. Например, в обыкновенном CD сигнал запоминается 44100 раз за секунду. Так как CD работает со стерео, мы запоминаем сигнал для левой и правой колонки параллельно. Для каждого замера используются 16-битовые числа. Поэтому одна секунда звучания занимает 2 × 2 × 44100 = 176 400 байт.

Сжатие (компрессия) аудиоданных представляет собой процесс уменьшения скорости цифрового потока за счет сокращения статистической и психоакустической избыточности цифрового звукового сигнала.

Методы сокращения статистической избыточности аудиоданных так же называют сжатием без потерь, а, соответственно, методы сокращения психоакустической избыточности - сжатием с потерями.

**1. Сжатие без потерь**

Сокращение психоакустической избыточности основано на учете свойств самих звуковых сигналов. Она определяется наличием корреляционной связи между соседними отсчетами цифрового звукового сигнала, устранение которой позволяет сокращать объем передаваемых данных на 15...25% по сравнению с их исходной величиной. Для передачи сигнала необходимо получить более компактное его представление, что возможно осуществить с помощью ортогонального преобразования. Важными условиями применения такого метода являются:

* возможность восстанавливать исходный сигнал без искажений
* способность обеспечивать наибольшую концентрацию энергии в небольшом числе коэффициентов преобразования
* быстрый вычислительный алгоритмом

Этим требованиям отвечает модифицированное дискретно-косинусное преобразование (МДКП).

Уменьшить скорость цифрового потока позволяют методы кодирования, учитывающие статистику звуковых сигналов, например, вероятности появления уровней разной величины. Одним из таких методов является код Хаффмана, где наиболее вероятным значениям сигнала приписываются более короткие кодовые слова, а значения отсчетов, возможность появления которых мала, шифруется кодовыми словами большей длины. Именно в силу этих двух причин в наиболее эффективных алгоритмах компрессии цифровых аудиоданных кодированию подвергаются не сами отсчеты звукового сигнала, а коэффициенты МДКП.

**2. Сжатие с потерями**

Сжатие аудиоданных с потерями основывается на несовершенстве человеческого слуха при восприятии звуковой информации. Неспособность человека в определенных случаях различать тихие звуки в присутствии более громких, называемая эффектом маскировки, была использована в алгоритмах сокращения психоакустической избыточности. Эффекты слухового маскирования зависят от спектральных и временных характеристик маскируемого и маскирующего сигналов и могут быть разделены на две основные группы:

* частотное (одновременное) маскирование
* временное (неодновременное) маскирование

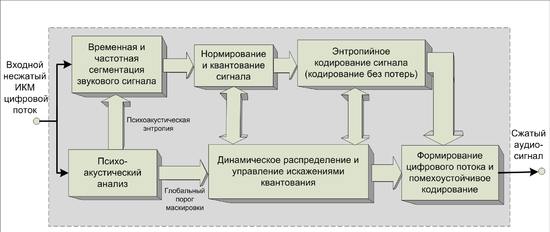
Эффект маскирования в частотной области связан с тем, что в присутствии больших звуковых амплитуд человеческое ухо нечувствительно к малым амплитудам близких частот. То есть, когда два сигнала одновременно находятся в ограниченной частотной области, то более слабый сигнал становится неслышимым на фоне более сильного.

Маскирование во временной области характеризует динамические свойства слуха, показывая изменение во времени относительного порога слышимости (порог слышимости одного сигнала в присутствии другого), когда маскирующий и маскируемый сигналы звучат не одновременно. При этом следует различать явления послемаскировки (изменение порога слышимости после сигнала высокого уровня) и предмаскировки (изменение порога слышимости перед приходом сигнала максимального уровня). Более слабый сигнал становится неслышимым за 5 − 20 мс до включения сигнала маскирования и становится слышимым через 50 − 200 мс после его включения.

