

32 kbps Audio Codecs Comparison



Spectral Comparison

*Руководитель проекта: Александр Жирков
Замеры, обработка, текст: Валентин Вербовой
Консультант: Дмитрий Ватолин*

**Произведено подробное сравнение шести
представительных кодеков на битрейте 32
kbps несколькими методиками**

September 2003

CS MSU Graphics&Media Lab

Audio Group

<http://www.compression.ru/audio/>

Сравнительное тестирование звуковых кодеков

1 Введение	3
1.1 Цель исследования	3
1.2 Описание файлов, использующихся для тестирования.....	4
2 Тестирование по сонограмме	7
2.1 Методы тестирования	7
2.2 MP3 (Blaze Media Convert).....	8
2.3 MP3 (RazorLame)	9
2.4 MP3pro	11
2.5 WMA.....	12
2.6 MPEG-4 AAC.....	13
2.7 GSM 6.10.....	14
3 Тестирование по спектру.....	18
3.1 Описание задач и методов	18
3.2 mp3 32 kbps (Blaze Media Convert)	19
3.3 mp3 32 kbps (RazorLame).....	20
3.4 mp3pro	21
3.5 WMA.....	22
3.6 AAC	24
3.7 GSM 6.10.....	26
4 Спектральный анализ разницы сигналов.....	27
4.1 Описание методов	27
4.2 mp3 (Blaze Media Convert).....	29
4.3 mp3 (RazorLame).....	30
4.4 mp3pro	31
4.5 WMA.....	32
4.6 MPEG-4 AAC.....	33
4.7 GSM 6.10.....	34
5 Тестирование по числовым характеристикам.....	35
5.1 Описание характеристик, по которым проводилось сравнение	35
5.2 Сравнение числовых характеристик	37
6 Сравнение полученных в результате тестирования данных	38
7 О форме звуковых волн в различных кодеках	40
8 Субъективное сравнение кодеков на слух	41
8.1 Оборудование для тестирования	41
9 Заключение	42
9.1 Результаты прослушивания music.wav	42
9.2 Результаты прослушивания speech.wav	43
9.3 Выводы	44
9.4 Планы	45
9.5 Благодарности	45

1 Введение

1.1 Цель исследования

На сегодняшний день существует множество различных кодеков, используемых для сжатия звукового сигнала с потерями.

В данном обзоре рассматриваются некоторые из наиболее популярных кодеков, а результаты их работы сравниваются по различным аналитическим критериям оценки качества сжатия, а также по субъективному качеству звучания результатов на слух.

Основная задача состояла в сравнении качества работы кодеков на низких битрейтах при кодировании музыки и речи, а также в выявлении особенностей кодирования, отражающихся на качестве звучания.

Вторая задача — показать, насколько соответствуют автоматические метрики результатам оценки качества при прослушивании.

Сравнение на слух по определению субъективно, поскольку основывается на восприятии человеком и даже двух одинаково воспринимающих людей не существует.

Например, рассматривая две картины объективно можно сказать, что на одной больше зеленых красок, чем на другой. Но сказать объективно, что одна лучше другой - нельзя, это можно определить только субъективно, и один человек скажет одно, а другой другое.

Со сравнением качества видео похожая ситуация – можно объективно сказать, что одно видео по среднеквадратичному отклонению ближе к исходному чем другое, но это не будет значить, что каждому человеку покажется, что, то, которое с меньшим отклонением ближе к оригиналу.

Вопрос отличия объективного и субъективному сравнения качества аудио ещё более остр по сравнению с видео, поскольку в нем любые посэмпольные метрики ещё менее адекватны человеческому восприятию.

Поэтому, в большинстве обзоров и статей пользуются метриками основанными на прослушивании музыкальными экспертами. Однако, совокупность различных объективных метрик так же может дать нам более-менее реалистичную оценку отличий одного кодека от другого и что существенно – автоматическую. Автоматические методы сравнения позволяют существенно ускорить разработку новых кодеков, в особенности для предотвращения появления ошибок в коде влияющих на воспроизведение.

Целью данного исследования как раз и является также осуществление многостороннего анализа композиции различными аналитическими методами для возможности получения оценки сравнимой с получаемой при её перцептивном восприятии.

Для тестирования нами были взяты следующие кодеки:

- Mp3 кодек из пакета Blaze Media Convert 1.4
- Mp3 кодек Lame 3.93 MMX (RazorLame V 1.1.5.1342)
- mp3pro Fraunhofer (CoolEdit 2.0 mp3pro audio coding technology licensed from Coding Technologies, Fraunhofer IIS and Thomson multimedia)

- WMA кодек из пакета Blaze Media Convert 1.4
- MPEG-4 AAC (производитель неизвестен)
- GSM 6.10.

Настройки всех кодеков брались по умолчанию.

Кодеки Mp3, Mp3pro, WMA, AAC принадлежат к классу кодеков, предназначенных для кодирования музыки, GSM – это стандартный формат для представления речи.

1.2 Описание файлов, использующихся для тестирования

Основная задача данного 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.

Было использовано несколько звуковых файлов, но изображения приведены, в основном, для фрагмента музыки с голосом солистки, т.к. наиболее часто возникающая задача — это кодирование именно музыки с вокалом.

Следующие изображения, характеризующие тестируемые файлы представляют собой сонограммы, т.е. характеристику распределения звука в частотно-временном пространстве. Чем громче звук в данной точке (по времени и по частоте) на сонограмме, тем больше ее значение в этой точке. Цвета переходят от белого к синему, потом к зеленому, красному и желтому, по мере увеличения значений.

Файлы, используемые для тестирования:

1. music.wav, в файле первую секунду в основном присутствует голос солиста, а потом голос и музыка. Длительность 9 секунд. (Рис 1.1)

Это основной файл в тестировании.

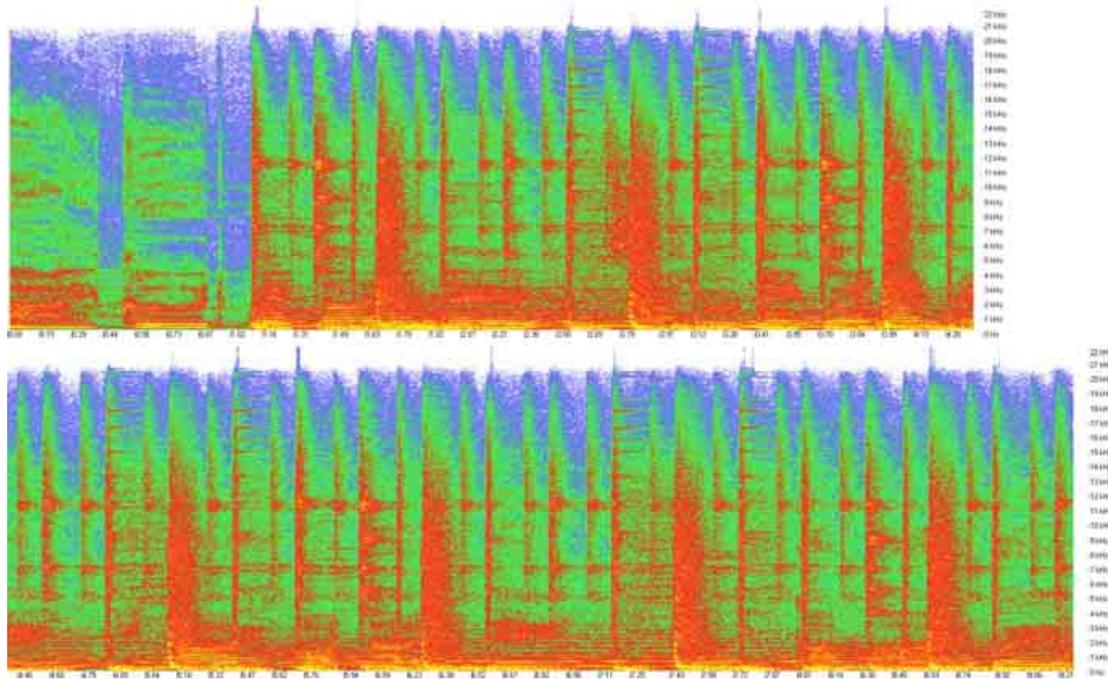


Рисунок 1.1 Спектрограмма тестового файла с музыкальной композицией

2. naturenoises.wav, файл содержит запись крика чаек (первая треть файла), вторая треть — шум воды, третья — щебетание птиц, разделенные тишиной. Длительность 10 секунд. (Рис 1.2)

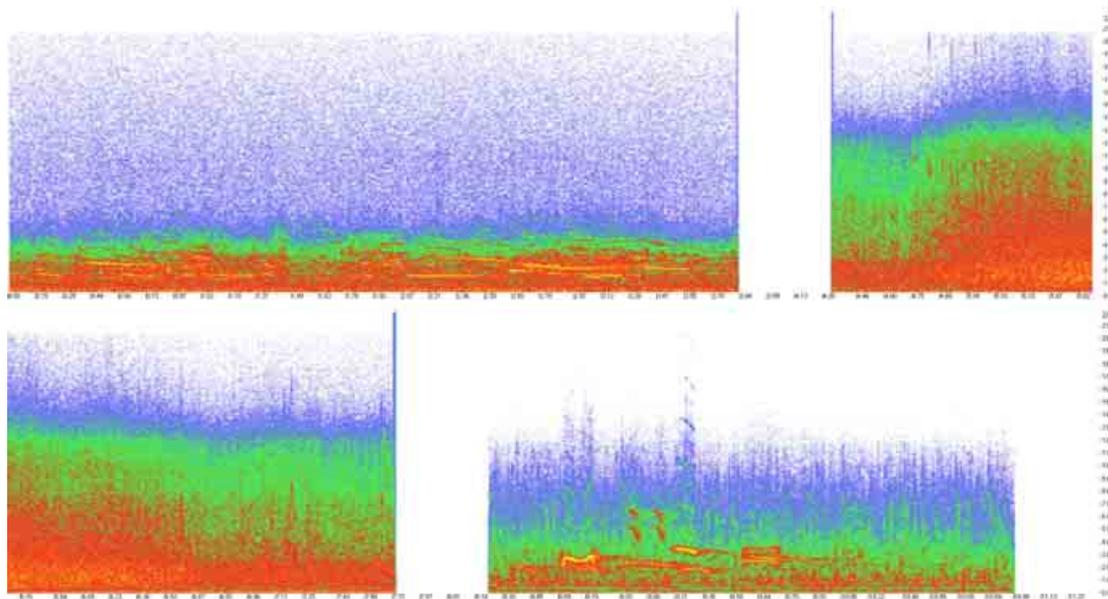


Рисунок 1.2 Спектрограмма тестового файла звуков природы

3. test.wav, в файле записаны синтетические звуки. На данном файле можно четко определить, насколько качественно кодек работает в различных частных областях и с различными сигналами. Синтетический сигнал достаточно прост, поэтому на нем гораздо проще определить проблемные участки для каждого тестируемого кодека, что достаточно тяжело сделать, анализируя сложную музыкальную композицию. Сигнал состоит из гармонических составляющих и коричневого шума.

Данный файл особенно пригодился при тестировании кодека GSM. Длительность 15 секунд. (Рис 1.3)

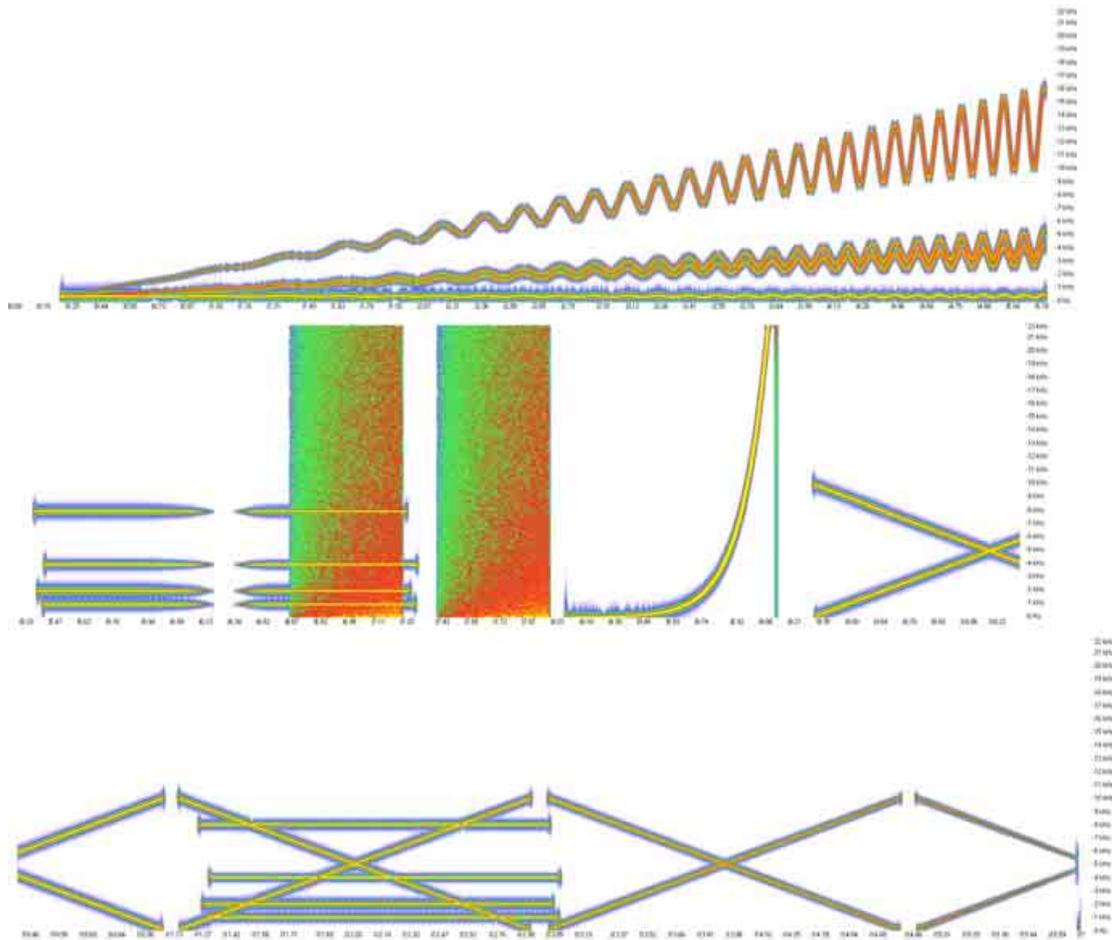


Рисунок 1.3 Спектрограмма синтетического файла

4. **speech.wav**, содержит образец речи диктора с радио.
Длительность 16 секунд. (Рис 1.4)

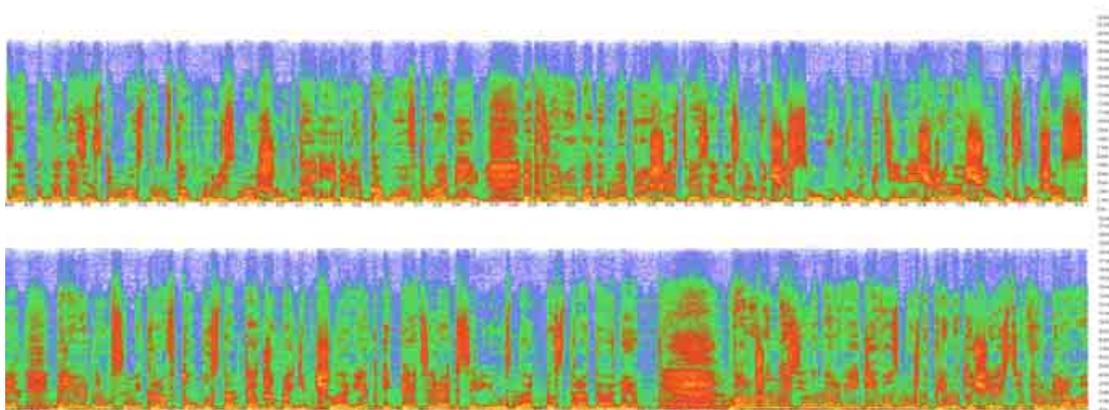


Рисунок 1.4 Спектрограмма файла с образцом речи диктора радио

5. **instrvoice.wav**, файл, содержащий фрагмент еще одной музыкальной композиции. Так же как и первая композиция, это сольная партия певицы с музыкальным аккомпанементом. В отличие от первого файла, в данном фрагменте голос солистки относительно музыкальных инструментов звучит значительно громче. Длительность 24 секунды. (Рис 1.5)

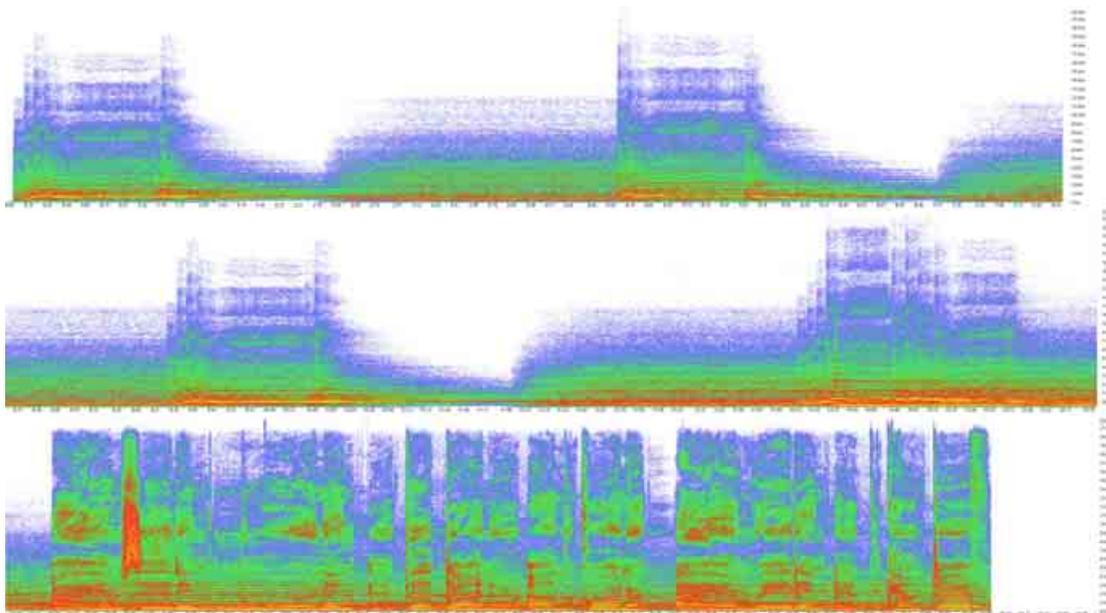


Рисунок 1.5 Спектрограмма музыкальной композиции (2)

2 Тестирование по сонограмме

2.1 Методы тестирования

Методика тестирования кодеков с потоком данных 32 кбит/с. по сонограмме.

Описываемый ниже метод сравнения сигналов проводился «на глаз» по полученным сонограммам. 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 по модулю оказывались получаемые значения, тем больше отличается 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, 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 изображения 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 $A_n[i][j]$ получается из дискретного преобразования Фурье от исходного сигнала следующим образом:

$$A_n[i][j] = \log\left(\sqrt{FFT[j].Im^2 + FFT[j].Re^2}\right), i \in [0, T],$$
 i -время. FFT-вектор, полученный в результате обработки сигнала быстрым преобразованием Фурье.

Данный метод учитывает только амплитуду сигнала и поэтому пригоден для тестирования кодеков не сохраняющих форму волны.

Перейдем непосредственно к тестированию.

2.2 MP3 (Blaze Media Convert)

Первое, что можно отметить при изучении результата работы кодека этот кодек — это то, что при кодировании он добавил 2212 сэмпла (эквивалентно 0,05 секунды) тишины в начале файла, что определил специально разработанный для этих целей детектор сдвига начала файла. Поэтому, с оригиналом сравнивался поправленный файл с подкорректированным смещением звуковой волны.

