

ГУАП

КАФЕДРА № 42

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ
ПРЕПОДАВАТЕЛЬ

Старший преподаватель
должность, уч. степень, звание

подпись, дата

Т. А. Суетина
инициалы, фамилия

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ

«Сравнительные исследования кодеков, позволяющие осуществить

сжатие звукового файла в формате MP3»

по курсу: Техника аудиовизуальных средств информации

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ гр. № 4128

подпись, дата

В.А. Воробьев
инициалы, фамилия

Санкт-Петербург 2024

1. Цель работы

Получить теоретические знания по представлению звуковой информации на электронных носителях в формате MP3, алгоритмах сжатия и кодеках, позволяющих осуществлять сжатие звукового файла в формате MP3.

2. Задание

1 Описать принципы построения кодеков звука с потерями, основанных на особенностях восприятия звука человеком.

2 Сформулировать принципы построения психоакустической модели кодека.

3 Выполнить обзор бесплатных кодеков звука формата MP3.

4 Рассмотреть принципы цифрового сжатия звука, а также достоинства и недостатки широко распространенных кодеков звука.

3. Ход работы

3.1 Перцептивные аудиокодеки

В современном мире существует два типа (парадигмы развития) аудиостандартов:

- Lossless – форматы
- Форматы на базе перцептивного восприятия

Первые являются стандартами, пытающимися максимально точно воспроизвести исходный аудио-файл, а сжатие в данном случае происходит лишь за счет статистической избыточности (redundancy) (пример – FLAC). Вторые используют теорию психоакустики для поиска тех частей аудиофайла, удаление которых не повлияет (или минимально повлияет) на восприятие человеком.

Общая структурная схема перцептивных кодеков (кодировщика и декодировщика) изображена на рисунке 1.

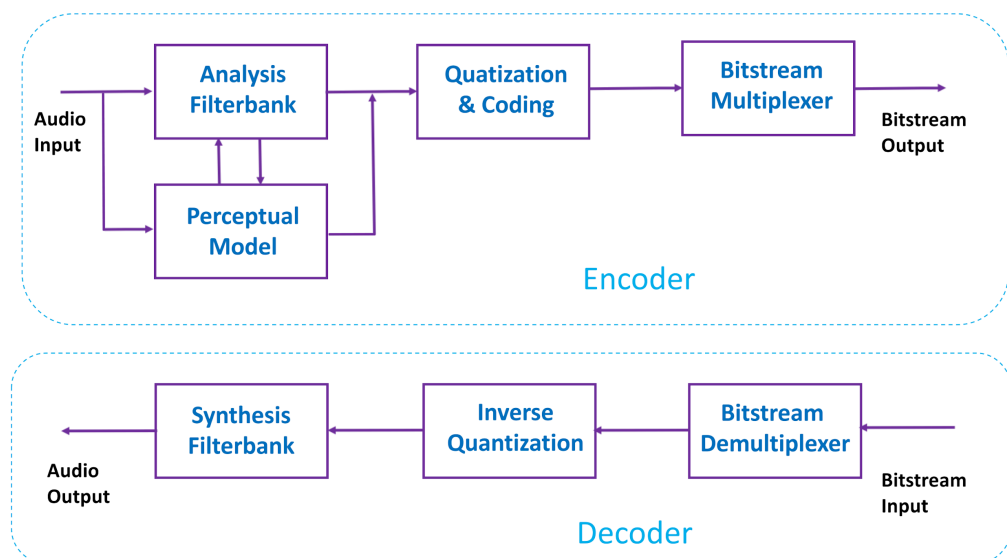


Рисунок 1 – структурная схема перцептивного кодирования

Разберем подробнее каждый блок кодировщика, так как именно кодировщик использует психоакустическую модель (ПАМ) в процессе кодирования.

1. Банк фильтров. Разбивает входной сигнал на частотные диапазоны, сумма которых равна исходному сигналу.
2. Модель восприятия человеком. Именно здесь используется психоакустическая модель и происходит большая часть сжатия.
3. Квантование, энтропийное кодирование и мультиплексирование. Основы цифровой обработки и передачи сигналов, не относятся к обсуждаемой теме. Однако, некоторые кодеки также используют ПАМ при квантовании (например, для избавления от шумов квантования).