Наилучшим методом кодирования звука, учитывающим эффект маскирования, оказывается полосное кодирование. Сущность его заключается в следующем: группа отсчетов входного звукового сигнала, называемая кадром, поступает на блок фильтров, который разделяет сигнал на частотные поддиапазоны. На выходе каждого фильтра оказывается та часть входного сигнала, которая попадает в полосу пропускания данного фильтра. Далее, в каждой полосе с помощью психоакустической модели, анализируется спектральный состав сигнала и оценивается, какую часть сигнала следует передавать без сокращений, а какая лежит ниже порога маскирования и может быть переквантована на меньшее число бит. Для сокращения максимального динамического диапазона определяется максимальный отсчет в кадре и вычисляется масштабирующий множитель, который приводит этот отсчет к верхнему уровню квантования. Эта операция аналогична компандированию в аналоговом вещании. На этот же множитель умножаются и все остальные отсчеты. Масштабирующий множитель передается к декодеру вместе с кодированными данными для коррекции коэффициента передачи последнего. После масштабирования производится оценка порога маскирования и осуществляется перераспределение общего числа битов между всеми полосами.

После устранения психоакустической избыточности звуковых сигналов их точное восстановление при декодировании уже невозможно. Методами устранения психофизической избыточности можно обеспечить сжатие цифровых аудиоданных в 10 − 12 раз без существенных потерь в качестве.

**2.1. Структура кодера сжатия аудиоданных с потерями**



*Обобщенная структура кодера звукового сигнала с компрессией цифровых аудиоданных*

Исходный цифровой звуковой сигнал разделяется на частотные поддиапазоны и сегментируется по времени в блоке временной и частотной сегментации.

Длина кодируемой выборки зависит от формы временной функции звукового сигнала. При отсутствии резких выбросов по амплитуде используется так назы­ваемая длинная выборка, обеспечивающая высокое разрешение по частоте. В случае же резких изменений амплитуды сигнала длина кодируемой выборки резко уменьшается, что дает более высокое разрешение по времени. Решение об изменении длины кодируемой выборки принимает блок психоакустического анализа, вычисляя значение психоакустической энтропии сигнала.

После сегментации сигналы частотных поддиапазонов нормируются, квантуются и кодируются. В наиболее эффективных алгоритмах компрессии кодированию подвергаются не сами отсчеты выборки звукового сигнала, а соответствующие им коэффициенты МДКП.

Учет закономерностей слухового восприятия звукового сигнала выполняется в блоке психоакустического анализа. Здесь по специальной процедуре для каждого частотного поддиапазона рассчитывается максимально допустимый уровень искажений (шумов) квантования, при котором они еще маскируются полезным сигналом данного поддиапазона.

Блок динамического распределения бит в соответствии с требованиями психоакустической модели для каждого поддиапазона кодирования выделяет такое минимально возможное их количество, при котором уровень искажений, вызванных квантованием, не превышал порога их слышимости, рассчитанного психоакустической моделью.

Ещё могут использоваться:

матрицирование стерео - сложение и вычитание левого и правого канала для устранения повторяющейся информации

специальные процедуры итерационных циклов, позволяющие управлять величиной энергии искажений квантования в поддиапазонах при недостаточном числе доступных для кодирования бит

процедуры линейного и обратного адаптивного предсказаний

техника сглаживания переходных шумов во временной области (Temporal Noise Shaping - TNS), позволяющая управлять микроструктурой искажений квантования внутри каждого поддиапазона кодирования

Многие другие приёмы могут послужить способом сокращения объёма данных звуковой информации. Даже простое сужение полосы частот сигнала вместе с уменьшением динамического диапазона может называться сжатием аудиоданных. Например, в стандарте сжатия звука в сотовой связи используется и то и другое. Стремясь удалить избыточность из звука, кодек при плохом качестве сигнала становится избирателен к определённым словам, упорно проглатывая их.

**3. Субъективная оценка качества**

Для сжатых аудиоданных существует субъективная оценка качества, оцениваемая как процент людей, почувствовавших разницу с оригиналом.

Соответствие процента заметивших на пробе сжатой MP3 кодеком и битрейтом, режим Stereo.

- 0 - 1 %% - 320 kb/s

- 5 - 30 %% - 256 kb/s

- 30 - 40 %% - 192 kb/s

- 40 - 70 %% - 128 kb/s

Следует учесть тот факт, что качество получившегося материала зависит от характера сжимаемых данных, от жанра, наличия фона, помех. После сжатия, например MP3, на средних битрейтах, слушатели отмечают оловянность перкуссионных. А на голосе сжатие (даже сильное) отражается мало.