На следующем изображении четко видно, что с частотами выше 10 kHz кодек обошелся очень грубо, хотя, формально, они присутствуют (рис 2.2.1).

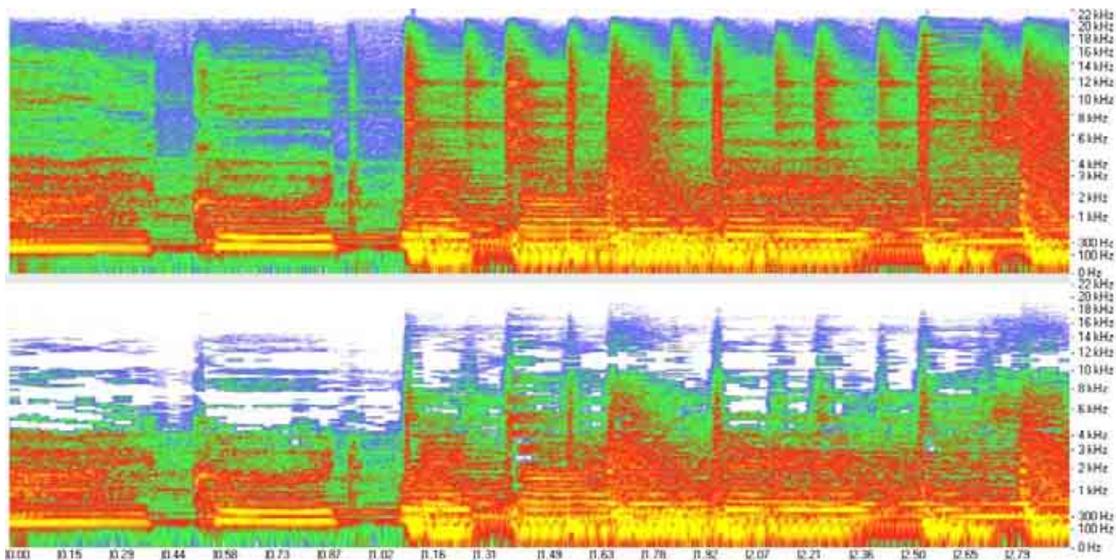


Рисунок 2.2.1 Верхняя спектрограмма – оригинал, нижняя – после BMC mp3

На сонограмме разницы (рис 2.2.2) сигналов можно увидеть, что изменения частотной картины произошли не только в высокочастотном диапазоне. Наличие красных и зеленых областей в диапазоне до 6 kHz свиде-

тсльствует о серьезных искажениях, добавленных кодеком в исходный сигнал, что, в свою очередь, отрицательно сказывается на качестве звучания.

Искажения в этой области особенно существенны, потому что человек наиболее восприимчив к звукам в среднечастотном звуковом диапазоне и этот диапазон обычно наиболее качественно воспроизводят колонки.

Интересен тот факт, что кодек неодинаково отнесся к различным частотным полосам выше 6 kHz. Так, например, на рис 2.2.2 видно, что на частотах в районе 7 и 12 kHz звук претерпел большие потери, нежели в остальном частотном диапазоне. Но в целом, начиная с 5-ти kHz и выше, энергия звука ослабилась на 10-30 дБ, и рассчитывать на то, что оставшиеся высокие частоты будут вносить заметный вклад в общее звучание, не приходится, так как в восстановленном сигнале их почти не будет слышно.

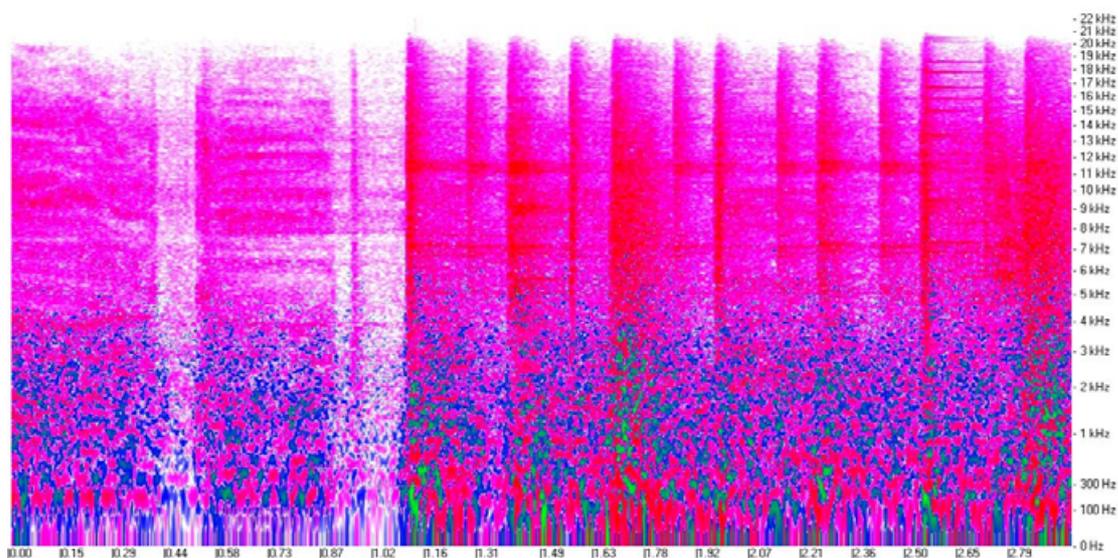


Рисунок 2.2.2 Разница спектрограмм исходного и обработанного mp3 BMC звуков

2.3 MP3 (RazorLame)

Сдвиг сигнала относительно исходного после цикла кодирования/декодирования составил 2 сэмпла.

На следующем изображении (рис 2.3.1) четко видно, что произошел срез частот на уровне 8 кГц.

Известно, что Lame, в отличие от BMC (Blaze Media Convert), уменьшает частоту дискретизации либо просто применяет фильтр низких частот до уровня 8 kHz для того, чтобы после этого более качественно закодировать только частоты ниже данного диапазона. В результате, как и ожидалось, в части 0-7 kHz, спектр изменился гораздо меньше, чем после обработки BMC.

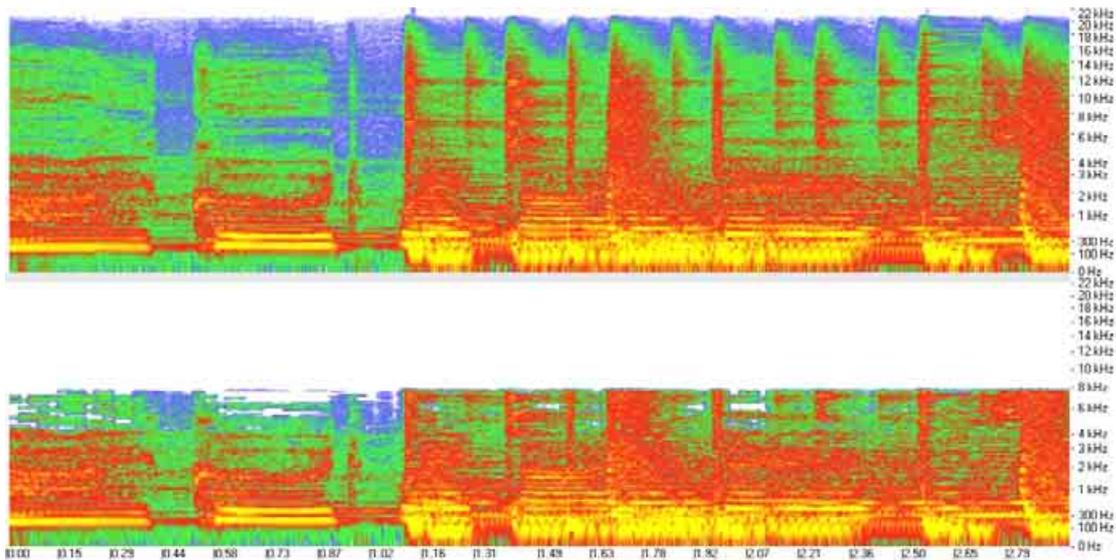


Рисунок 2.3.1 Верхняя спектрограмма – оригинал, нижняя – после Lame mp3

На рис 2.3.2 видно, что зеленых и красных областей меньше, чем у VMS и они меньше по площади — это значит, что звуки, по частоте находящиеся ниже 7,5 kHz, Lame изменил намного слабее, чем VMS. В частотном диапазоне до 7 kHz, в основном, лежит речь (это хорошо видно на рис 1.4), поэтому можно предположить, что Lame на 32 kbps больше подходит для кодирования речи, нежели музыки.

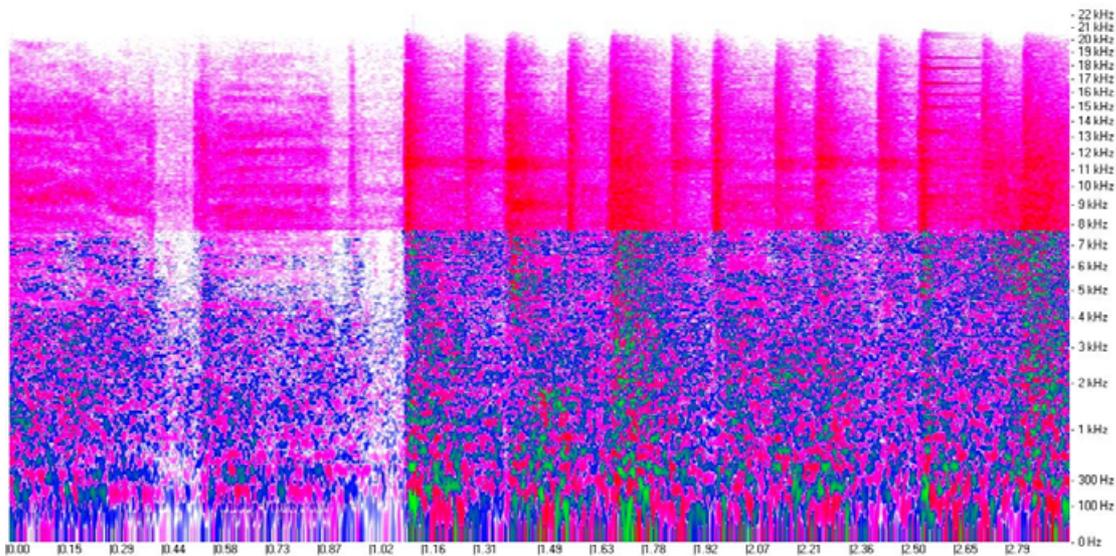


Рисунок 2.3.2 Разница спектрограмм исходного и обработанного mp3 lame звуков

Конечно, сказать точно, что лучше — полный частотный диапазон с большими искажениями или урезанный с маленькими, можно на слух, т.к. неизвестно что больше понравится среднему слушателю.

2.4 MP3pro

Сдвиг звука в обработанных файлах составил 3541 сэмпла, что эквивалентно 0,08 секунды тишины в начале файла.

На рис 2.4.1 видно, что срез частоты произошел, на уровне 16 kHz, что существенно лучше по сравнению с остальными кодеками, к тому же, считается, что более высокие частоты сохранять не обязательно, поскольку большинство людей их просто не слышит.

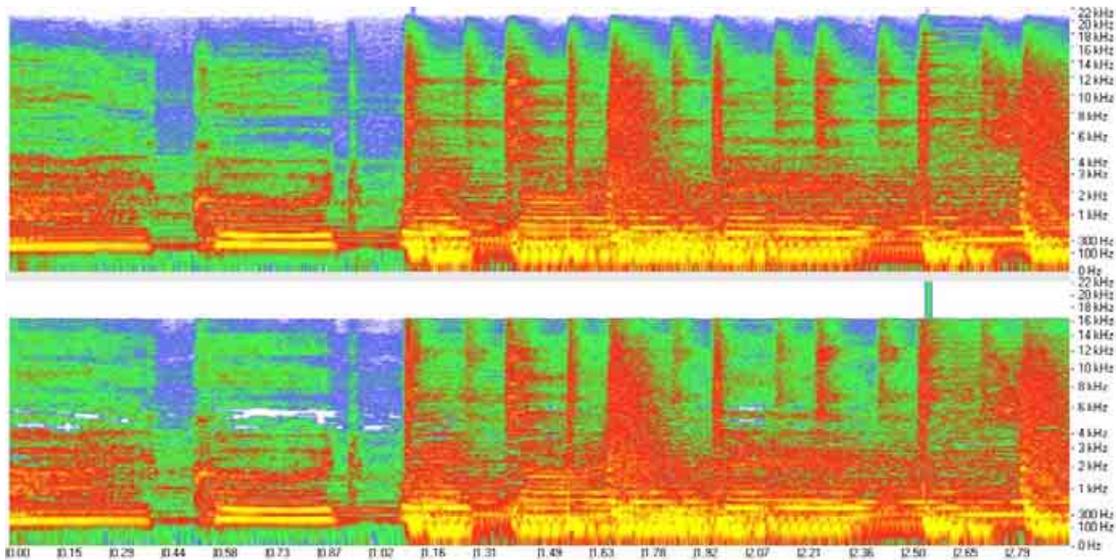


Рисунок 2.4.1 Верхняя спектрограмма – оригинал, нижняя – после mp3pro

Изучая рис 2.4.2, можно заметить, что в диапазоне ниже 6 kHz отличия от исходного сигнала немного больше, чем у Lame, а в диапазоне до 3 kHz результат работы mp3pro даже иногда уступает VMS.

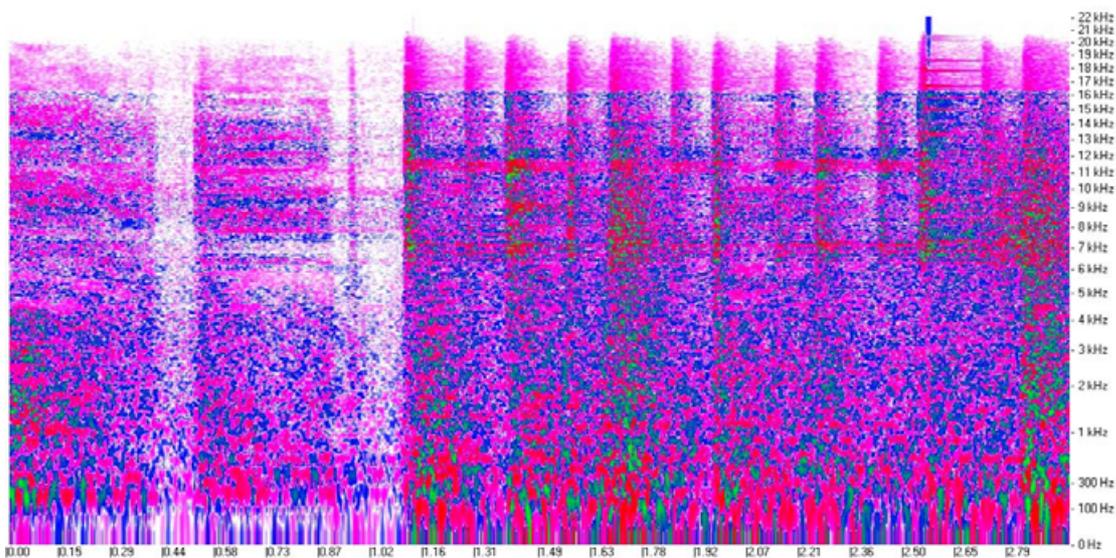


Рисунок 2.4.2 Разница спектрограмм исходного и обработанного mp3pro звуков

В частотном диапазоне 6-16 kHz с кодированием частот mp3pro справился намного лучше, чем оба вышеприведенных кодека, т.к. Lame просто отфильтровал звук выше 8 kHz, а VMS хоть и кодировал этот диапазон, но справился с этим крайне плохо. Стоит отметить, что MP3-Pro применяет хитрость и не хранит форму волны на высоких частотах явным образом, а хранит только их энергетическую огибающую.

Необходимо учитывать, что степень искажений во многом зависит от интенсивности исходного звука. Поскольку в данном случае в низкочастотной области звук громче, чем в высокочастотной области, то отличия в сонограммах там могут быть сильнее, но это будет меньше заметно.

2.5 WMA

На рисунке 2.5.1 видно, что срез частотного диапазона произошел в районе 15 kHz. Правда, на частотах 12-15 сохранены лишь некоторые частотные всплески.

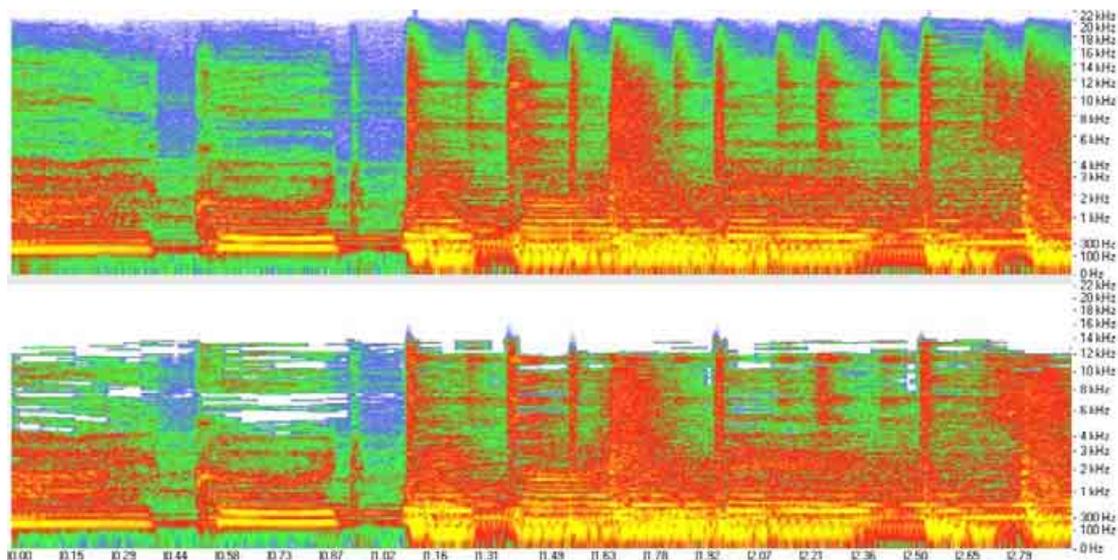


Рисунок 2.5.1 Верхняя спектрограмма – оригинал, нижняя – после WMA

Сравнивая результаты работы WMA (рис 2.4.2) и mp3pro (рис 2.2.2), видно, что WMA сильно выигрывает у mp3pro в диапазоне 6-12 kHz, т.к. количество искажений в этой частотной области у него значительно меньше, но при этом проигрывает в области 12-16 kHz, т.к. фактически не сохраняет звуки с этими частотами. В области около 7 kHz нет роста искажений, наблюдающегося у mp3pro, что свидетельствует о том, что в среднечастотной области величина искажений мало зависит от амплитуды исходного сигнала. Интересная ситуация наблюдается в целом по диапазону до 12 kHz. Все предыдущие кодеки немного подрезали почти все кодируемые частотные составляющие звука, тем самым, добавив искажения. Соответственно на сонограммах разницы появлялось большее количество красных областей. В результате WMA наоборот преобладают зеленые области, показывающие, что после кодирования сигнала в нем добавились некоторые звуки, особенно заметные в области до 1 kHz. Однако в целом, нельзя сказать, что искажений в низкочастотной области продукции данного кодека больше, чем в описанных ранее. Хорошо видно, что только Lame справился с этим частотным диапазоном немного лучше.