3.2 Принципы построения ПАМ аудиокодеков

Хоть разные аудиокодеки и могут использовать разные ПАМ для кодирования аудиофайлов, в них содержатся общие положения и принципы, в основе которых лежат особенности восприятия человеком звука.

Первая особенность – устройство слуховой системы человека. Она состоит из нескольких составляющих, наиболее важные из которых (в данном контексте):

- барабанная перепонка (eardrum): в виде вибраций передает колебания воздуха (звуковые волны) слуховым косточкам;
- слуховые косточки среднего уха (ossicular bones): молоточек, наковальня, стремя — передают механические колебания улитке;
- улитка (cochlear structure): индуцирует бегущие волны по длине базилярной мембраны;
- нервные рецепторы (neural receptors): преобразовывают колебания в химические и электрические сигналы (имеют соединения по всей длине базилярной мембраны).

Основной интерес здесь представляют улитка с базилярной мембраной. Благодаря форме базилярной мембраны (сужается к основанию) и тому, что к разным участкам данной мембраны подсоединяются клетки, отвечающие за

восприятие разных частот, можно утверждать, что ушная улитка — это нелинейная система с частотной избирательностью (рисунок 2).

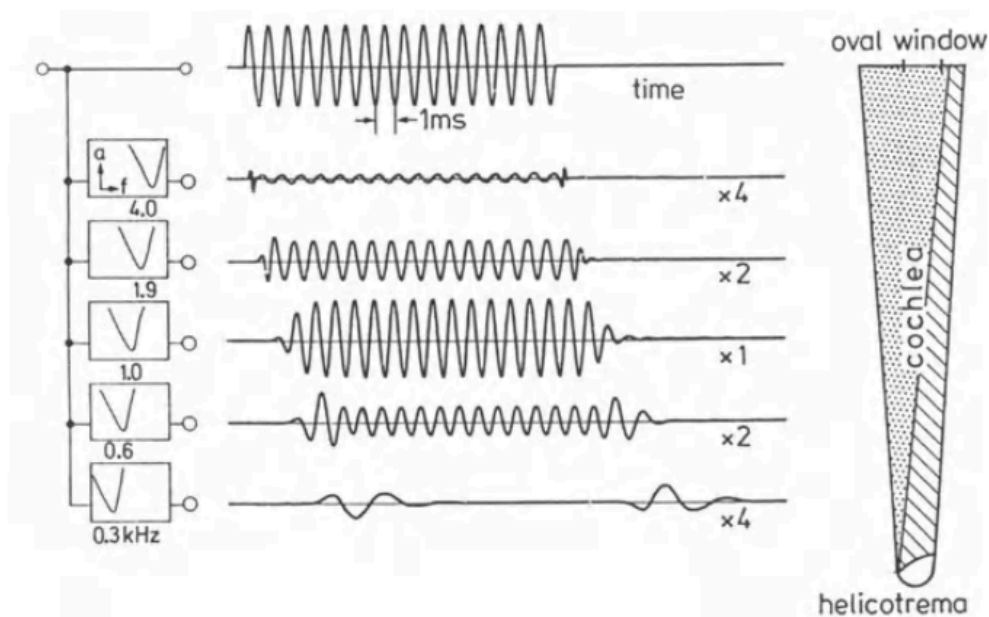


Рисунок 2 – отклики тона в разных местах базилярной мембраны

Это значит, что с точки зрения цифровой обработки сигналов, ушная улитка — это банк полосовых фильтров. При этом фильтры сильно перекрывают друг друга. Именно поэтому целесообразно при перцептивном кодировании разбивать сходной сигнал на частотные диапазоны, а именно — частотные диапазоны, соответствующие диапазонам базилярной мембраны (рисунок 3).