**Алгоритм Райса**

Одним из базовых алгоритмов сжатия аудио – алгоритм Райса, который состоит из нескольких частей – подготовки и самого кодирования соответственно.

**1. Преобразование координат L, R → X, Y**

Первым шагом в алгоритме сжатия будет представление правого и левого каналов в форме, более удобной для кодирования и последующего декодирования – L R каналы необходимо представить в виде некоторых чисел X Y, согласно след. алгоритму:

Для дробных чисел такого рода преобразования не теряют информации и остаются эквивалентными оригинальным. Но для целых чисел конвертация в X теряет 0.5 при переводе в целочисленный формат, если L и R имеют разную четность, однако заранее проверив четность можно восполнить потерю в 0.5.

**2. Предиктор**

Следующий шаг - пропустить X и Y через алгоритм, который максимально эффективно уберёт весь избыток информации в представлении X, Y.

В данном случае процесс направлен на преобразование массивов X и Y таким образом, чтобы они были представлены минимально возможными числами с возможностью обратить процесс. Существует несколько различных вариантов осуществить такое преобразование. Одним из простых являются преобразования с помощью линейной алгебры:

PX = (2 \* X−1) − X−2  
PY = (2 \* Y−1) − Y−2  
  
Если X = (2, 8, 24, ?), то в ряду PX на четвертом месте будет

P4 = (2 \* X4-1) − X4-2 = (2 \* 24) − 8 = 40  
То есть, если X = (2,8,24,27,25,28,21,17), то PX = (2,8,14,40,30,…)

При проведении преобразований необходимо учитывать особенности изначальных X и Y, так как может возникнуть ситуация, при которой на какой-то части спектра в какой-то момент преобразований может возникнуть разница между соседними частотами, что может усложнить кодирование.

Пусть число m лежит в диапазоне 0 … 1024. Для массива PX выполняется серия преобразований с разными значениями m следующим образом:  
  
X = (2, 8, 24, ?), тогда соответственно  
PX = (2 \* X−1) − X−2 = (2 \* 24) − 8 = 40  
  
Если ? = 45 и m = 512, тогда конечное значение = ? − ( P X ∗ m / 1024 ) = 45 − ( 40 ∗ m / 1024 ) = 45 − ( 40 ∗ 512 / 1024 ) = 45 − 20 = 25 {\displaystyle ?-(PX\*m/1024)=45-(40\*m/1024)=45-(40\*512/1024)=45-20=25}

Идентичным образом проводится поиск оптимального значения m, так как большие значения могут быть эффективнее.

После получения для определенного m массив данных, происходит увеличение или уменьшение m в зависимости от того, была ли последняя попытка в алгоритме удачной.

При использовании различных уравнений и большого колличества проходов по разным коэффициентам можно добиться хорошего сжатия данных.

**3. Кодирование. Алгоритм Райса**

Идея сжатия аудио заключается в представлении чисел, соответствующих потоку минимально возможным образом, убрав предварительно любую корреляцию данных. После этого можно записывать поток закодированных данных на диск. Одним из самых эффективных способов является кодирование Райса.

Меньшие числа предпочтительней тем, что их представление в бинарном представлении короче. Например, необходимо закодировать следующий ряд:

Базис по основанию 10: 10, 14, 15, 46

Или тот же ряд в бинарном виде

Базис по основанию 2: 1010, 1110, 1111, 101110

Теперь если требуется представить этот в виде строки, где для каждого числа зарезервировано 32 бита диапазон всех возможных значений, это будет неэффективно, поскольку понадобится 128 бит. Однако существует более эффективный метод. Наилучшим решением было бы просто записать бинарные числа 1010, 1110, 1111, 101110 без запятых, получив ряд вида 101011101111101110. Проблема в том, что после нет возможности узнать границы каждого числа. В качестве решения подобной задачи, как правило, используется алгоритм Райса. Кодирование Райса - это способ представить маленькие числа одной строкой, сохраняя способность их различать. Примечание: алгоритм тем эффективнее, чем меньше числа, поэтому необходимо изначально позаботиться об этом

На каком-то этапе кодирования данные представлены в виде числа n. Закодированное, оно добавляется справа к строке уже закодированных чисел таким образом, чтобы был возможен обратный процесс.