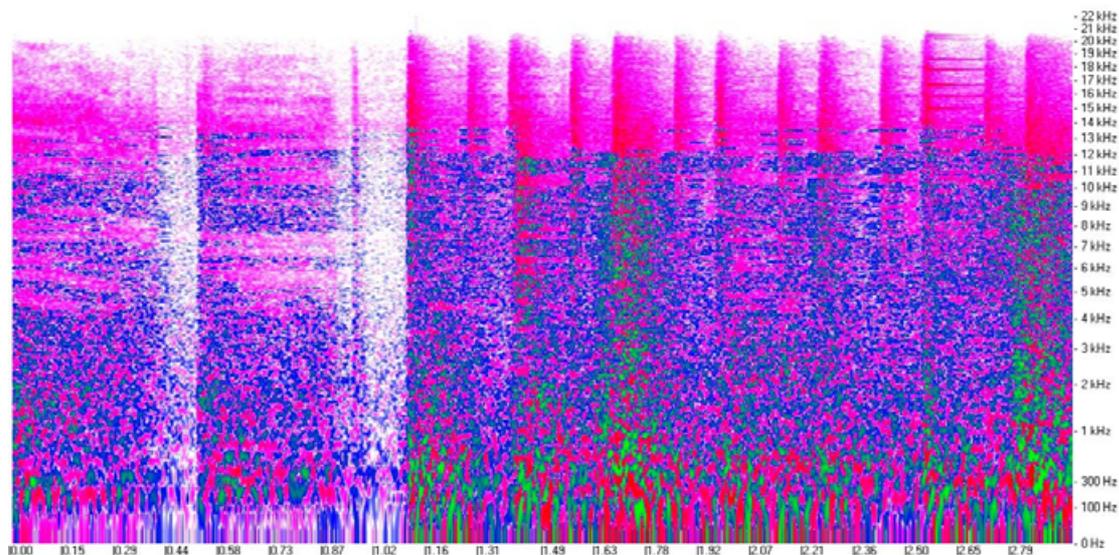


Рисунок 2.5.2 Разница спектрограмм исходного и обработанного WMA звуков
2.6 MPEG-4 AAC

При кодировании звук был срезан на отметке 10 kHz, т.е. частот сохранено больше чем у mp3, но меньше, чем у mp3pro и WMA. (рис 2.6.1).

Четко видно, что кодек усилил некоторые звуки, которые в исходном сигнале были значительно слабее. Также наблюдается искажение отдельных мелких деталей звука. Так, к примеру, слабый звон, присутствующий в исходном файле на 1.31, 1.91, 2.15 секунде, был сильно размыт во времени, а местами просто исчез. Следует заметить, что размывание резких атак, подобных ударам литавр и звону во времени отрицательно сказывается на восприятии этих звуков.

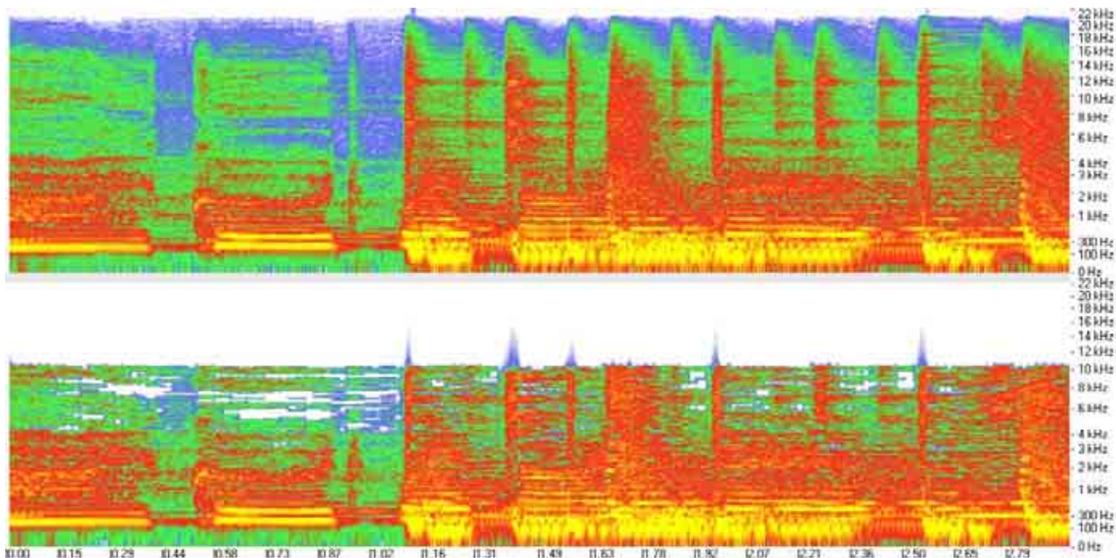


Рисунок 2.6.1 Верхняя спектрограмма – оригинал, нижняя – после аас

По сонограмме разницы (рис 2.6.2) видно, что AAC, как и WMA, усиливал многие закодированные звуки, но делал это в гораздо большей степени, чем WMA. Искажения в звуке в области 1-10 kHz наиболее велики среди вышеописанных звуковых кодеков. Причем области, где амплитуда отличий велика, имеют большую площадь, что делает эти искажения гораздо более заметными.

В области от 300 Hz до 1 kHz кодек выделяется хорошими результатами. Вероятно, это связано с использованием психоакустической модели, построенной по экспериментально полученным данным, которые свидетельствуют о том, что человек наиболее чувствителен к изменениям в звуке именно в области 300-1000 Hz.

В низкочастотной области (до 300 Hz) ситуация немного лучше, чем в диапазоне 1-10 kHz.

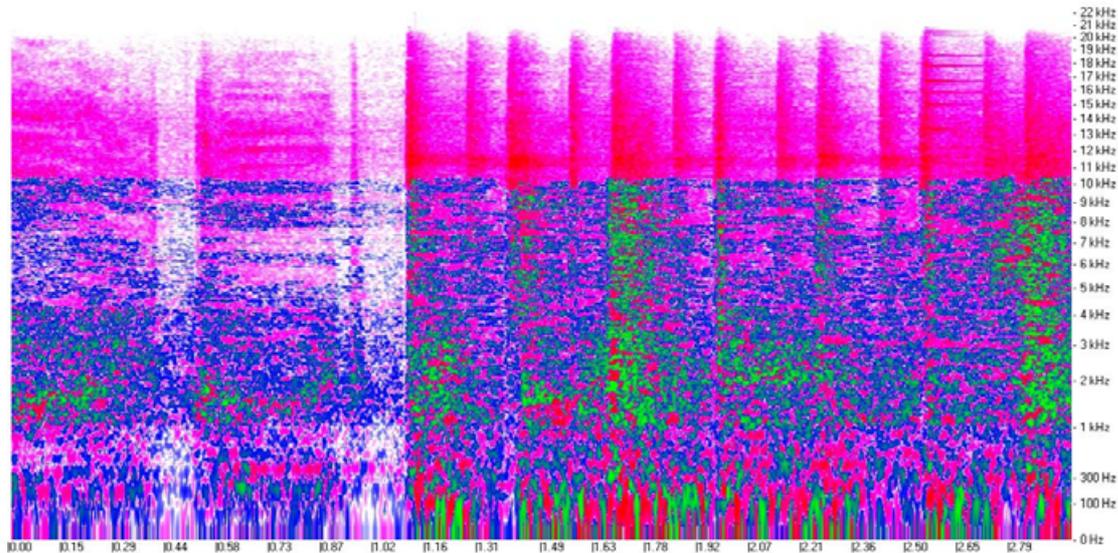


Рисунок 2.6.2 Разница спектрограмм исходного и обработанного AAC звуков

2.7 GSM 6.10

Результаты тестирования показаны для файлов `music_gsm_32.wav` и `music_22_gsm_32.wav`, частота дискретизации которых 44100 Hz и 22050 Hz соответственно.

Данный кодек разрабатывался для сжатия речи.

При беглом взгляде на рис 2.7.1, где представлены сигналы, взятые из файлов `music.wav` и `music_gsm_32.wav`, сразу заметно, что нет среза на высоких частотах. Более того, оказалось, что высоких частот стало заметно больше, чем было до этого.

Частота дискретизации в 32 kbps потоке в GSM 6.10 всего 22 kHz, то есть, таких частот, как 18 kHz, там теоретически не может присутствовать.

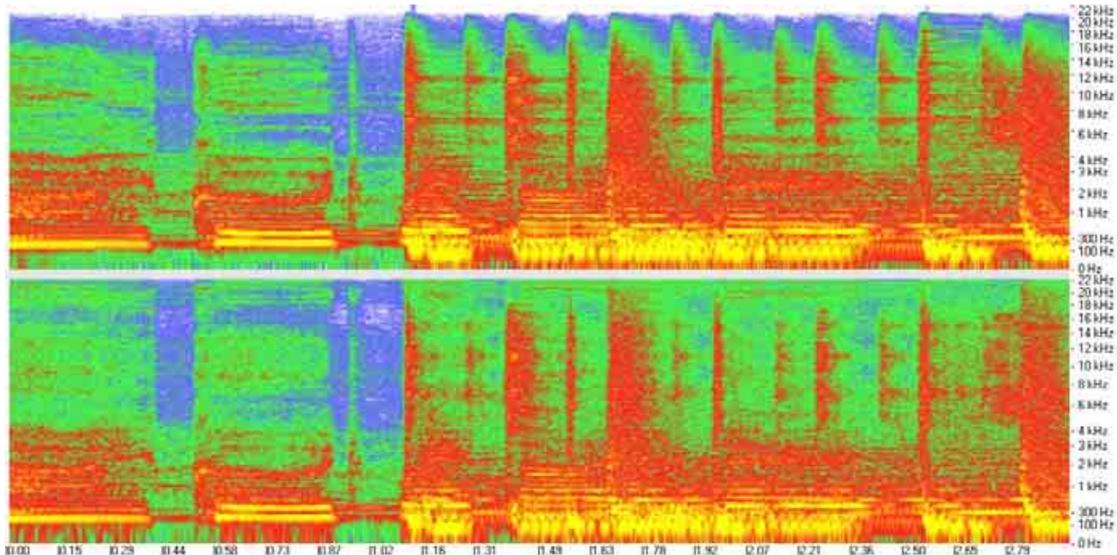


Рисунок 2.7.1 Верхняя спектрограмма – оригинал, нижняя – после gsm 6.10

Ситуацию разъясняет рис 2.7.2, на котором представлен фрагмент исходного и восстановленного файла test.wav.

По всей видимости, при кодировании кодек отражает верхнюю половину частотного диапазона в нижнюю, после чего складывает с оригинальной нижней половиной спектра. При декодировании кодек восстанавливает сначала то, что было закодировано в нижней части спектра, а потом отражает в верхнюю половину с уменьшением интенсивности сигнала по мере приближения к частоте 22 kHz.

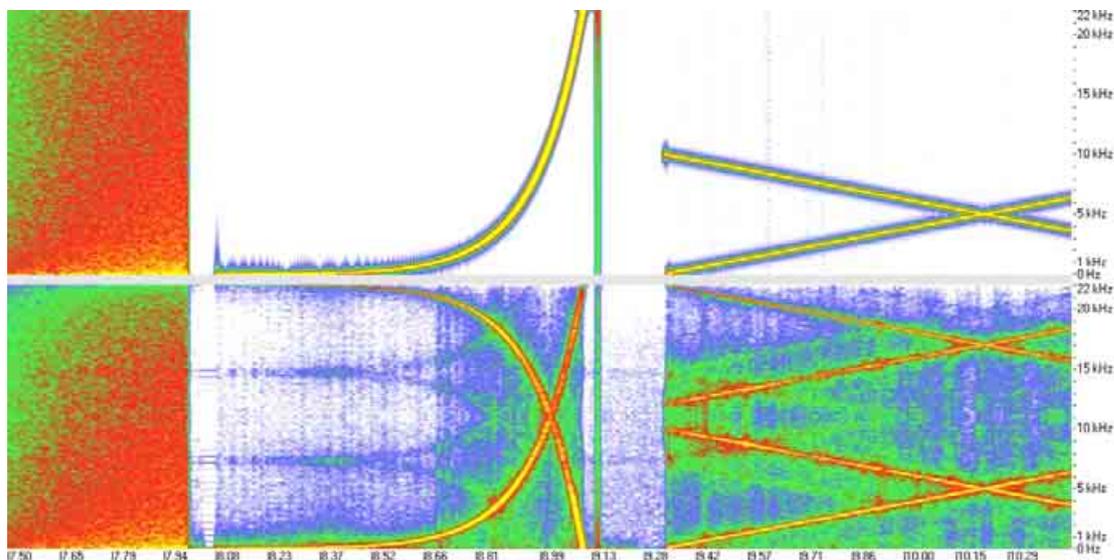


Рисунок 2.7.2 Верхняя спектрограмма – синтетический файл, нижняя – после gsm 6.10 частота дискретизации 44100 Hz

Если же распаковать сигнал с частотой дискретизации 22050 Hz (в прошлом тесте сигнал сразу восстанавливался до 44.1 kHz), а потом вручную привести его к 44.1 kHz, то получится только нижняя половина спектра (рис 2.7.3).

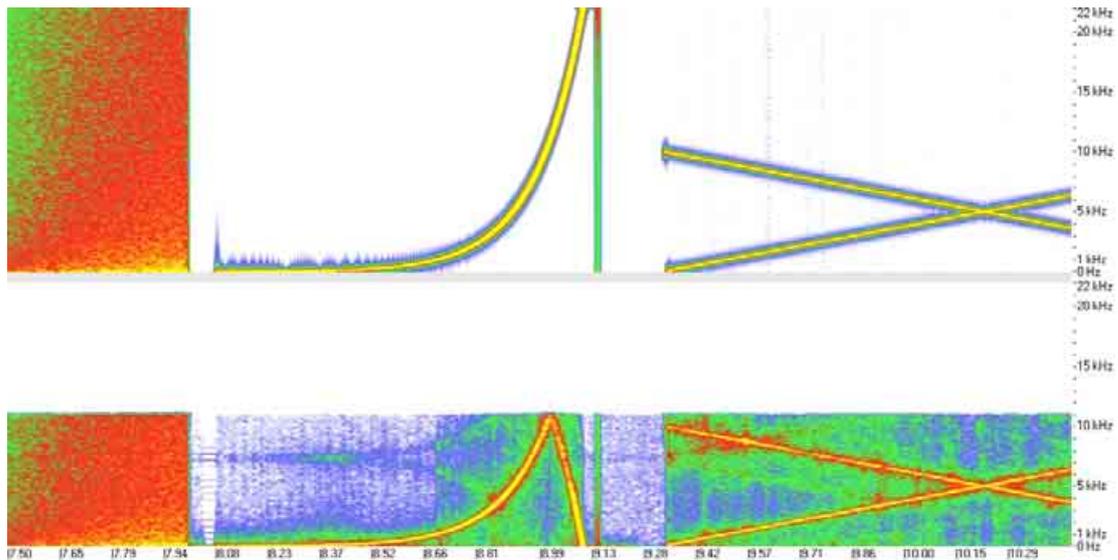


Рисунок 2.7.3 Верхняя спектрограмма – синтетический файл, нижняя – после gsm 6.10 частота дискретизации 22050 Hz расширенный до 44100 Hz

Видно, что в области до 12 kHz результат не стал ни хуже, ни лучше. Все то же отражение верхних частот в нижние, которое, несомненно, сильно портит исходный сигнал.

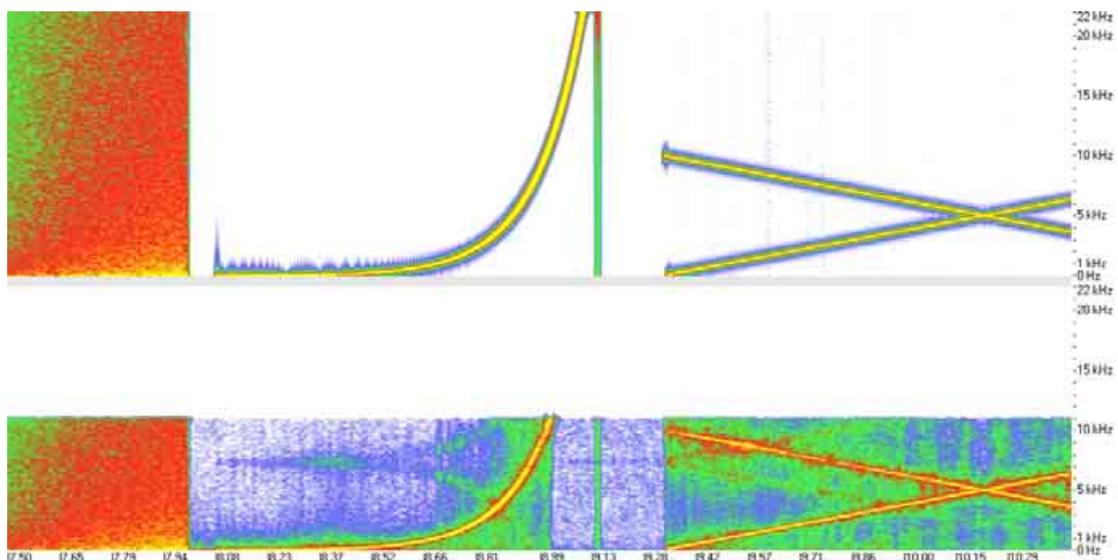


Рисунок 2.7.4 Верхняя спектрограмма – синтетический файл, нижняя – после gsm 6.10 частота дискретизации 22050 Hz на входе GSM кодека

На рис 2.7.4 показан результат работы кодека, когда ему на вход был подан файл с частотой дискретизации 22050 Hz. Результат кодирования частотного диапазона до 11 kHz от этого не улучшился.

Таким образом, установлено, что кодек на данном битрейте не подходит для кодирования исходных сигналов с частотой дискретизации 44100 Hz, по этому для сравнения сигнала был взят исходный файл music.wav и файл music_22_gsm_32.wav. Этот файл был получен путем преобразования music.wav к частоте дискретизации 22050 Hz, последующим кодированием/декодированием кодеком GSM 6.10 с битрейтом 32 kbps и восстановлением файла до частоты 44100 Hz.

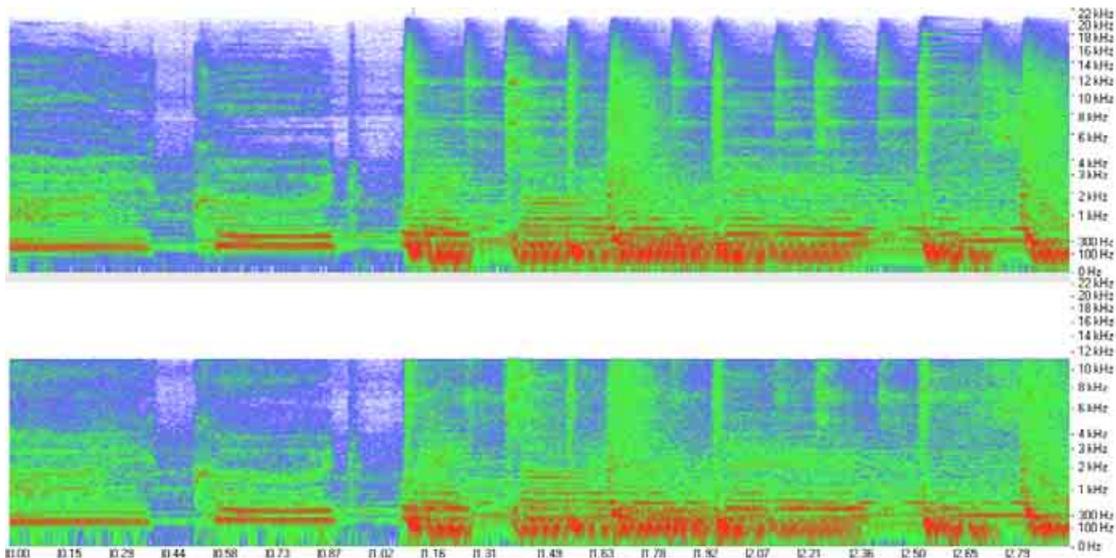


Рисунок 2.7.5 Верхняя спектрограмма – оригинал, нижняя – после gsm с частотой дискретизации 22050 Hz

На рис 2.7.5, где представлены файлы music.wav и music_22_gsm_32.wav, виден срез в районе 11 kHz, но, как было написано выше, это произошло из-за того, что частота дискретизации файла поданного на вход кодека была ровно 22050 Hz. Зато в целом, частотно-временная картина в области до 11 kHz изменилась не сильно.