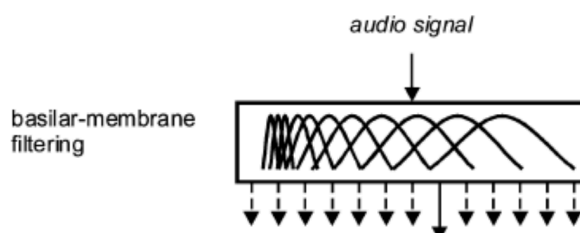


Рисунок 3 – часть схемы модели восприятия, описывающая фильтрацию базилярной мембраны

Отмечено также, что: у частотных групп, на которые базилярной мембраной разбивается аудио-сигнал, фиксированная ширина полосы; ширина полосы частотной группы зависит от средней частоты группы нелинейно.

Вторая особенность восприятия звука человеком состоит в пороге слышимости. Исследования порога слышимости и зависимости его от частоты привели к результатам, изображенным на рисунке 4.

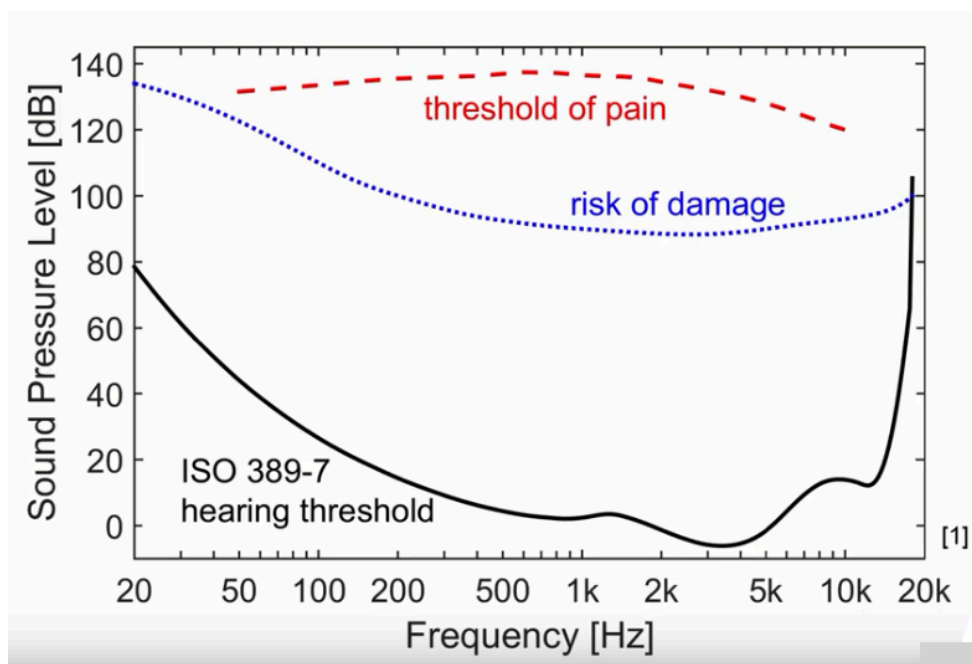


Рисунок 4 – порог слышимости в тишине (threshold in quiet, hearing threshold), уровень риска повреждения органов слуха (risk of damage), уровень болевых ощущений (threshold of pain)

Следовательно, чтобы какой-либо звук мог быть услышан, он должен превысить значение "порога в тишине". Можно увидеть, что лучше всего человек слышит звуки в частотном диапазоне 2-4 кГц. Видно также, что человеку все сложнее услышать какой-либо звук с увеличением или уменьшением частоты.

При этом, при появлении шума, порог слышимости поднимается с увеличением мощности шума. К примеру, чтобы услышать звук на частоте 1кГц при белом шуме в примерно 40дБ, необходимо, чтобы этот звук был на уровне более 60дБ. В случае узкополосных шумов, порог слышимости сильно поднимается на частоте узкополосного шума, и плавно снижается в обе стороны от него. Данный эффект называется эффектом маскирования – более сильные сигналы маскируют собой более слабые в данной частотной области (рисунок 5).