Основной идеей данного метода является представление числа n как , чтобы

Лучшие алгоритмы связаны со сжатием аудио в стандартах MPEG

**Форматы**

*MP3*

МРЗ − аудио MPEG1 уровня 3 (level 3), как его обычно называют, предлагает сжатие с коэффициентом примерно 10:1, сохраняя высокое качество. МР3 имеет собственный файловый формат, в котором сжатый аудиопоток расщепляется на фрагменты, именуемые кадрами. Каждый из них имеет заголовок, в котором указываются скорость передачи, частота дискретизации и другие параметры. Файл также может включать теги метаданных, ориентированные на музыкальное содержимое и указывающие заголовок трека, его исполнителя, альбом, из которого взята композиция, и т. д. Файлы МР3 широко используются для загрузки и хранения музыки на компьютерах и mp3-плейерах, сотовых телефонах. Этот формат широко распространен в Интернете.

*MP3pro*

Данный формат был создан вовсе не для того, чтобы заменить mp3, позволяя лишь добиться приемлемого качества звучания на низких битрейтах. Если "классический" битрейт в 128 Кb/s некоторыми слушателями и воспринимается как дающий качество, близкое к идеальному, то даже незначительное его понижение вызывает появление большого количества отчётливо слышимых искажений. Для передачи музыки в Интернете используются обычно именно низкие битрейты, которые являются далеко не сильной стороной "обычного" mp3. Здесь MP3pro и проявляет себя с лучшей стороны. Для хранения музыки высокого качества MP3pro совершенно не годится: даже при использовании максимально доступного для большинства кодеков битрейта 96 Кb/s слышны искажения, хотя по сравнению со многими другими форматами, поддерживающими низкие битрейты, результаты работы кодека MP3pro заметно лучше. При повышении битрейта качество файлов падает по сравнению с остальными форматами, и уже при битрейтах 128 Кb/s разумнее использовать mp3.

*ACC*

Формат Advanced Audio Coding (расширенное аудиокодирование), также известный как MPEG2, является преемником формата mp3. В отличие от MP3 у ААС большее сжатие при меньших скоростях передачи данных. Качество ААС всегда ставится выше качества MP3 при одинаковых скоростях передачи битов, а одинаковые оценки оба кодека получают, только если ААС используется при меньших скоростях передачи, чем MP3. Кодек ААС был встроен в MPEG4, где он является основой для кодирования природного аудио (в противоположность речи и синтезированным звукам). Тем не менее, в использовании AAC есть свои трудности: алгоритмы кодирования, используемые в данном формате, достаточно сложны, поэтому для создания AAC-файла требуется значительное количество времени и системных ресурсов.

*AIFF*

Это стандартный формат файлов для сохранения аудиоданных на платформе Macintosh. Расшифровывается как Audio Interchange File Format (формат обмена звуковыми файлами). Если вам когда-нибудь потребуется пересылать аудиофайлы между персональным компьютером и компьютером Macintosh, используйте именно этот формат. Он поддерживает 8- и 16-битные монофонические и стереофонические аудиоданные. В этом формате можно хранить не только сами звуковые файлы, но и информацию об используемых частотах и разрешении дискретизации.

*WAVE*

Формат Wave является форматом для операционной системы Windows, а это означает, что любой компьютер с операционной системой Windows может воспроизводить файлы Wave. Формат поддерживает множество различных типов аудиоданных, в том числе 8- и 16-битные, моно и стерео. Так же, как и в AIFF, в этом формате можно хранить звуковые файлы вместе со всеми частотами и разрешениями дискретизации звука. Wave эквивалентен AIFF по качеству и коэффициенту сжатия, и выбор из них зависит только от платформы, на которой создаётся звуковой файл. Файлы в формате Wave имеют расширение wav.