На разнице сонограмм, представленной на рис 2.7.6, хорошо заметно, что ослаблены низкие частоты в районе 100-300 Hz, но усилены частоты до 50 Hz, которые совершенно не нужны в данном случае.

Т.к. кодек создан для сжатия речи, то в первую очередь он должен хорошо сохранять частотный диапазон 1-4 kHz, что подтверждается изображениями. В частотной полосе 4-10 kHz искажение сильнее, чем у большинства протестированных кодеков, за исключением AAC и mp3 (VBR).

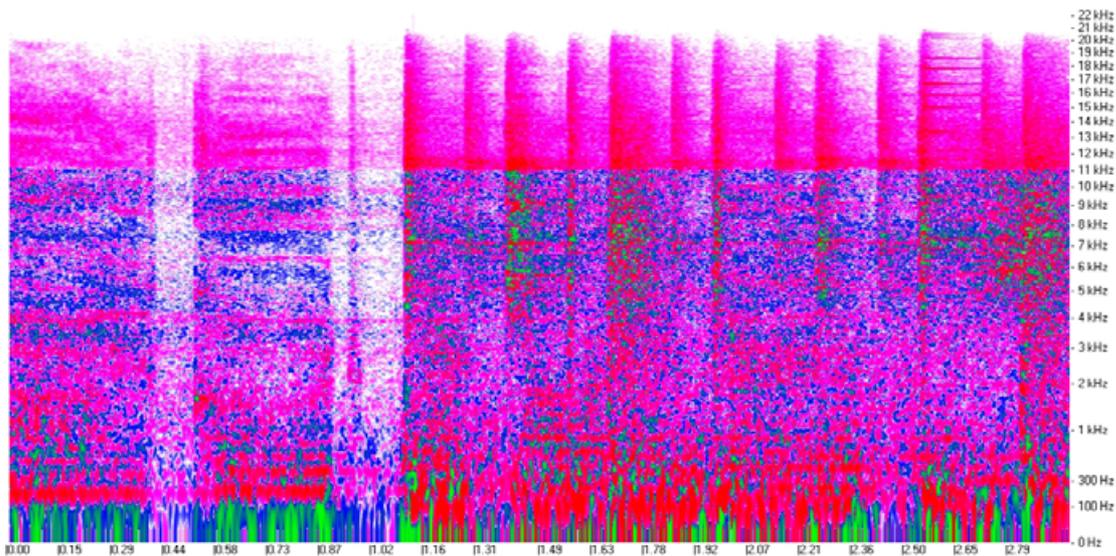


Рисунок 2.7.6 Разница спектрограмм исходного и обработанного GSM звуков

3 Тестирование по спектру

3.1 Описание задач и методов

В этой части рассмотрим различия в полных спектрах закодированных файлов. Кроме того, попытаемся обобщить результаты, полученные при исследовании сонограмм и при замере среднего значения спектра для всего файла по логарифмической шкале.

Объединение области зеленого цвета с областью синего цвета дает спектр исходного сигнала, а объединение областей красного и синего цвета дает спектр сигнала, подвергшегося циклу кодирования/декодирования. В следующих изображениях по вертикальной оси отложена амплитуда сигнала в децибелах, а по горизонтальной - частота. Спектр строился по всему входному файлу.

3.2 mp3 32 kbps (Blaze Media Convert)

На рис 3.1.1 видно, что начиная с 3 kHz и выше, сигнал после обработки ВМС начинает ослабевать по мере увеличения частоты и на отметке 4 kHz падение мощности составляет уже порядка 5 дБ. Напомним, что различие громкости в 1.5 дБ уже заметно на слух.

Таким образом, суммируя полученные данные об изменении частот в спектре в целом и по сонограмме (в пункте 2.2), можно считать, что сигнал, обработанный ВМС, сохранил энергию исходного сигнала только в диапазоне до 3 kHz.

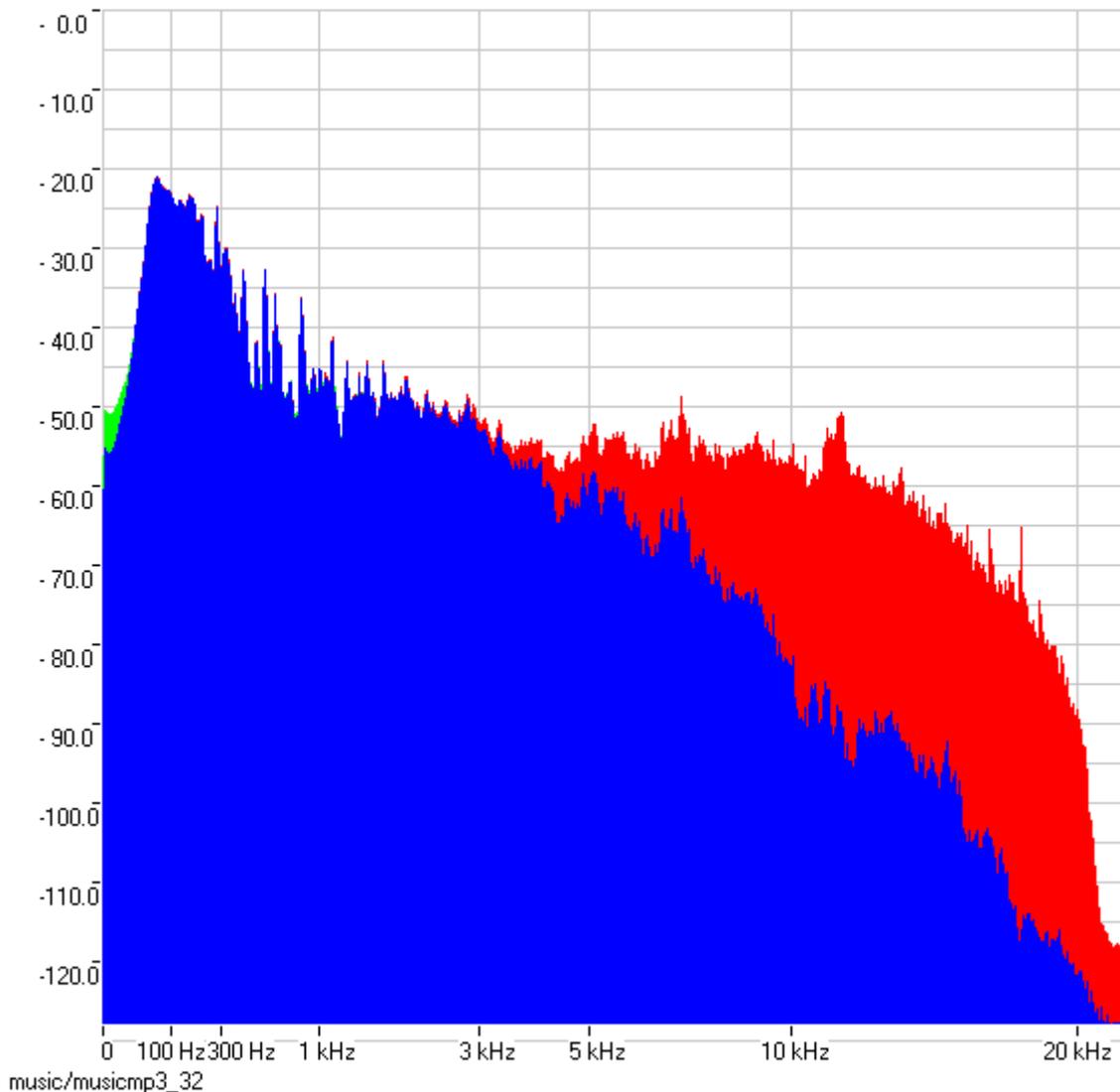


Рисунок 3.2.1 Спектры исходного файла и файла обработанного кодеком ВМС mp3

3.3 mp3 32 kbps (RazorLame)

Спектральный анализ файла показал почти полную схожесть результата работы mp3 с исходным сигналом в диапазоне частот 0-8 kHz. Необычная ситуация наблюдается в диапазоне от 8 до 11 kHz, где построенный спектр показывает присутствие звуков с амплитудой сигнала -120 ДБ. Это ничтожно малая громкость сигнала и она не похожа ни на попытки кодека сохранить исходный звук, ни на артефакты преобразования Фурье. Если считать, что перед кодированием применялся только фильтр низких частот, настроенный на 8 kHz, то амплитуды более высокой части спектра должны быть или небольшими и равномерным, или постепенно спадающими по мере увеличения частоты. Вероятнее всего эти звуки являются шумом, возникшем в результате перевода звукового сигнала перед кодированием на другую частоту дискретизации (например, в 22050 Hz), либо шумы, возникающие при MDCT преобразовании.

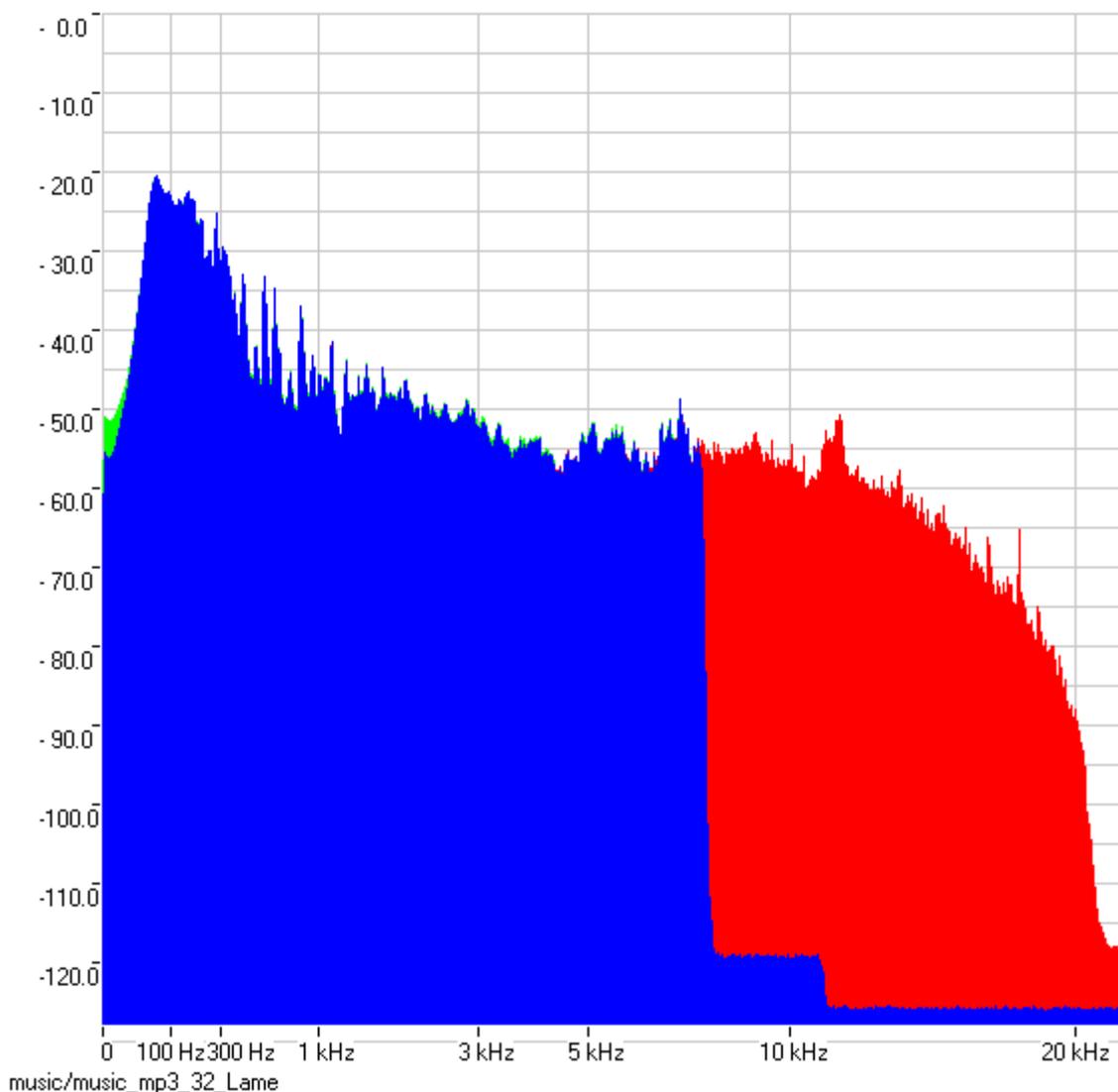


Рисунок 3.3.1 Спектры исходного файла, и файла обработанного кодеком Lame mp3

3.4 mp3pro

По графику спектра на рис 3.4.1 видно, что среднее распределение частот до 6 kHz очень похоже на распределение в исходном сигнале. Звук от 6 до 17 kHz подвергся более серьезным искажениям, характер которых хорошо виден на рис. 2.4.1 и рис. 2.4.2 и подробно описан при тестировании по сонограмме в пункте 2.4. На уровне семнадцати килогерц в mp3pro применен фильтр низких частот, в результате чего частоты выше 17 kHz отсутствуют в декодированном звуке.

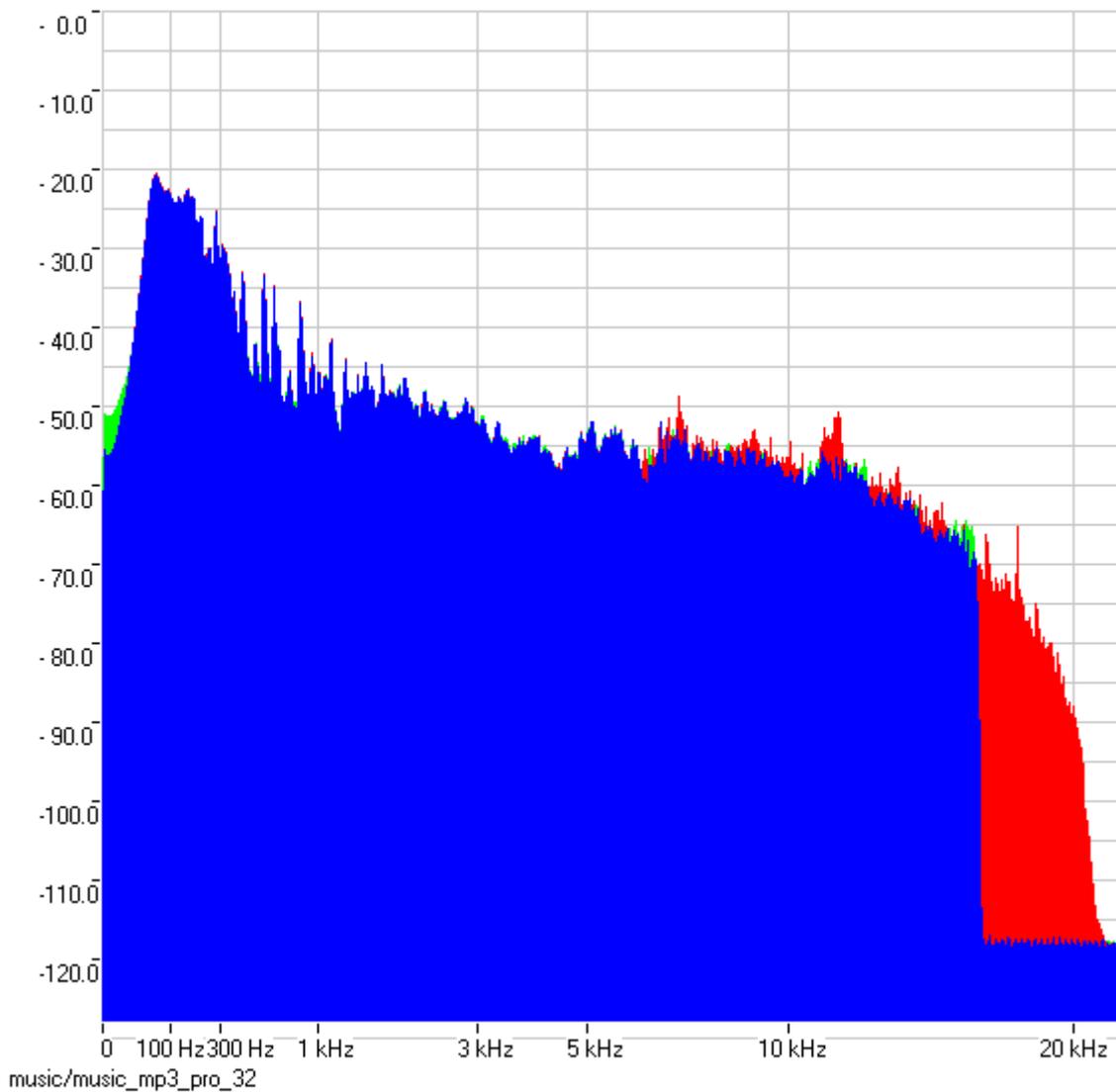


Рисунок 3.4.1 Спектры исходного файла, и файла обработанного кодеком mp3pro

3.5 WMA

На рис 3.5.1 заметно, что спектр до 8 kHz изменился достаточно слабо, а спектральный диапазон от 8 до 12 kHz претерпел серьезные изменения, но искажен меньше, чем у результата работы кодека mp3го.

Верхние частоты обрезаются начиная с 13 kHz.

Интересно отметить, что спектр обработанного кодеком сигнала в районе частоты 12 kHz очень похож на спектр исходного сигнала. Если также заметить, что именно в районе частоты 12 kHz и частоты 7 kHz наблюдается усиление амплитуд частотных составляющих, то напрашивается вывод, что настолько правильное сохранение частот близких к отметке 12 kHz не случайно.

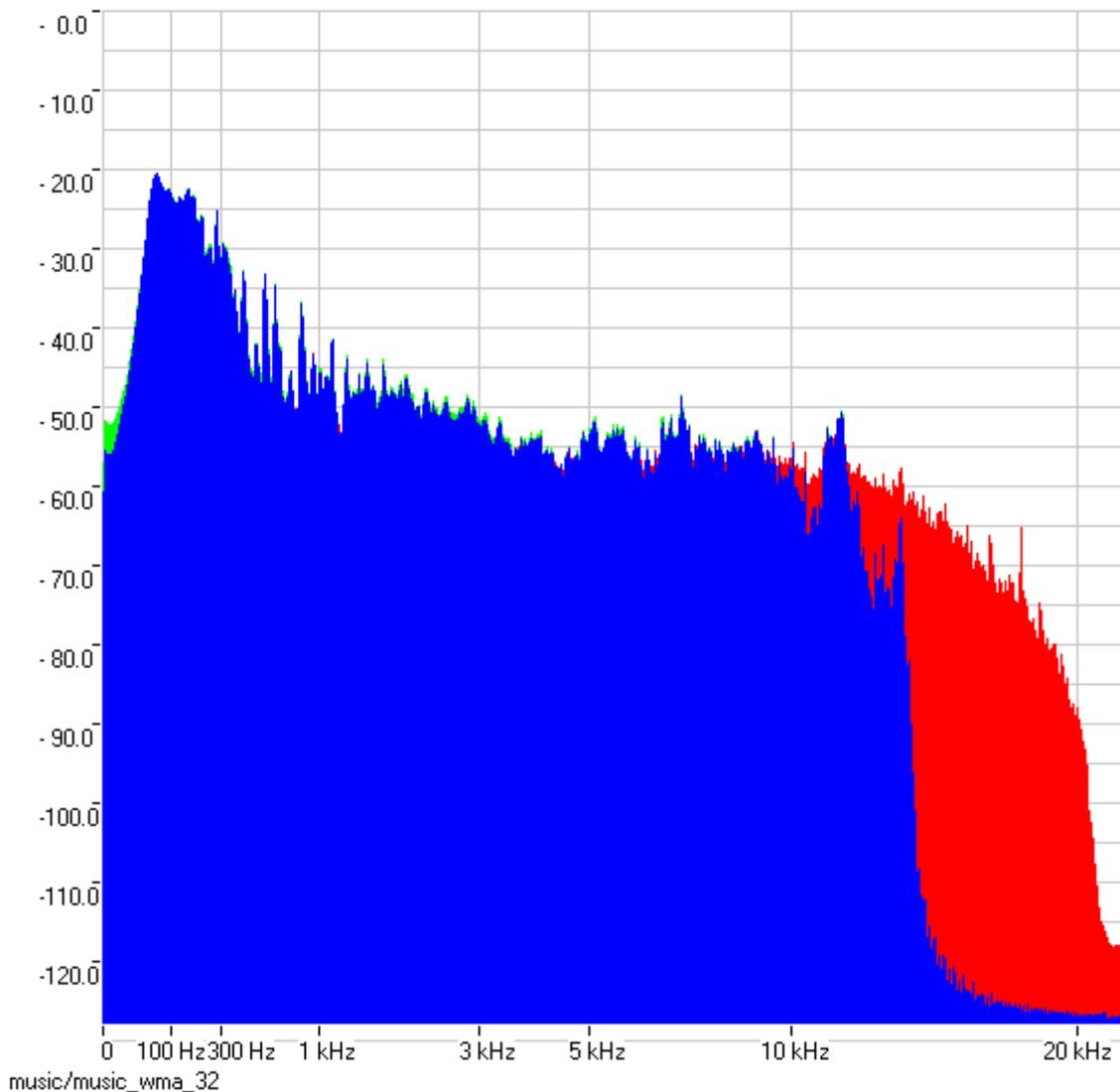


Рисунок 3.5.1 Спектры исходного файла, и файла обработанного кодеком WMA

Это подтверждают и результаты исследования спектра другого фрагмента музыкального файла, имеющего большую энергию сигнала на частоте, близкой к 9 kHz. (рис 3.5.2)

Видно, что в данном случае лучше всего сохранились частоты именно близкие к отметке в 9 kHz, а на частотах 7 и 12 kHz наоборот заметны явные отличия в общем спектре всего звукового сигнала.