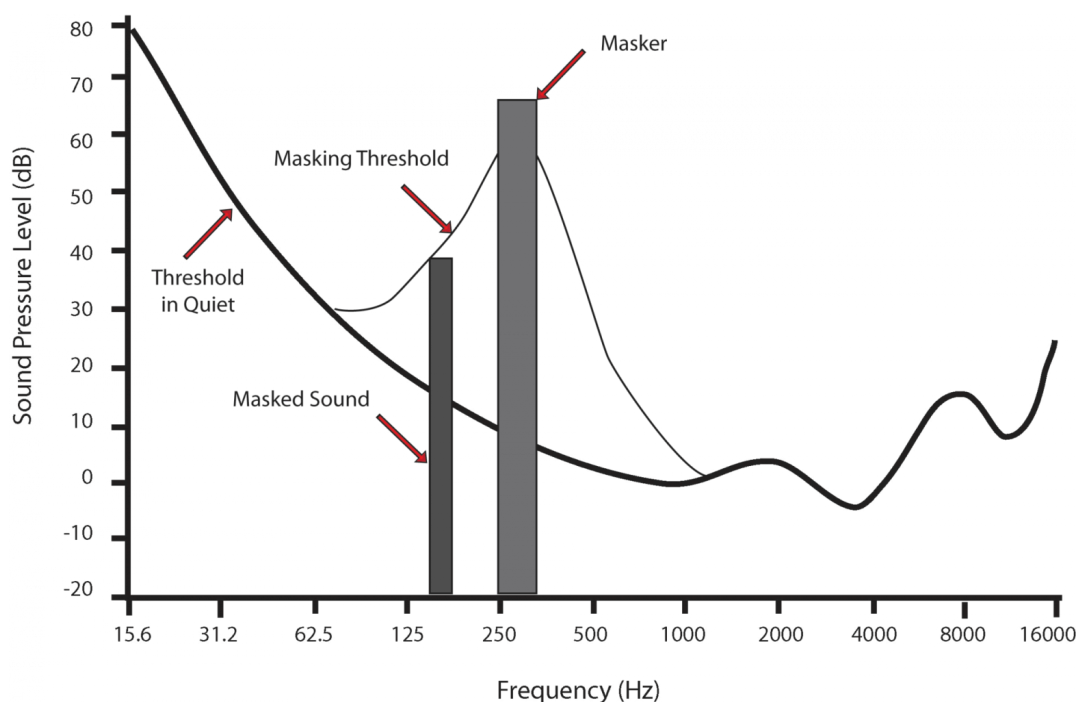


Рисунок 5 – пример маскирования одного тона другим тоном более высокой частоты

Функция, которая определяет порог маскирования, называется функцией распространения (spreading function) и вычисляется на основе эмпирически полученных коэффициентов и шкалы Барков.

Кроме того, существует также временное маскирование: громкий сигнал маскирует собой как следующий за ним более слабый, так и предшествующий ему более слабый сигналы.

Эффект маскирования – как частотного (чаще), так и временного (реже) – основа большинства ПАМ, так как, рассчитав порог маскирования в пределах частотной группы, появляется возможность исключения избыточности с точки зрения модели восприятия путем очистки нерелевантных звуков на этапе квантования.

3.3 Обзор MP3-аудиокодеков

В данном разделе будет проводиться сравнение трех популярных бесплатных аудиокодеков MP3-формата. Будут описаны особенности их психоакустических моделей, субъективного восприятия и объективных характеристик.

В качестве программ для сравнения взяты:

- MP3 Producer от Fraunhofer IIS. Это кодек от Института интегральных схем общества Фраунгофера, которым и разработан формат MP3.
- Lame. Это один из самых популярных кодеков, использующих собственную психоакустическую модель GPSYCHO.
- BladeEnc. Этот кодек является консольной программой, он популярен благодаря своей простоте, скорости кодирования и высокой степени сжатия, однако критикуется за низкое качество звука.

Рассмотрим каждый из кодеков по отдельности. Для оценки субъективного восприятия и объективных характеристик (спектрограммы) возьмем небольшой файл в формате WAV. Спектрограмма этого файла представлена на рисунке 6. Произведена конвертация данного файла вышеописанными кодеками в MP3.

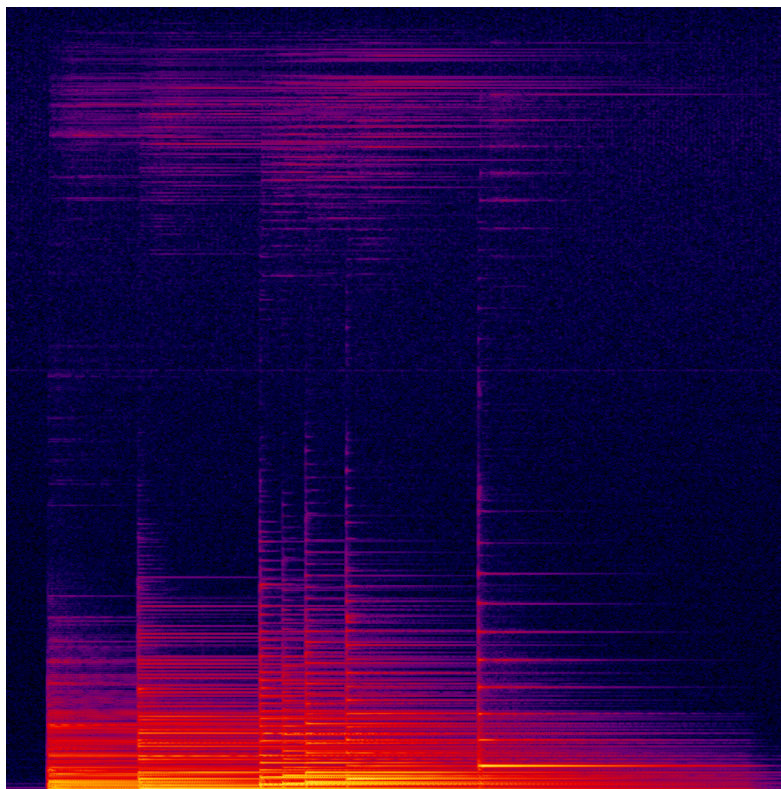


Рисунок 6 – спектрограмма исходного файла

Настройки всех кодеков выставлены одинаково, насколько это позволяют сделать сами программы. Кроме того, специально выбран низкий битрейт (48kbps), так как при таком значении битрейта лучше слышны артефакты сжатия и различия звучания. Выбрано также постоянное значение битрейта (CBR).

Кодек *Lame* в ходе тестов показал самые лучшие результаты, во многом благодаря собственной ПАМ GPSYCHO. Следует выделить качественных алгоритм VBR, использующийся в этой модели, хоть эта особенность и не использовалась в сравнении кодеков.

Кроме этого, данная ПАМ использует технологию mid/side stereo, Технология Mid/Side (M/S) рассматривает стереосигнал не в виде самостоятельных правого и левого каналов, а в виде центрального и бокового элементов. Центральный сигнал (середина) — это монофоническая сумма правого и левого каналов. Здесь отображаются те звуки, которые есть в обоих. Боковой сигнал (сторона) отображает различия между этими каналами и содержит элементы, влияющие на стереофонию и определяющие её. Следовательно, баланс между центральным и боковым сигналами определяет ширину стереокартины. Такой подход позволяет, во-первых, отводить меньше места side-каналу, во-вторых, более разумно кодировать информацию на основе перцептивной модели.

Также, в ПАМ GPSYCHO реализовано эффективное управление предэхо – модель учитывает временное понижение чувствительности уха к сигналу текущего фрейма, вызванное мощным сигналом предыдущего фрейма.

Спектрограмма mp3-файла, созданного кодеком Lame, представлена на рисунке 7.