Вероятно, при сжатии кодек определяет какие частотные полосы вносят наибольший вклад в звучание композиции и уделяет им наибольшее внимание.

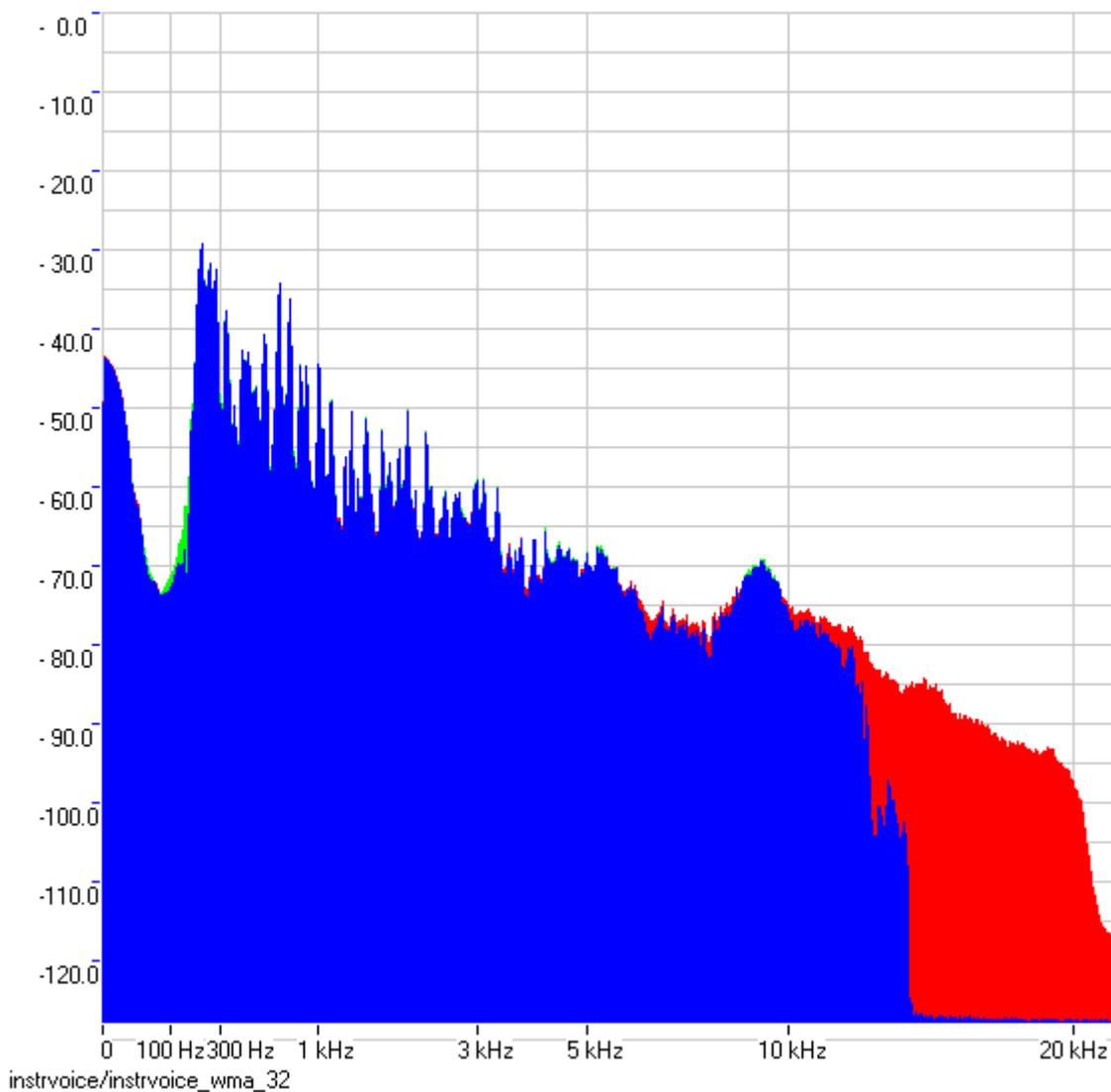


Рисунок 3.5.2 Спектры исходного музыкального файла, и файла обработанного кодеком WMA

3.6 AAC

По спектральной характеристике на рис. 3.6.1, так же как и по сонограмме, видно, что кодек применил фильтр высоких частот, убрав звуки находящиеся выше частоты 10 kHz. При сравнении сонограмм в сигнале, обработанном кодеком AAC, было заметно наличие звуков с частотой выше 10 kHz в местах, где на исходном сигнале присутствовали резкие удары. В спектральной характеристике всего файла выше 10 kHz наблюдается плавный спад интенсивности звука по мере увеличения частоты.

Наличие этих посторонних звуков в высокочастотном диапазоне не похоже на попытки кодека сохранить какие-либо высокочастотные составляющие (слишком уж мала и равномерна амплитуда получающегося сигнала). Скорее эти звуки возникают в связи с особенностями алгоритма кодирования AAC.

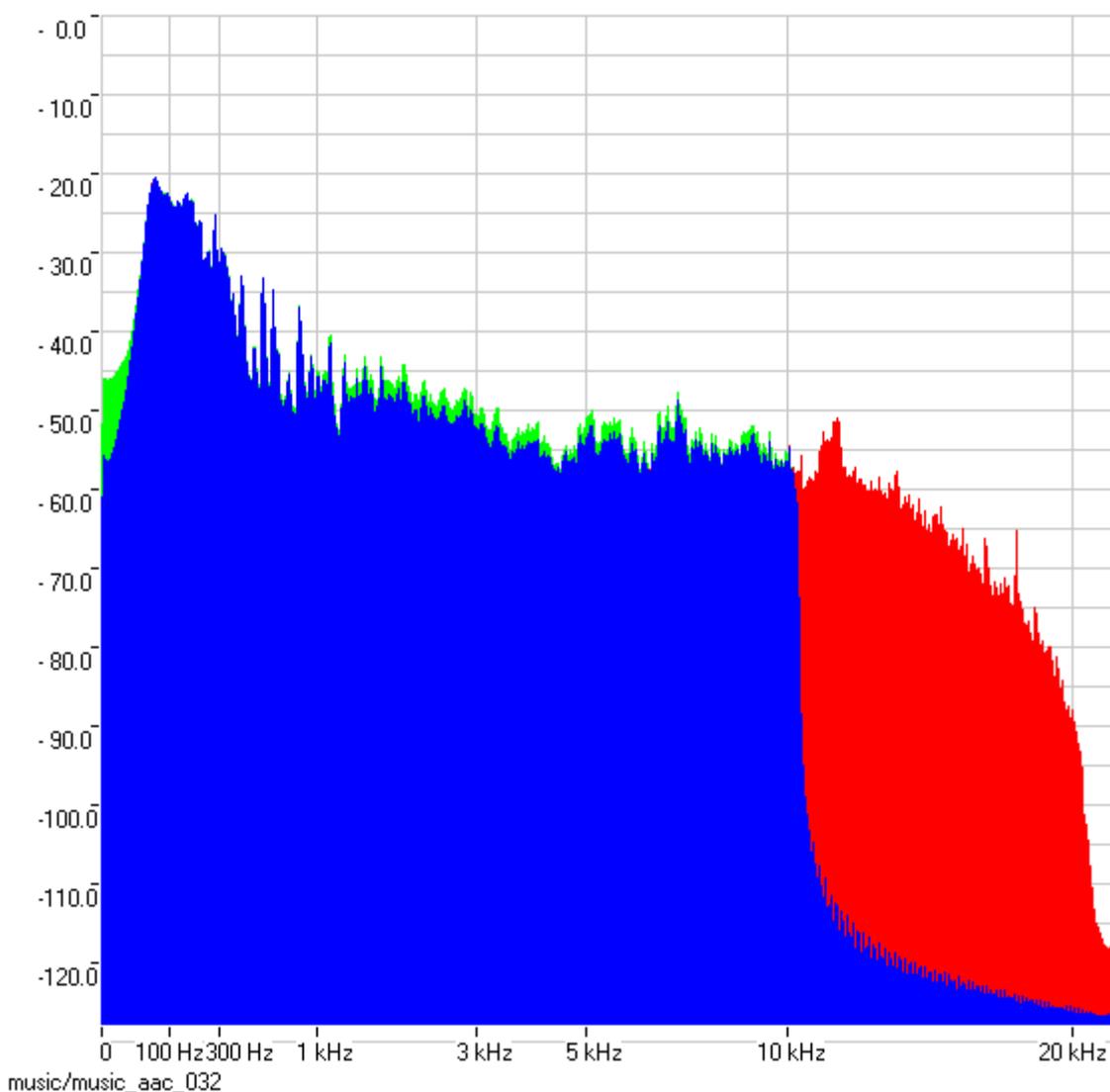


Рисунок 3.6.1 Спектры исходного файла, и файла обработанного кодеком AAC

По спектру четко видно, что в диапазоне 1.5-10 kHz амплитуда частотных составляющих в среднем увеличилась на 3-4 ДБ, что, вероятно, должно отразиться на общем звучании, которое должно стать более сухим, по сравнению с исходным звуком. На практике же, вероятно, получится не так, поскольку в сравнении на слух используется исходный файл, в кото-

ром присутствуют все частоты вплоть до 22 kHz и звук без высоких частот (выше 10 kHz), но с завышенной амплитудой у частотных составляющих в районе 2-9 кН. Про то, как именно измениться звук на слух по спектрам точно сказать нельзя, очевидно только то, что он будет более глухим за счет практически полного отсутствия высоких частот.

3.7 GSM 6.10

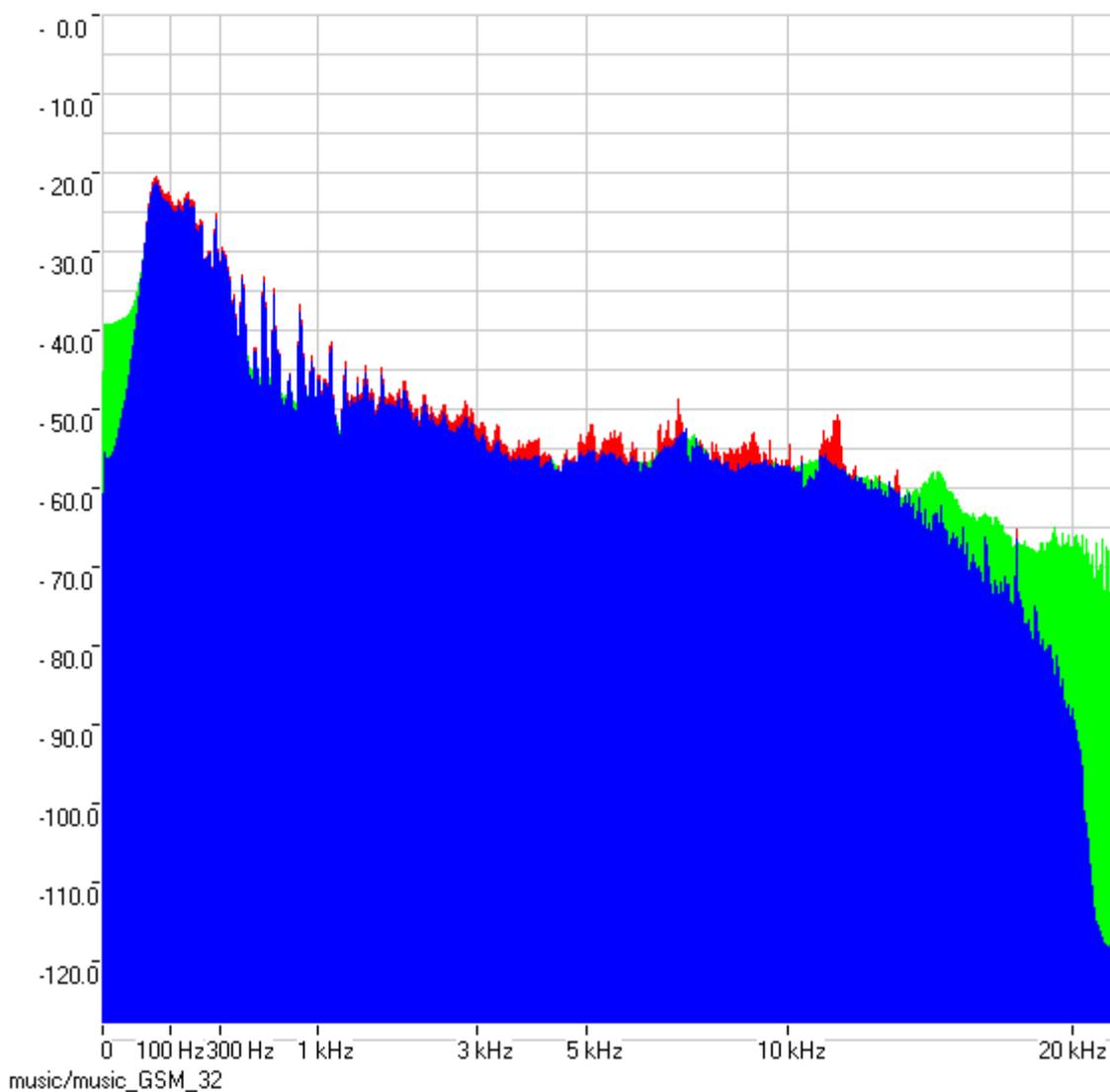


Рисунок 3.7.1 Спектры исходного файла, и файла обработанного кодеком GSM

Как было установлено при изучении сонограммы, если кодеку на вход подать моно файл 44100 Hz @ 16 bit, то он его исказит до ужасного состояния, хотя это почти незаметно по спектральной характеристике, не считая того, что сильно увеличилась мощность высоких частот, начиная от 13 kHz (рис 3.7.1)

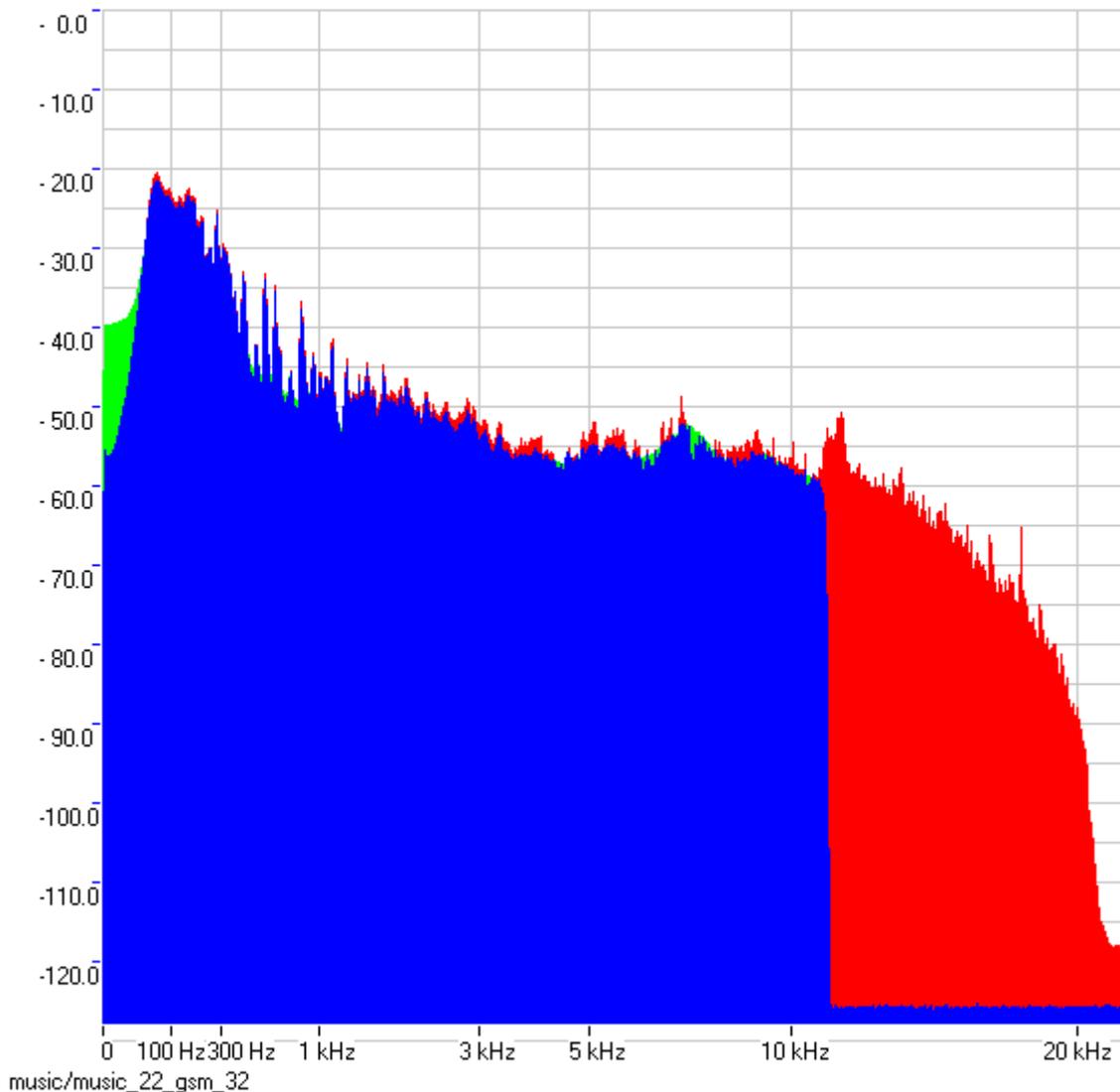


Рисунок 3.7.2 Спектры исходного файла, и файла обработанного кодеком GSM на частоте дискретизации 22050 Hz

Файл music_22_gsm_32 получен в результате ресэмплинга исходного файла, к 22050 Hz, последующему кодированию — декодированию кодеком GSM, и возврату в формат 44100 Hz @ 16 bit.

При сравнении исходного файла, с файлом music_22_gsm_32 и сопоставлении с результатами анализа файла music_gsm_32, видно, что спектр файлов в диапазоне до 11 kHz отличается очень мало, в то время, как уже установлено, что сигнал в файл music_gsm_32 был записан неправильно. Таким образом, можно говорить о том, что сравнение кодеков только по спектру полученных файлов дает очень мало информации о реальном качестве восстановленного звука.

4 Спектральный анализ разницы сигналов

4.1 Описание методов

Данный принцип тестирования также основан на изучении спектра всего сигнала целиком. Сигнал, по которому строится спектр, является разностью исходного и декодированного. Данный способ учитывает как сохра-

нение амплитуды и формы сигнала, так и его фазы. Чем меньше значение амплитуд частотных составляющих разностного сигнала (график ниже относительно основного спектра), тем большую схожесть имеют входные сигналы.

В качестве примера можно привести разность двух сигналов одной частоты (1 kHz) с взаимным фазовым сдвигом на $\pi/2$ (рис 4.1.1)

Красная линия обозначает спектр исходного файла а зеленая — обработанного. Синий – спектр разницы сигнала.

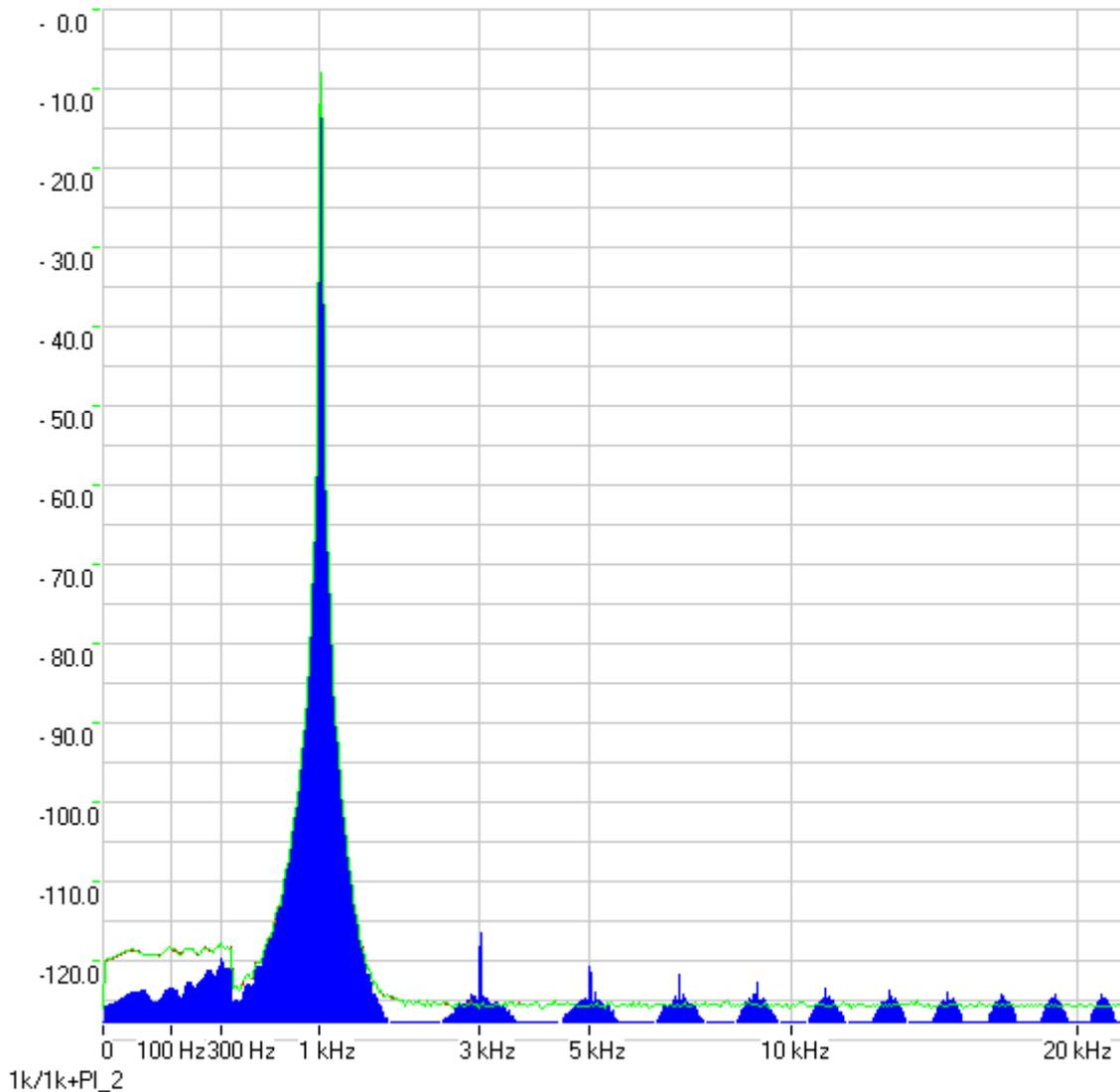


Рисунок 4.1.1 Спектр разницы сигналов в виде синусоиды на частоте 1 kHz, со сдвигом относительно друг друга на $\pi/2$

4.2 mp3 (Blaze Media Convert)

Как видно на рис. 4.2.1, лучше всего кодек справился с частотами от 30 до 300 Hz. Отличие между исходным сигналом (обозначен красной линией) и сигналом разницы в этой области составляет около 30 ДБ. Несколько худший результат кодек показал в области 300-5000 Hz, где разница составила всего лишь около 10 ДБ.

В области 5-22 kHz данный кодек, по сравнению с остальными тестируемыми, справился плохо, что означает, что фазы волн в этом диапазоне сохранены хуже всего. Но, в целом, это уже не так важно, поскольку в предыдущих тестах было установлено, что mp3 кодек от VMC показывает нормальные результаты только в области до 3 kHz. Данный тест подтверждает, что нормальное качество звука кодек дает только на средних и низких частотах.

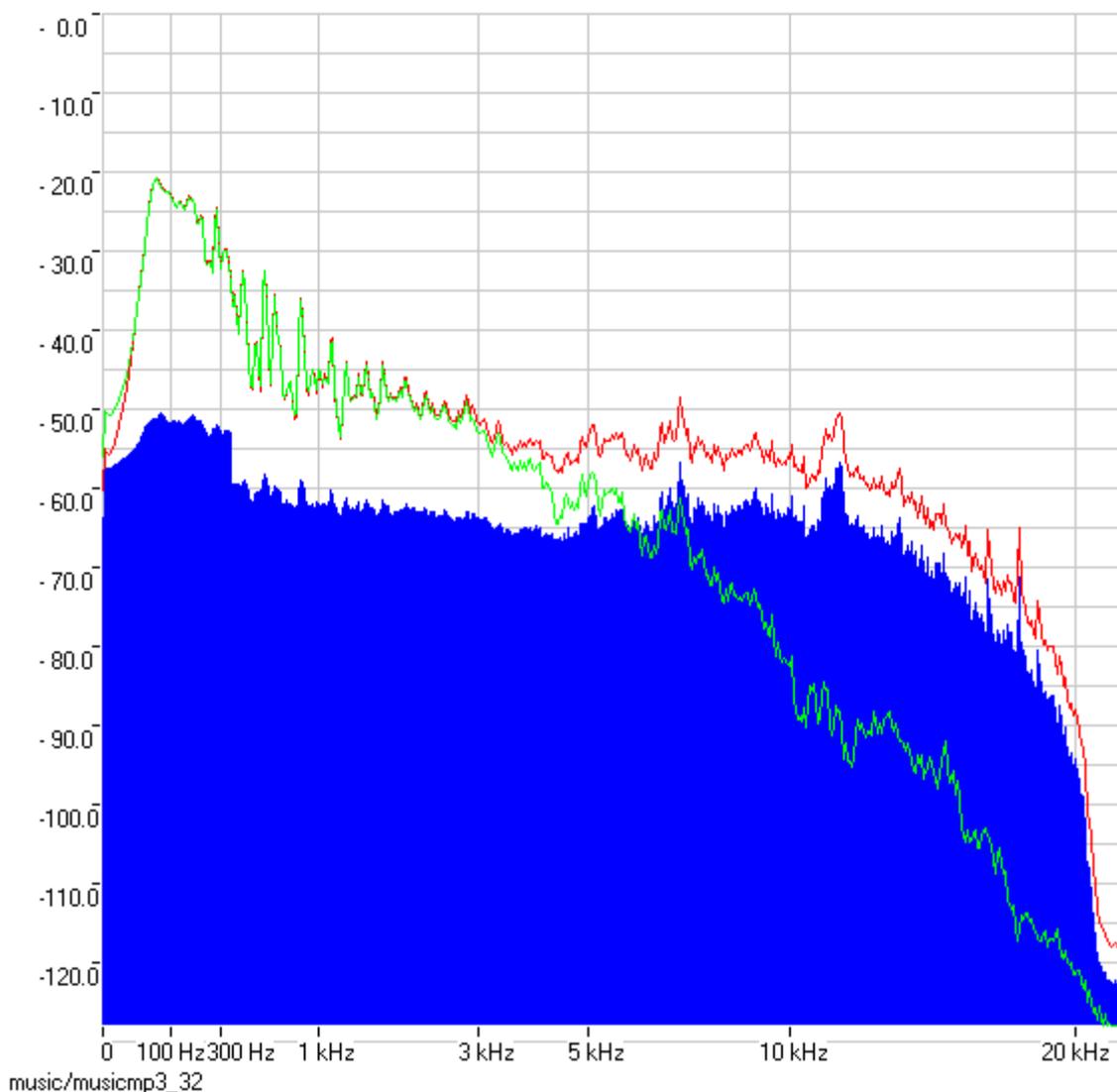


Рисунок 4.2.1 Спектры исходного файла (красный), файла, обработанного кодеком VMC mp3 (зеленый), и спектр разницы (синий).

4.3 mp3 (RazorLame)

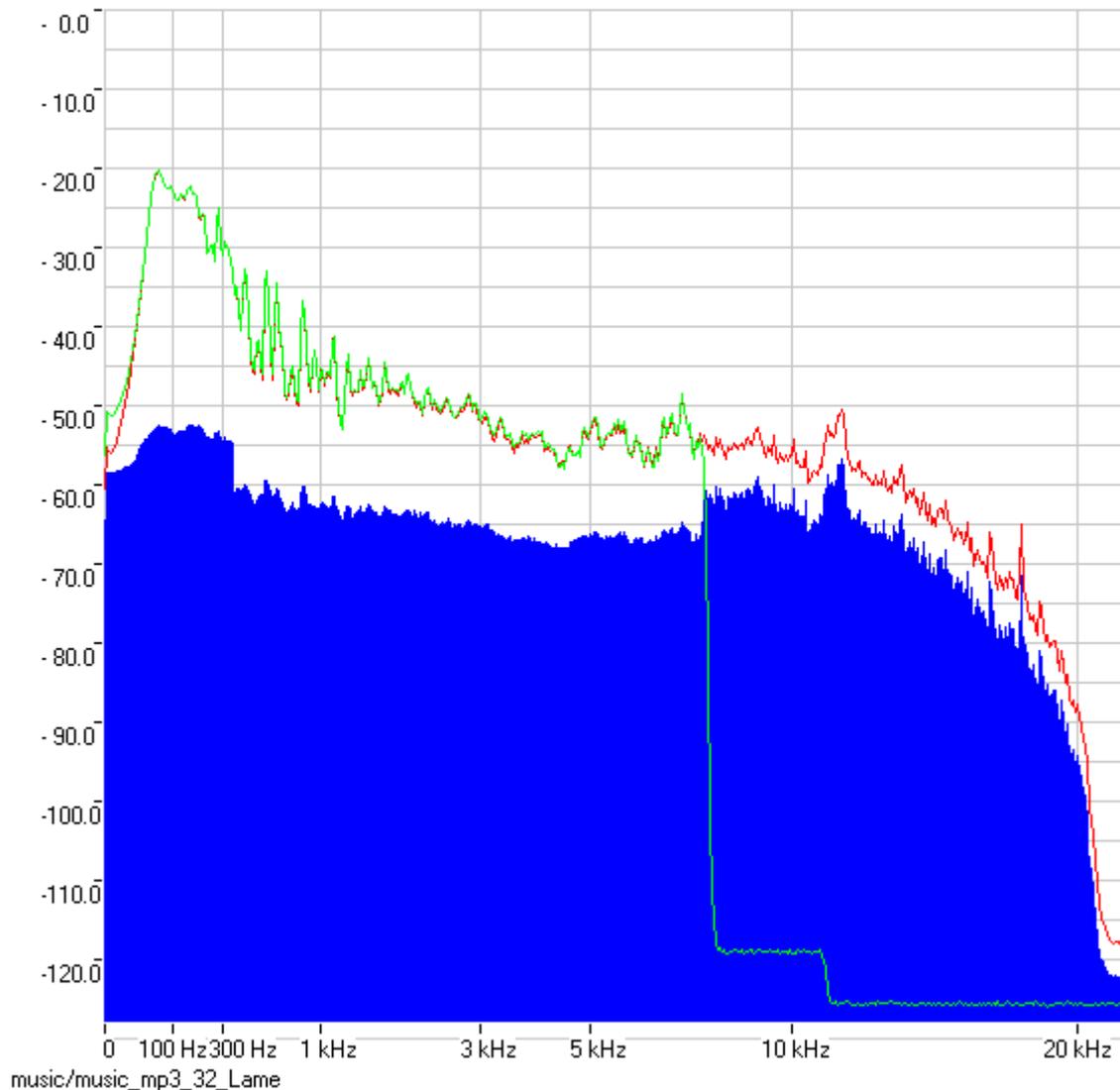
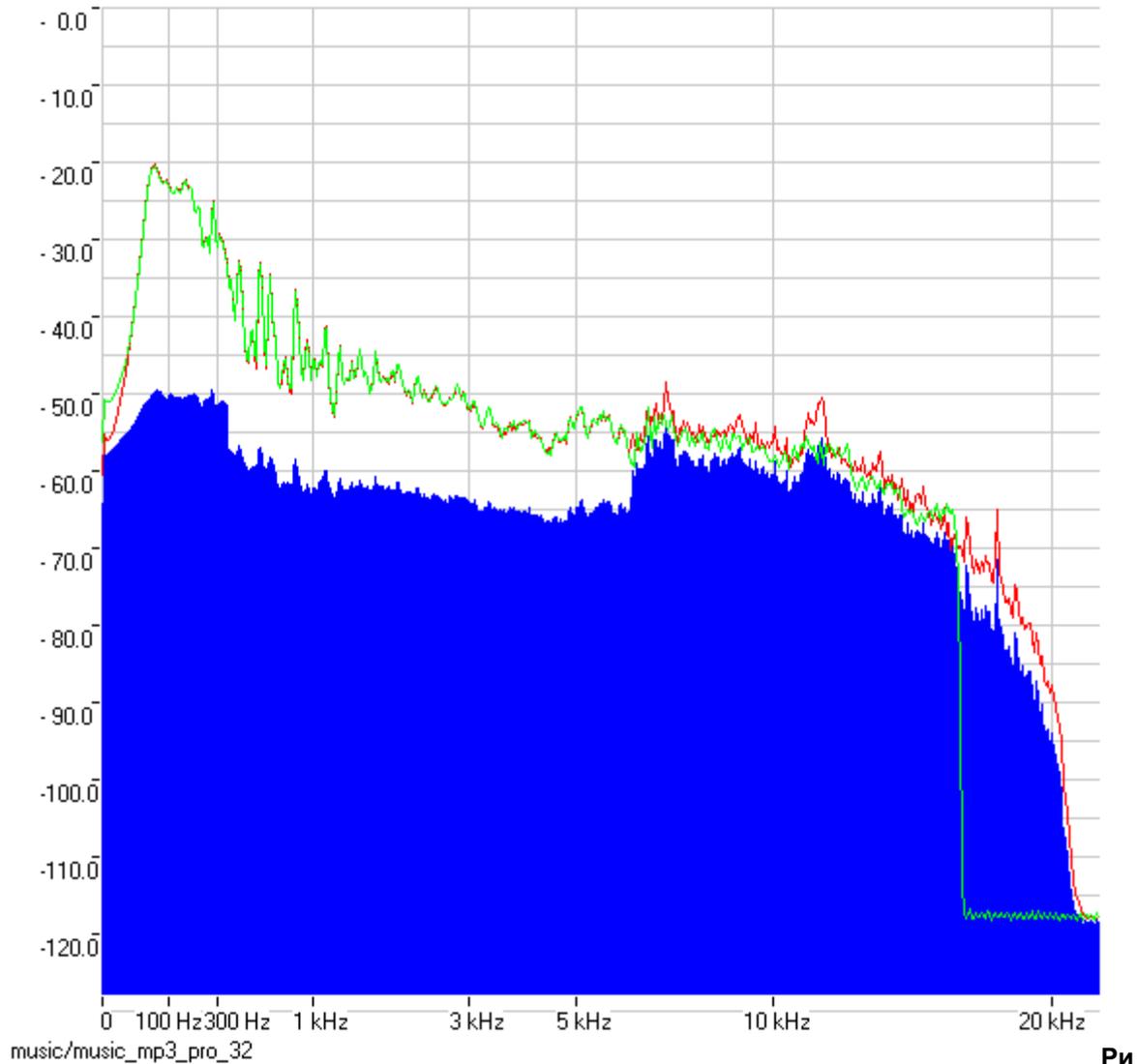


Рисунок 4.3.1 Спектры исходного файла (красный), файла, обработанного кодеком Lame mp3 (зеленый), и спектр разницы (синий).

По графикам на рис 4.3.1 видно, что, так же как и VMC, Lame наиболее качественно справился с кодированием частот (и фаз) в области 30-300 Hz, но результаты в этой области в среднем несколько лучше, чем у VMC (на 3-5 ДБ). В области 300 Hz-8 kHz Lame также справился немного лучше, чем VMC — его результат 10-15 ДБ. В диапазоне выше 8 kHz рассматривать данные нет смысла, так как они являются практически копией исходного сигнала, с амплитудой уменьшенной в 2 раза, потому что у Lame в данном диапазоне звук был отфильтрован.

4.4 mp3pro



сунок 4.4.1 Спектры исходного файла (красный), и файла обработанного кодеком mp3pro (зеленый), и спектр разницы (синий).

С кодированием области 30-300 Hz mp3pro справился лучше, чем все вышеприведенные кодеки mp3. В области 300-6000 Hz кодек показал несколько более скромный результат, чем Lame и примерно равный результату VMC. А вот в частотном диапазоне от 6 до 17 kHz кодек показывает неожиданно большую разницу с исходным сигналом. Здесь и проявляется ключевое отличие MP3-PRO от других кодеков, поскольку он как видно из графиков сохраняет энергию, но не сохраняет фазы сигнала.