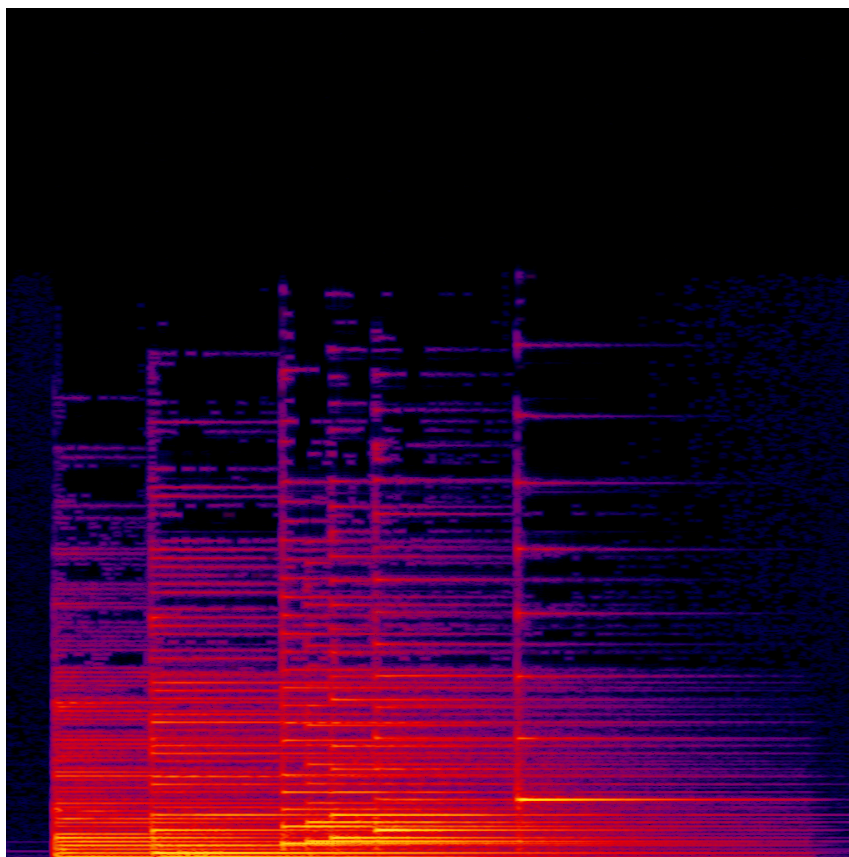


Рисунок 7 – спектрограмма файла, созданного кодеком Lame

Уже видно, что спектрограммы Lame mp3, и оригинального файла отличаются. Тем не менее, субъективно звучание практически не изменилось, только появились еле слышимые высокочастотные шумы квантования. Одна из особенностей кодека Lame состоит именно в качественной работе квантователя. В модели учитывается, что потеря одного бита ведёт к внесению шума квантования величиной порядка 6 dB.

Следующий исследуемый кодек – *MP3 Producer*, разработанный институтом интегральных схем общества Фраунгофера, авторами формата MP3.

Данный кодек использует основные принципы построения психоакустической модели MP3. В данной ПАМ также используется алгоритм абсолютного порога слышимости (ATH), позволяющий отбросить больше избыточных частот, не слышимых человеком. В отличие от GPSYCHO, где биты распределяются на основе различий между порогами

маскирования каналов, в модели, созданной Fraunhofer IIS, биты распределяются на основе мощности сигнала. Для кодировки стерео-файлов в этой ПАМ в основном используется технология Joint Stereo, смешивающая левый и правый каналы для большего сжатия. Может также закодировать левый и правый каналы как отдельные моно-каналы.

Спектрограмма mp3-файла, созданного кодеком от Fraunhofer IIS, представлена на рисунке 8.