4.5 WMA

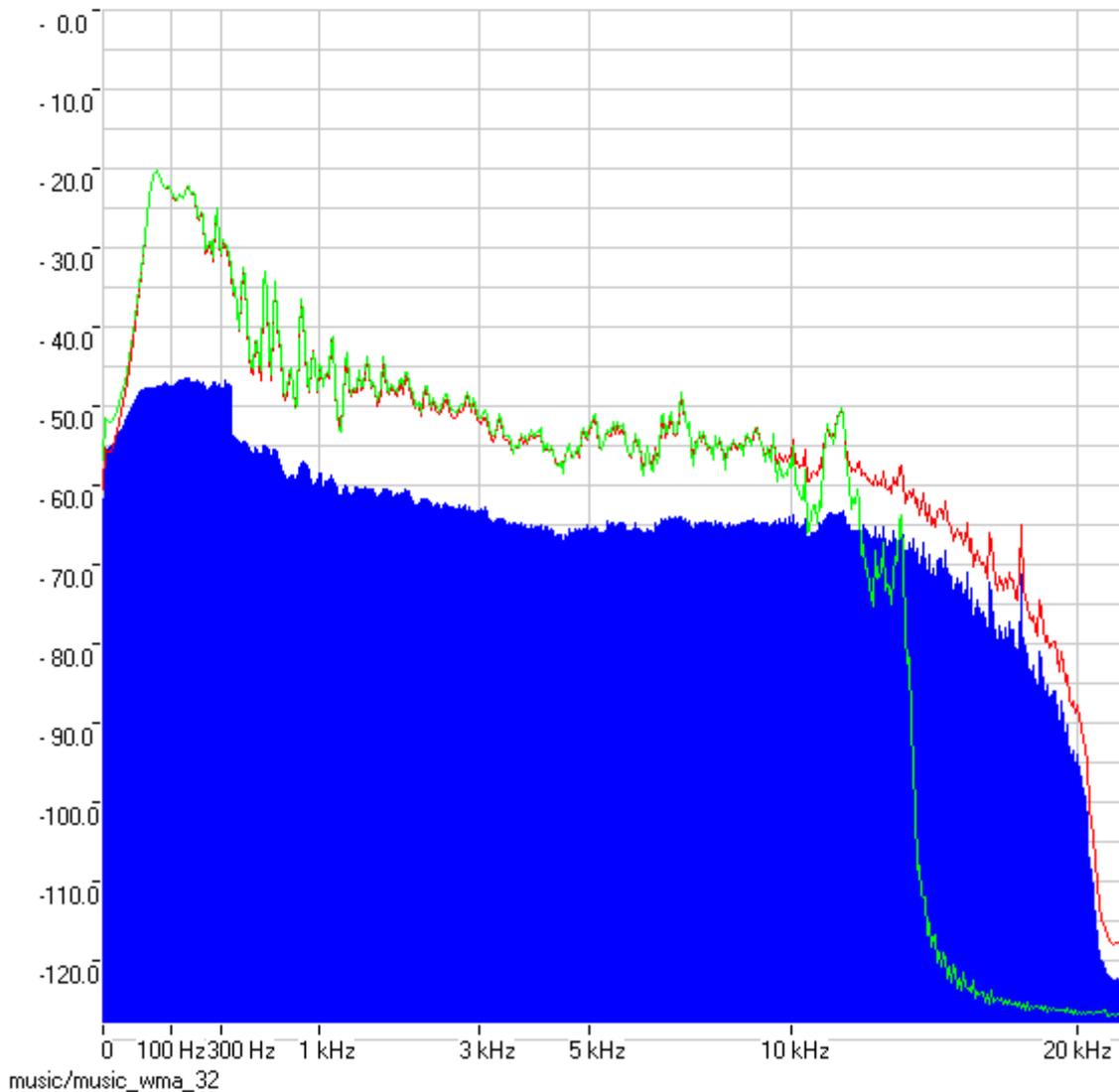


Рисунок 4.5.1 Спектры исходного файла (красный), и файла обработанного кодеком wma (зеленый), и спектр разницы (синий).

В области до 5 kHz кодек показал один из худших результатов среди уже рассмотренных кодеков (сравнимый с VMS). Зато в диапазоне 5 - 6 кГц результат по разнице формы волны лучше, чем у mp3pro и VMS, а на частотах 6 - 12 kHz кодек отработал лучше всех вышеперечисленных. Если оценивать кодирование частот (и сохранение их фаз) до 12 kHz, то WMA в целом работает лучше из всех тестируемых кодеков. Другое дело — нужно ли сохранять все фазы при кодировании или это не будет заметным. Некоторые разработчики придерживаются стратегии не всегда сохранять фазу, так как, например, белый шум будет ощущаться одинаково, не зависимо от конкретных значений фаз.

4.6 MPEG-4 AAC

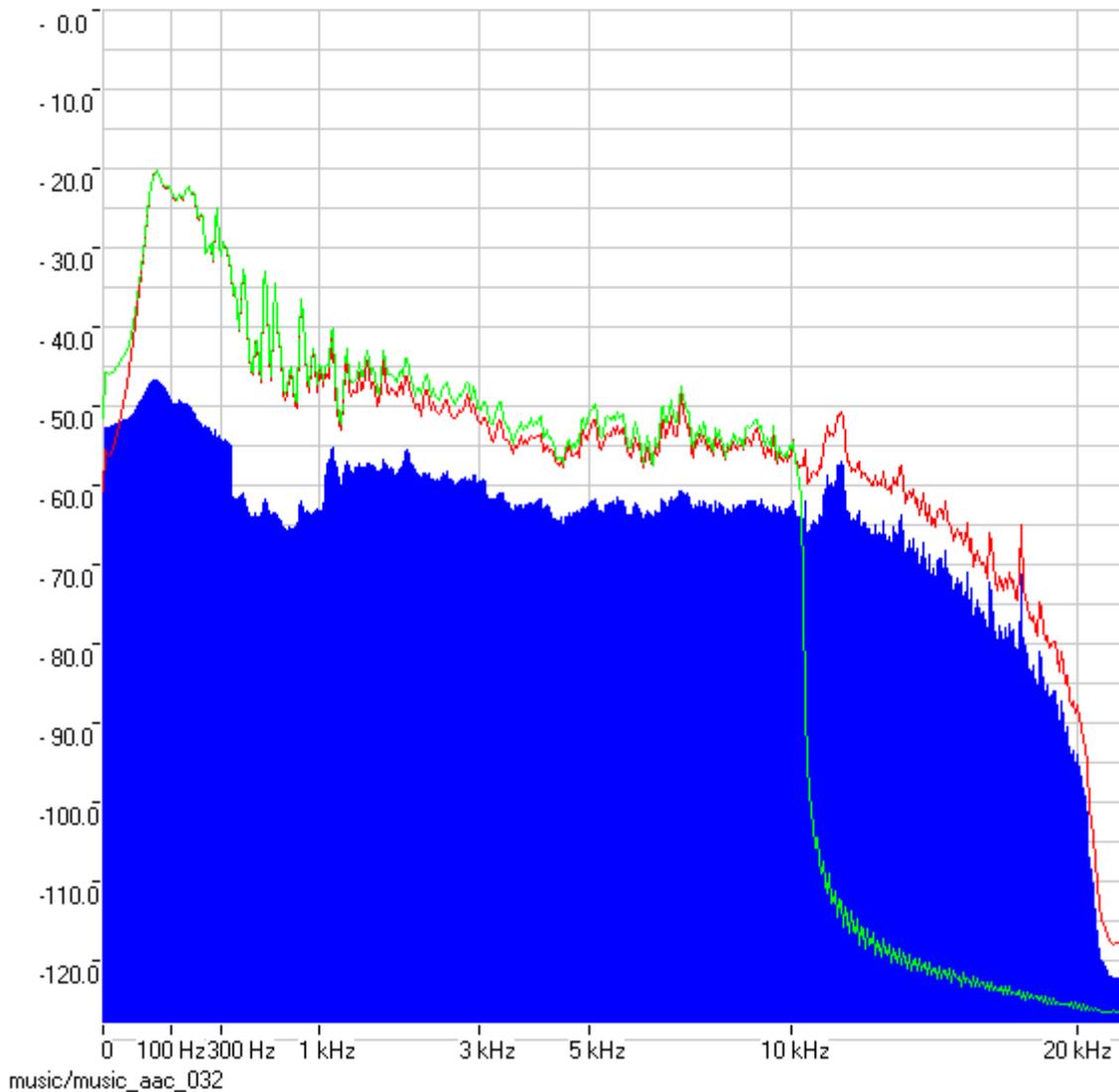


Рисунок 4.6.1 Спектры исходного файла (красный), файла, обработанного кодеком AAC (зеленый), и спектр разницы (синий).

Результат, показанный кодеком AAC на рис 4.6.1, довольно интересен — видно достаточно неплохое сохранение фаз низкочастотного диапазона, и очень хорошее, по сравнению с другими кодеками, сохранение диапазона от 300 до 1000 Hz, где отличие между исходным сигналом и сигналом разницы составляет порядка 15-20 ДБ. В остальном же частотном диапазоне результат получился хуже, чем у всех вышеописанных кодеков. Возможно, что более точное сохранение формы волны в области 300–1000 Hz - это правильный подход, ведь именно в этом диапазоне человек лучше всего различает звуки, однако в целом результат кодека не очень хорош.

4.7 GSM 6.10

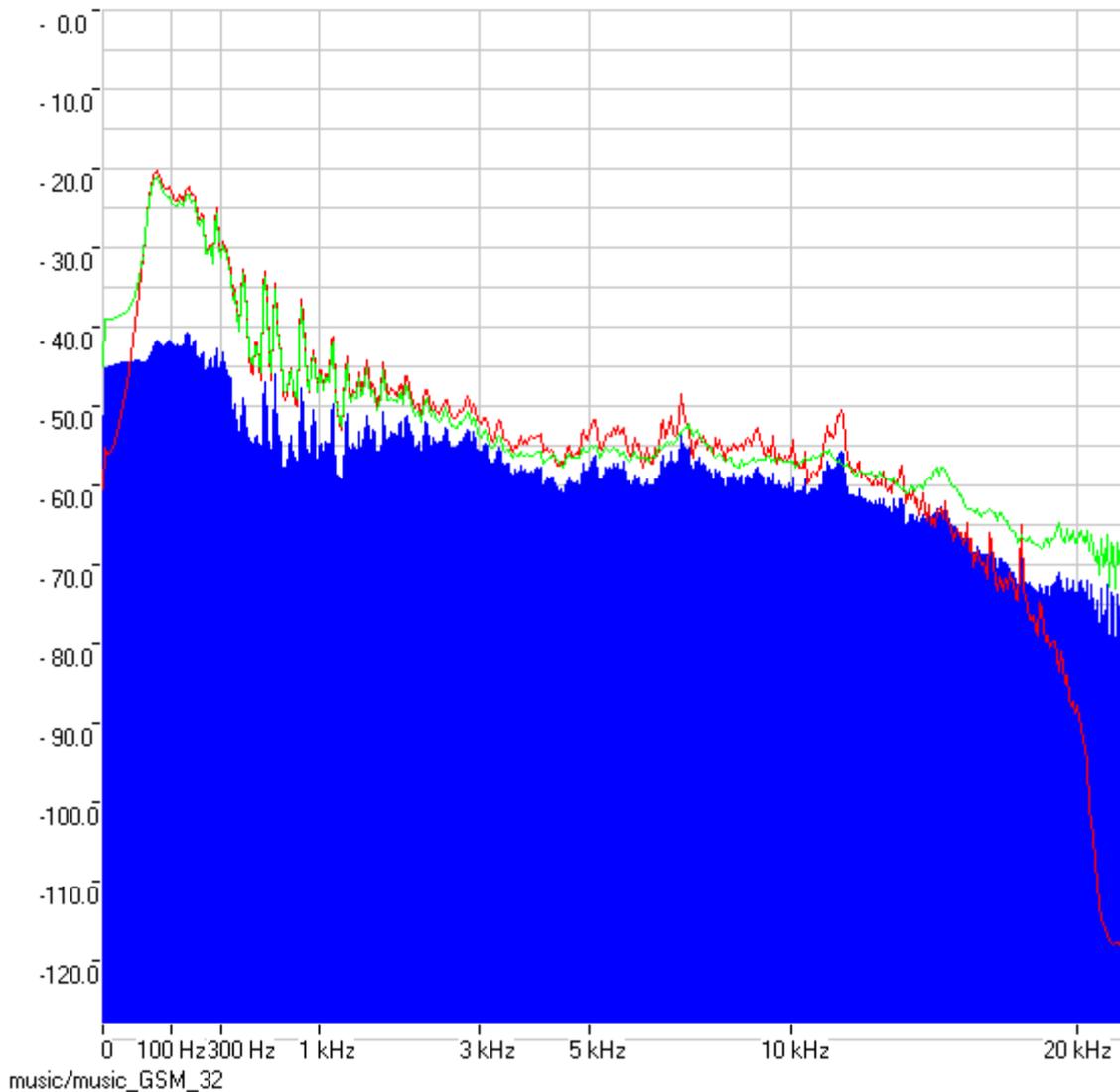


Рисунок 4.7.1 Спектры исходного файла (красный), файла, обработанного кодеком GSM 6.10 (зеленый), и спектр разницы (синий).

На рис 4.7.1 показан результат сравнения некорректно обработанного кодеком GSM файла (music_GSM_32.wav 44100 Hz @ 16 Bit) с оригинальным. Видно, что результат оказался очень плохим.

На рис 4.7.2 показан результат сравнения корректно обработанного кодеком GSM файла с оригиналом (подробное описание проблем и необходимых действий для их решения см. в пункте 3.7). Но в этом случае, во всех частотных областях видно явное отставание данного кодека от остальных.

В совокупности с результатами остальных тестов можно четко сказать, что данный кодек справляется со своей задачей хуже всех протестированных, что не удивительно, поскольку он специализирован только для сжатия речи, а не музыки.

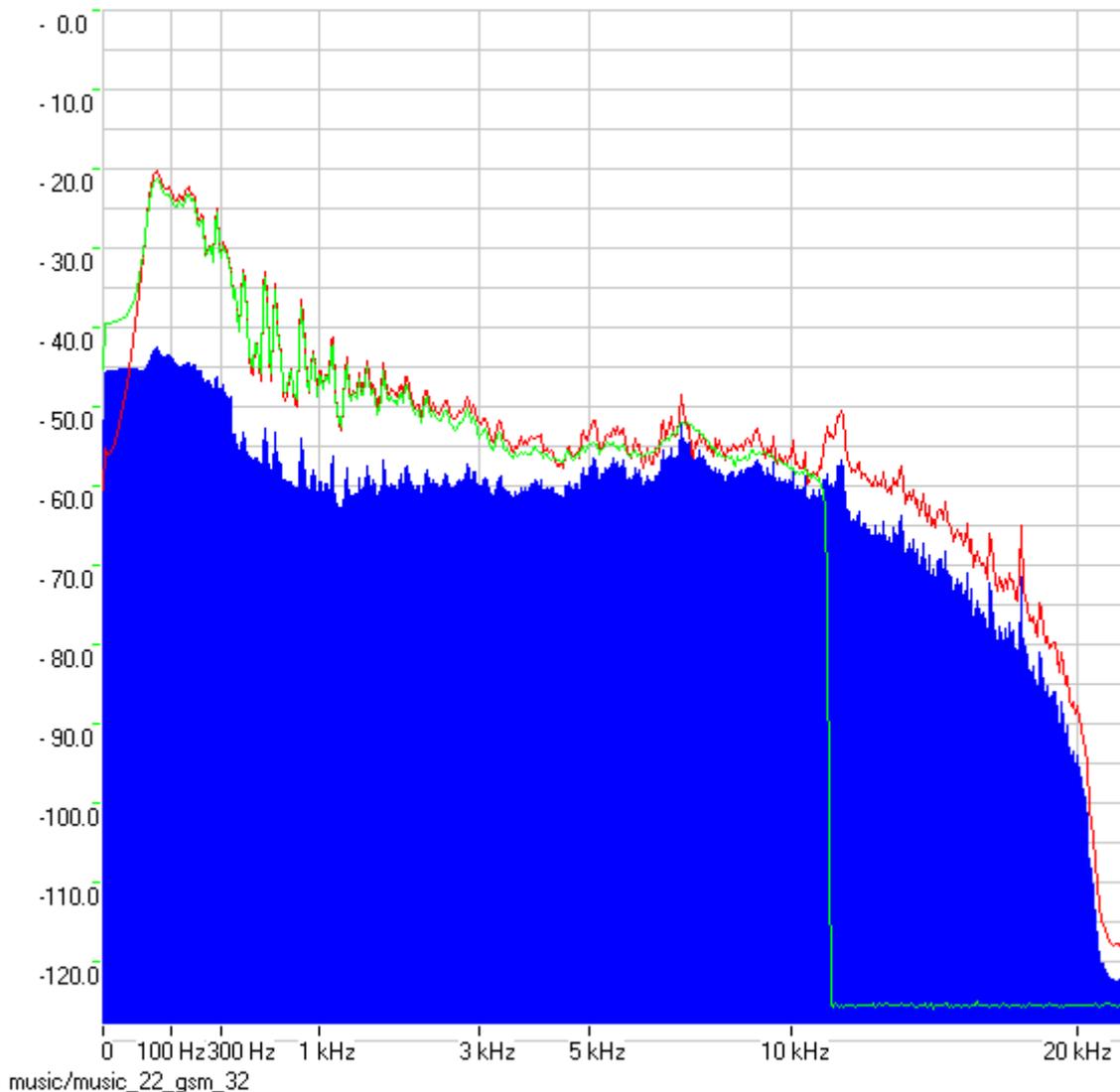


Рисунок 4.7.2 Спектры исходного файла (красный), файла, обработанного кодеком GSM 6.10 на частоте дискретизации 22050 kHz (зеленый), и спектр разницы (синий).

5 Тестирование по числовым характеристикам

Поскольку, как уже упоминалось выше, многие кодеки сдвинули сигнал в файлах на некоторое временное значение по сравнению с исходным, то в данном разделе, так же как и раньше, сравниваются сигналы, у которых скорректировано значение временного сдвига. В противном случае, если сдвиг оставить, то все характеристики сравнения, описанные далее, были бы не верны.

5.1 Описание характеристик, по которым проводилось сравнение

PSNR — величина, характеризующая отклонение амплитуды одного сигнала от другого.

Вычисляется по формуле:

$$\text{PSNR} = 20 * \log_{10} \left(32768 \times \sqrt{\frac{N}{\sum_{i \in N} (\text{Signal}_i^1 - \text{Signal}_i^2)^2}} \right)$$

и измеряется в децибелах. Signal1и Signal2 – это, соответственно, амплитуды первого и второго сигналов, а N – это число временных отсчетов в файлов (для разных пар файлов это число может быть разным, так как на его значение влияет как сдвиг одного файла относительно другого, так и длина раскодированного файла, которая различается в зависимости от применяемого кодека).

Максимальное отклонение - это максимальная разница в значении между амплитудами семплов, стоящих на одной временной позиции в файлах.

$$\text{MD} = \text{Max}_{i \in N} \left(\left| \text{Signal}_i^1 - \text{Signal}_i^2 \right| \right)$$

Среднее отклонение характеризует среднюю величину отклонения сигнала по всему сигналу и высчитывается следующим образом:

$$\text{SrD} = \frac{\sum_{i \in N} \left| \text{Signal}_i^1 - \text{Signal}_i^2 \right|}{N}$$

Среднее интегральное сигналов характеризует сохранение энергии сигнала в файлах.

Среднее интегральное значение первого файла — это значение, посчитанное для исходного не кодированного файла.

$$SI^k = \frac{\sum_{i \in N} |Signal_i^k|}{N}$$

Тем не менее, как видно из таблицы 5.1, оно немного отличается от строки к строке. Связано это с тем, что, как уже упоминалось выше, число N, по которому ведется суммирование значений, различается в зависимости от длины и сдвига файла, обработанного кодеком. Соответственно, так как значения исходного файла каждый раз пересчитывались с различными значениями N, то и результаты подсчета немного отличаются (и это вполне нормально).

Значение разницы между средними интегральными значениями высчитывается по формуле:

$$SID = |SI^1 - SI^2|$$

характеризует среднее изменение энергии всего сигнала (в целом).

5.2 Сравнение числовых характеристик

	PSNR (ДБ)	Макс. Откло- нение (MD)	Среднее отклонение (SrD)	Среднее инт- ое первого файла (SI1)	Среднее инт-ое вто- рого файла (SI2)	Разница между сред- ними инт- ыми значениями (SID)
ВМС mp3	27.16	15308	922.0105	3625.0046	3414.553	210.45135
Lame mp3	23.55	22234	1426.815	3690.3518	3630.920	59.431421
mp3 pro	24.9	22128	1185.331	3690.4405	3598.730	91.710370
WMA	27.57	13672	942.7687	3690.7552	3758.105	67.350286
AAC	26.1	16289	1097.900	3692.2582	3818.388	126.12979
GSM 44k	20.75	32977	2008.176	3690.4405	3318.562	371.87772
GSM 22k	23.11	25518	1502.358	3690.4405	3301.201	389.23938

Таблица 5.1

	BMC mp3	Lame mp3	mp3 pro	WMA	AAC	GSM 44kHz	GSM 22kHz
PSNR	1	4	3	1	2	7	4
MD	1	2	2	1	1	3	2
SrD	1	5	2	1	2	10	5
SID	6	1	2	1	3	12	12
Итого	2,25	3	2,25	1	2	8	5,75

Таблица 5.2

В таблице 5.2 сведены результаты анализа таблицы 5.1 по различным способам тестирования и расставлен условный балл.

Методика построения шкалы оценок.

Выбирается самый лучший результат и ему присваивается единица, дальше, по мере ухудшения результатов, балл увеличивается пропорционально ухудшению результата теста. Так, например, для PSNR увеличение его значения на единицу относительно лучшего результата, соответствовало увеличению балла на единицу, после чего баллы для каждого кодека суммировались по вертикали, и делились на количество номинаций сравнения, т.е. на 4 и таким образом получался итоговый балл. Во всех случаях, чем меньше балл — тем лучше.

По итогам сравнения можно видеть, что в данных тестах лучший результат показал WMA. Второе место можно разделить между AAC, BMC mp3 и mp3pro, правда BMC mp3 показывает невысокий результат при анализе среднего интегрального значения по сравнению с AAC и mp3pro. Третье место следует отдать Lame mp3. Кодеки показывают не очень хороший результат по PSNR, но отлично (лучше всех) проявляет себя в тесте среднего интегрального отклонения.

Четвертое место занял GSM 22kHz. Хороший результат по PSNR, но по среднему интегральному отклонению кодек находится на последнем месте.

GSM 44 kHz во всех тестах является аутсайдером, и, как уже было установлено, происходит это из-за ошибки в способе сжатия кодеком файлов с частотой дискретизации 44100 Hz, поэтому в дальнейшем работа GSM 44kHz рассматриваться не будет.

6 Сравнение полученных в результате тестирования данных

По результатам тестирования по сонограмме, лучшим следует признать кодек WMA. За ним следует mp3pro, потом AAC и Lame, далее, в зависимости от цели использования, GSM — если предполагается кодировать речь и BMC — если нужно кодировать музыку.

По спектральному методу тестирования также лидирует WMA, дальше идет mp3pro, а потом Lame mp3. За ним идет AAC. Несмотря на то, что у AAC не сохранена энергия звука в среднечастотном диапазоне, контуры спектра до кодирования и после кодирования практически идентичны, что является хорошим показателем. Относительно неплохой спектр получился у GSM, и совсем плохой у BMC mp3.

В тесте спектрального анализа разницы сигналов хороший результат показывает Lame mp3, а за ним следуют кодеки BMC mp3, WMA и mp3pro. Mp3pro и BMC mp3 лучше справились на нижнем и среднем частотном диапазоне, а WMA показывает результат несколько хуже в низкочастотном диапазоне, но зато равномерно неплохой во всей частотной области, которую он кодирует. У AAC результат несколько худший, и совсем плохой у GSM.

Распределив места, полученные кодеками по результатам тестирования тремя вышеописанными методами в общую таблицу вместе с данными о качестве, полученными из таблицы 5.2, можно построить следующее сводные результаты качества работы:

	BMC mp3	Lame mp3	Mp3 pro	WMA	AAC	GSM
Сонограмма	7	5	3	1	5	7
Спектр	7	3	2	1	4	5
Спектр разницы	2	1	2	2	3	6
Тест числовых характеристик	2	3	2	1	2	4
Итоговый балл	4,5	3	2,25	1,25	3,5	5,5

Таблица 6.1

Из таблицы видно, что по результатам всех тестов лучшие показатели у mp3pro и WMA, второе место делят AAC и Lame mp3.

Третье место занимает BMC mp3 и самый плохой результат у GSM.

Следует отметить, что первые три строки таблицы базируются на визуальном анализе изображений, а не на сопоставлении чисел, полученных из этих изображений или прослушивании данных записей, а четвертая строка основана на сопоставлении числовых характеристик.

7 О форме звуковых волн в различных кодеках

Отличия в принципах кодирования звуковой информации музыкальными кодеками и речевыми кодеками также сказываются на форме получающейся волны. Это, в свою очередь сказывается как на восприятии сжатого сигнала человеком, так и на результатах всех описанных методов сравнения.

На рисунке 7.1 представлена исходная музыкальная волна, на рис 7.2 результат сжатия mp3 и на рис 7.3 сжатие кодеком GSM.

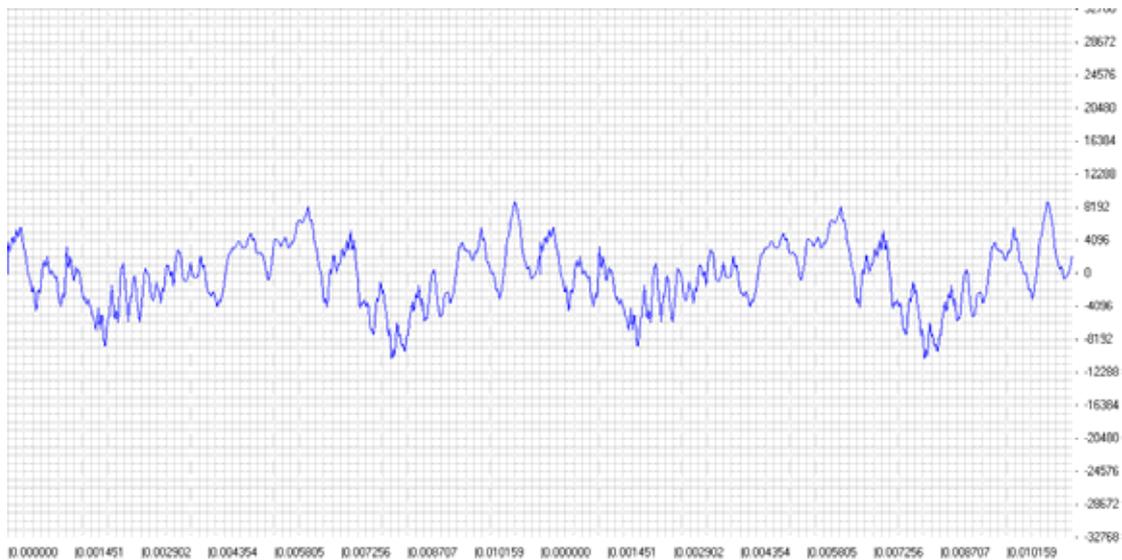


Рисунок 7.1 Исходный

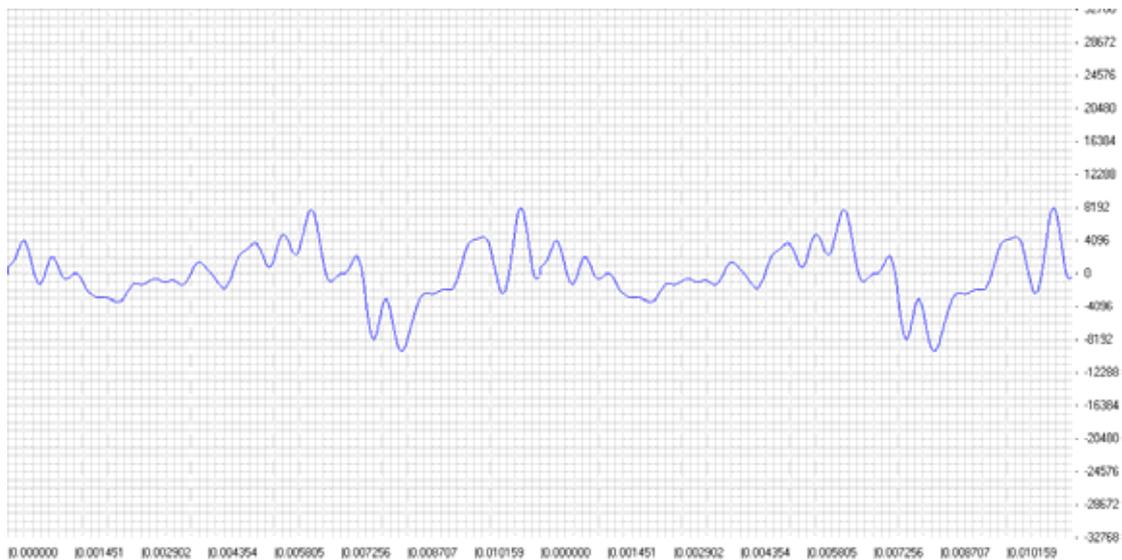


Рисунок 7.2 MP3

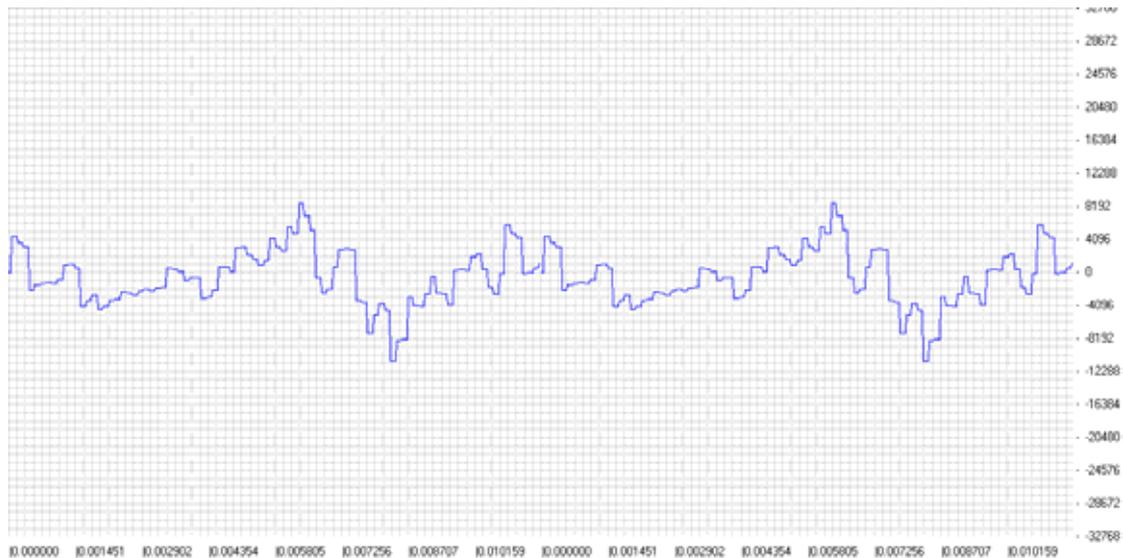


Рисунок 7.3 GSM

У GSM кодека сильно заметна ступенчатая структура формы волны связанная с существенно меньшим размером базового фрейма при сжатии и отсутствие перекрытий между блоками.

8 Субъективное сравнение кодеков на слух

В данном разделе описывается впечатление о прослушивании композиций. Включено описание результатов прослушивания двух звуковых фрагментов: music.wav (запись голоса солиста с музыкой) и speech.wav (запись речи диктора), т.к. на примере music.wav были описаны все методы тестирования и сигналы данного типа наиболее часто используются для кодирования. А образец речи, представленный в файле speech.wav, это совершенно другой тип звукового сигнала, нежели музыка и также является востребованным — например, в телеконференциях.

Тестирование производилось «вслепую». На каждой итерации тестирования выбирался наиболее плохо звучащий файл и удалялся из набора для тестирования. Когда оставался только один, лучший файл, все звуковые фрагменты располагались в порядке улучшения их звучания и сравнивались еще раз между собой и с оригиналом для определения особенностей их звучания и записывались впечатления от их прослушивания.

При таком тестировании часто применяют методы ABX, основанные на выставлении слепого рейтинга при прослушивании множества звуковых фрагментов, сжатых с разными параметрами или разными кодеками.

8.1 Оборудование для тестирования

Тест производился поочередно на двух наборах звуковоспроизводящей техники, результаты тестов в целом совпадают.

Первый набор звуковоспроизводящей техники:

- Аудио карта: Creative SB Live! 5.1

- Усилитель: Hi-Fi усилитель на микросхемах tda2030a (18 Вт) + tda7294(80 Вт)(НЧ канал)
- Акустические системы: Колонки 9 литров с ФИ RMS 50 Вт + Саб-вуфер RMS 200 Вт. Динамики рш130, 6гдв-1, 35гдн

Второй набор:

- Аудио карта: Creative SB Live! 5.1
- Усилитель: Усилитель от музыкального центра technics SA-EH60
- Акустические системы: Колонки technics SB-EH60 (80Вт (DIN))

9 Заключение

9.1 Результаты прослушивания music.wav

Как и следовало ожидать, самым плохим оказался GSM 6.10, которому на вход был подан сигнал с частотой дискретизации 44100 Hz.

Наблюдалось большое количество разнообразных призвуков, как на высоких, так и на низких частотах, в результате чего создавалось впечатление, что файл проигрывается через колонки, которые испытывают перегрузки, как будто на них подали сигнал значительно большей мощности, чем их номинальная и от этого они хрипят.

Звук в Lame mp3 имеет неприятную высокочастотную окраску, и при этом верхние частоты звучат очень плохо. Наблюдается хаотичность появления высокочастотных призвуков, мало похожих на высокочастотные составляющие в исходном звуке.

Более качественно звучат файлы, прошедшие обработку кодеками VMC mp3 и GSM 6.10 на частоте дискретизации 22050 Hz.

Сразу не ясно, что именно хуже звучит, т.к. воспринимаются они совсем по-разному.

В GSM слышны звуки, похожие на помехи (шум), а VMC просто звучит очень глухо, хотя и неплохо. Тем не менее, наличие шума сильно раздражает, поэтому третье место с конца все-таки следует отдать GSM 22 kHz, а четвертое место имеет смысл отдать VMC mp3. На примере этих кодеков становится хорошо видно, что в некоторых случаях лучше убрать высокие частоты совсем, чем сохранить их, пожертвовав качеством звука в целом.

Далее идет пара файлов, которая звучит намного качественнее предыдущих, но разницу между которыми было очень сложно определить. Похоже, что AAC звучит несколько хуже, чем WMA. Во-первых, голос закодированный первым кодеком больше отличается от оригинала, чем вторым. Во-вторых, призвуки на высоких частотах, которые отсутствуют в оригинале, есть и в том, и в другом файле, но в WMA они менее выражены и более приятны на слух.

Звук WMA лучше, чем в AAC, но сильно не дотягивает по качеству до звука полученного в mp3 pro. Высокочастотные составляющие в WMA потеряли свою мелодичность и иногда слышны резкие, неприятные призвуки на верхних частотах.

По сравнению со всеми описанными выше кодеками, в mp3 pro звук остался очень мелодичным, хотя явно не дотягивает до звучания оригинала. Высокие частоты присутствуют в нормальном виде, голос также приятен на слух. Отсутствуют сильные искажения в высокочастотном диапазоне, наблюдающиеся в остальных кодеках. Есть небольшие потери в высокочастотных составляющих, так что некоторые инструменты звучат немного по-другому, чем в оригинале — и это основное отличие от исходного звука.

Сведем места занимаемые кодеками при оценки качества на слух в одну таблицу вместе с спектральными методами сравнения:

	VMC mp3	Lame mp3	mp3 pro	WMA	AAC	GSM 44kHz	GSM 22kHz
Оценка на слух	4	6	1	2	3	7	5
Итоговый балл спектрального тестирования	4,5	3	2,25	1,25	3,5		5,5

Таблица 9.1

Как видно из таблицы 9.1, результат тестирования по метрикам существенно отличается от результатов спектрального сравнения. Попробуем разобраться в причине наиболее существенных отличий:

- Первое место в оценке на слух занял mp3-pro, а при аналитическом тестировании - второе. Как уже отмечалось, это обуславливается тем, что Mp3-pro сохраняет не форму волны на высоких частотах, а только её энергетическую огибающую. Это очень эффективный метод представления высоких частот с позиции именно нашего восприятия, для которого известно, что на высоких частотах регистрируются только амплитуды. Поэтому, не удивительно, что при тестировании без учета данного феномена нашего слуха, mp3-pro занял только второе место. Отметим, что наиболее эффективный на сегодня универсальный низкобитрейтный кодек HE AAC (он же AAC Plus) использует ту же идею что и в MP3-Pro.
- Кодек Lame MP3 оказался хуже VMC mp3 на слух и лучше по совокупности спектральных тестов. Это можно объяснить разной психоакустикой и разными типами артефактов у разных MP3-кодеров, которые сложно сравнивать даже на слух.

9.2 Результаты прослушивания speech.wav

Так же как и в предыдущем случае, худший результат у кодека GSM, кодирующего сигнал на частоте дискретизации 44100 kHz. Звук хрипит на всех частотах, начиная от низких и заканчивая высокими.

Несколько лучший результат у того же GSM при кодировании файла с частотой дискретизации 22050 Hz. Тем не менее, на фоне голоса по-прежнему отчетливо слышен треск, который очень раздражает.

Значительно лучший результат у VMC mp3, но звук слишком глухой и слышны призвуки с металлическим оттенком.

Mp3pro в работе с речью показал себя хуже, чем в работе с музыкой, и выигрывает только у VMC mp3. Звук, тем не менее, достаточно приятен на слух, но слишком глухой и с небольшими призвуками.

Lame mp3 тоже выдал достаточно глухой звук, но, тем не менее, более звонкий, чем mp3pro. Также присутствуют небольшие артефакты в сред-нечастотном диапазоне в тот момент, когда слышен голос диктора.

Звук в WMA достаточно звонкий и четкий, есть небольшие присвисты на высоких частотах, которых нет у AAC. В целом, звук более звонкий, чем в AAC.

Результат работы MPEG-4 AAC наиболее похож на оригинальное звучание, но, тем не менее, отличия хорошо заметны. После AAC звук более глухой и несколько менее живой (более механический), чем оригинальный.

9.3 Выводы

Как итог проделанной 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, протестированных 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. Отмечу, что если кодек не сохраняет форму волны, то это совершенно не значит, что кодек некачественно сжимает файл, т.к., например, шум можно сохранить как волну сложной формы, а можно просто записать параметры его 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 его 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, но не можем определить кто лучше в общем случае. 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.

При необходимости тестирования большого объема звуковой информации, для выявления, например лучшего результата мы можем сделать автоматичную первоначальную выборку результатов на основе некоторых метрик, которые явно плохо сохранились, чтобы в дальнейшем не тратить на них время. Остальные же результаты обрабатывать более подробно, например, с помощью музыкальных экспертов и специальных программ, аналогичных ABX и ABCHR (ABC Hidden Reference).

Также данные метрики целесообразно использовать при разработке самих кодеков, поскольку они позволяют определить такие ошибки, как потеря энергии или локальные выбросы в форме волны.

Данное сравнение кодеков не претендует на полноту и однозначность выводов о качестве звучания того или иного кодека, а описывает возможности той или иной метрики при сравнении звуковой информации различного характера, а так же результаты тестирования нескольких представительных кодеков с помощью данных методов сравнения.

9.4 Планы

Для получения более адекватной картины аналитическо-визуальной картины аудио-перцептивному восприятию, необходимо вычислять более тонкие акустические свойства сигнала. В первую очередь к ним относятся оценки гармоничности, шумности, звуковые удары и атаки. Эти свойства сигнала очень существенны для человеческого восприятия, однако качественность их передачи невозможно оценить используя рассматриваемые тут спектральные метрики.

9.5 Благодарности

Авторы выражают особую признательность за помощь в проверке и подготовке данной статьи **Сергею Лукьянову**, а также Николаю Труничкину, Сергею Саблину, Артему Титаренко, Александру Паршину, Алексею Лукину и Александру Ратушняку.