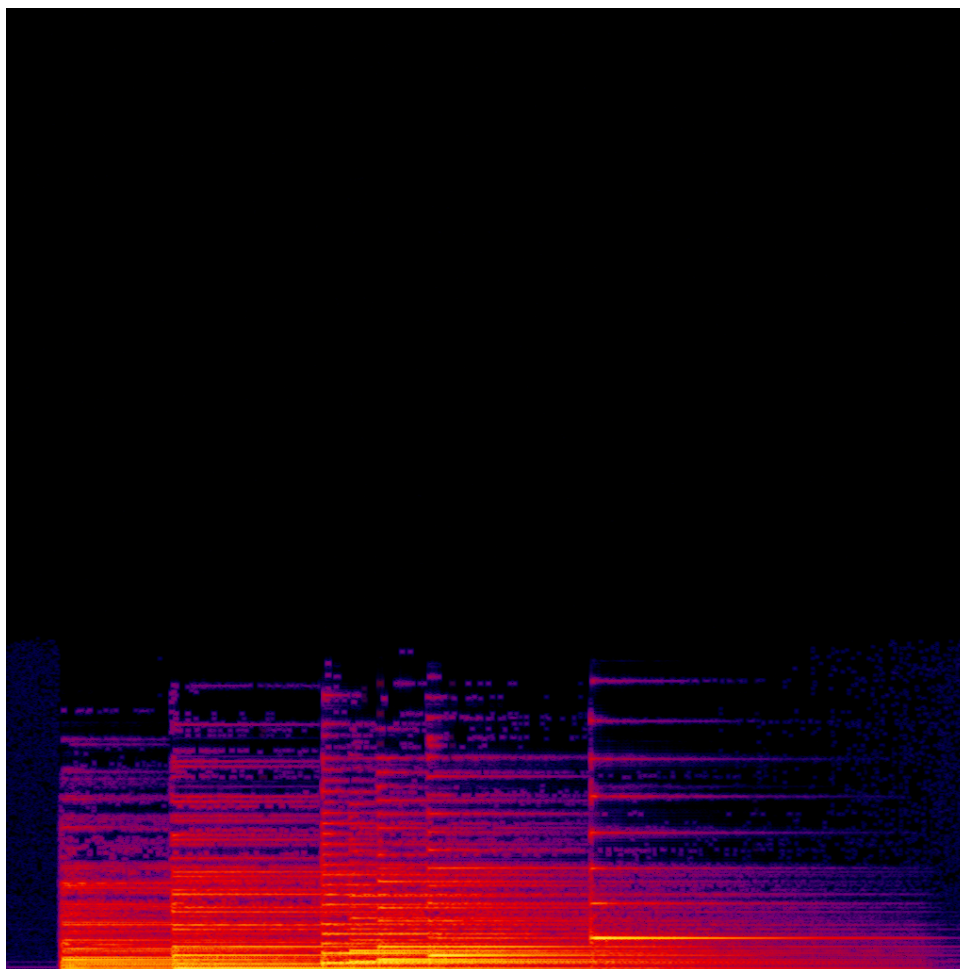


Рисунок 8 – спектрограмма файла, созданного кодеком Fraunhofer IIS (MP3 Producer)

Здесь спектрограмма начинает уже сильно отличаться от изначальной. Заметен резкий срез по частотам, который заметен при прослушивании (появляется «глуховатость» в звуке). Шум стал громче, но, тем не менее, такой файл может быть использован, например, для передачи по сети, либо прослушивания подкастов или других файлов с небольшим спектром частот.

Психоакустическая модель (ПАМ) в *Blade Encoder* основана на демонстрационном варианте ISO-кода и отличается значительной оптимизацией по скорости кодирования. ПАМ ищет громкие звуки, которые могут маскировать тихие звуки; шум, который может влиять на уровень соседних звуков; звуки, которые слишком тихие для восприятия человеком и должны игнорироваться. Информация от ПАМ используется, чтобы определить, каким частям спектра нужно больше битов и, следовательно, кодироваться с большим качеством, а каким частям — меньше битов.

Таким образом, отличительная особенность Blade Encoder отражается именно в алгоритмах скоростного кодирования в ущерб ПАМ. Кроме этого, ISO-код и психоакустическая модель, описанная там, сильно критикуется за неэффективность. Уклон в данном кодеке сделан именно на степень сжатия (mp3-файл, созданный этим кодеком, действительно оказался меньше остальных), а также на скорость кодировки.

Спектрограмма mp3-файла, созданного кодеком от Blade Encoder, представлена на рисунке 9.

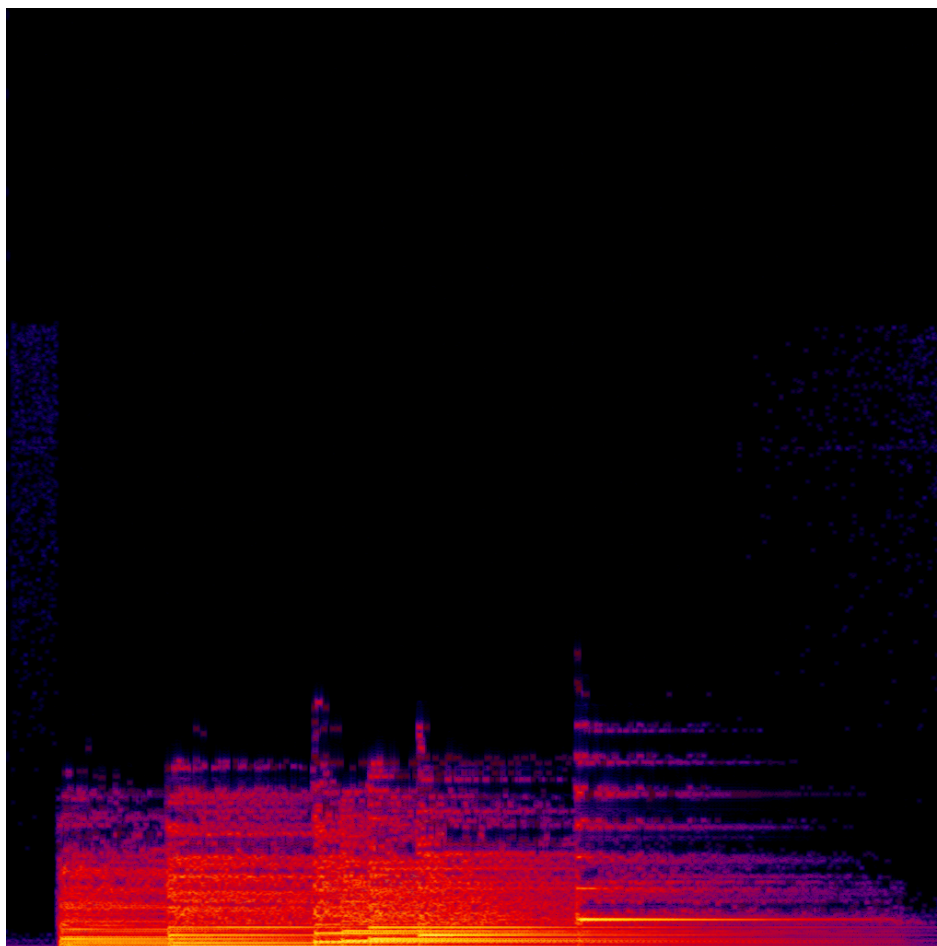


Рисунок 9 – спектрограмма файла, созданного кодеком Blade Encoder

В данном случае уже сильно заметна «квадратность» спектрограммы, она разительно отличается от исходной. В ней представлено еще меньше частот, чем в предыдущей, несмотря на то, что частота дискретизации в этом кодеке неизменяема и равна 48кГц (у остальных кодеков выставлена 44.1кГц).

По субъективному восприятию это также худший из вариантов, шумы становится невозможно игнорировать, а сам сигнал звучит очень глухо.

3.4 Принципы цифрового сжатия. Достоинства и недостатки популярных аудиокодеков.

Проведя тесты нескольких кодеков, можно вернуться к цифровому сжатию в целом. Как было сказано, компрессия аудиоданных представляет собой процесс уменьшения скорости цифрового потока за счет сокращения статистической и психоакустической избыточности ЦЗ сигнала. Методы

сокращения статистической избыточности аудиоданных также называют сжатием без потерь, а, соответственно, методы сокращения психоакустической избыточности – сжатием с потерями. Сокращение статистической 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 преобразования являются: возможность восстанавливать исходный сигнал без искажений. способность обеспечивать наибольшую концентрацию энергии в небольшом числе коэффициентов преобразования.

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 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.

Вывод

В ходе лабораторной работы получены теоретические знания по представлению звуковой информации на электронных носителях в формате MP3, алгоритмах сжатия и кодеках, позволяющих осуществлять сжатие звукового файла в формате MP3.