Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ _____ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ_____

КАФЕДРА _____СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ_____

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К КУРСОВОЙ РАБОТЕ

НА ТЕМУ:

Эргономический сравнительный анализ алгоритмов сжатия для различных видов аудио и форматов

Студент _____ИУ5И-32М___ _____ _____**Сюй Хаоюй**_____
(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель курсовой работы _____ _____**Б.С. Горячкин**____
(Подпись, дата) (И.О.Фамилия)

2025

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ
Заведующий кафедрой ___ИУ5____
(Индекс)
_____ ___В.И. Терехов__
(И.О.Фамилия)
«_01_» ____сентября_____ 2025 г.

З А Д А Н И Е
на выполнение курсовой работы

по дисциплине ___ _Эргономический анализ систем обработки информации_____

Студент группы _ИУ5И-32М__

_____Сюй Хаоюй_____
(Фамилия, имя, отчество)

Тема курсовой работы __Эргономический сравнительный анализ алгоритмов сжатия для различных видов аудио и форматов_____

Направленность КР (учебная, исследовательская, практическая, производственная, др.)
_____УЧЕБНАЯ_____
Источник тематики (кафедра, предприятие, НИР) _____КАФЕДРА_____

График выполнения работы: 25% к ___ нед., 50% к ___ нед., 75% к __ нед., 100% к ___ нед.

Задание ___Выполнить экспериментальное сравнение алгоритмов аудиосжатия для_____ __музыкальных и речевых сигналов. Задать правила подготовки входных WAV и_____ __формирования условий кодирования. Рассчитать и представить объективные показатели__ __качества и эффективности сжатия и сформулировать рекомендации по выбору кодека и___ __битрейта._____

Оформление курсовой работы:

Расчетно-пояснительная записка на _*48*_ листах формата А4.

Дата выдачи задания «_01_» ____сентября_____ 2025 г.

Руководитель курсовой работы _____ **Б.С. Горячкин**
(Подпись, дата) (И.О.Фамилия)

Студент _____ **Сюй Хаоюй**
(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре

Реферат

Данный документ представляет собой расчетно-пояснительную записку к курсовой работе по дисциплине «Эргономический анализ систем обработки информации». Цель курсовой работы – экспериментально оценить влияние алгоритмов сжатия аудио на качество сигнала и эффективность передачи и сформулировать рекомендации по выбору кодека и битрейта.

Содержание

Техническое задание

НАИМЕНОВАНИЕ РАЗРАБОТКИ

Эргономический сравнительный анализ алгоритмов сжатия для различных видов аудио и форматов

ОСНОВАНИЕ ДЛЯ РАЗРАБОТКИ

Основанием для разработки является учебный план, утвержденный кафедрой ИУ5 МГТУ им. Н. Э. Баумана.

ИСПОЛНИТЕЛЬ

Студент группы ИУ5И-32М Сюй Хаоюй

НАЗНАЧЕНИЕ И ЦЕЛЬ РАЗРАБОТКИ

Выполнить экспериментальное исследование влияния алгоритмов и параметров аудиосжатия на объективные показатели качества аудиосигнала и эффективности передачи данных и на этой основе сформулировать эргономические рекомендации по выбору кодека и битрейта для музыкальных и речевых сигналов.

СОДЕРЖАНИЕ РАБОТЫ

Провести подготовку входных аудиоданных и привести записи к единому эталонному формату (48 кГц, 16 бит, моно, унификация длительности).

Сформировать условия эксперимента кодирования для MP3, AAC и Opus в диапазоне 48–256 кбит/с и выполнить кодирование/декодирование.

Рассчитать объективные метрики качества и эффективности сжатия и выполнить статистическую агрегацию результатов по группам.

Представить результаты в виде таблиц и графиков и сформулировать рекомендации по выбору кодека и битрейта с учётом различий между музыкальными и речевыми сигналами.

ТРЕБОВАНИЯ К ДОКУМЕНТАЦИИ

По окончании работы предъявляются следующие документы:

1. Техническое задание (ТЗ);
2. Расчетно-пояснительная записка (РПЗ). РПЗ содержит 35 - 50 листов формата А4;
3. Приложения (листы формата А4), содержащие графическую часть в объеме пяти листов формата А4.

Графическая часть должна содержать следующие материалы:

лист 1.

ПОРЯДОК КОНТРОЛЯ И ПРИЕМКИ

Прием работы осуществляется путем проверки соответствия выполненной работы пунктам технического задания.

Введение

Алгоритмы аудиокодирования предназначены для получения компактных цифровых представлений широкополосных аудиосигналов с целью их эффективного хранения и передачи. Основной задачей аудиосжатия является снижение требуемой скорости передачи данных при максимально возможном сохранении воспринимаемого качества звука. В идеальном случае декодированный сигнал должен быть неотличим от исходного для человеческого слуха, несмотря на существенное уменьшение объёма данных.

Переход к цифровому представлению звука, реализованный в системах первого поколения, таких как компакт-диск (CD) и цифровые магнитофоны (DAT), обеспечил высокую верность воспроизведения и широкий динамический диапазон. Однако эти преимущества сопровождались высокими скоростями передачи данных, обусловленными использованием импульсно-кодовой модуляции (PCM) с частотами дискретизации 44,1–48 кГц и разрядностью 16 бит. Для многих современных приложений, особенно сетевых и беспроводных, такие скорости оказываются избыточными или практически недопустимыми.

В ответ на данные ограничения были разработаны перцептивные алгоритмы сжатия с потерями, основанные на свойствах слухового восприятия человека. Кодеки второго поколения, такие как MP3 и AAC, позволили существенно снизить битрейт при сохранении приемлемого или близкого к «CD-качеству» уровня звучания. В более поздний период развитие интерактивных сетевых сервисов привело к появлению универсальных и масштабируемых кодеков, ориентированных как на речь, так и на музыку, одним из наиболее распространённых примеров которых является Opus.

Несмотря на широкое практическое использование различных алгоритмов аудиосжатия, выбор конкретного кодека и битрейта на практике часто осуществляется эмпирически. При этом объективная оценка влияния параметров сжатия на свойства аудиосигнала, значимые с точки зрения восприятия и инженерного анализа, представляет особый интерес в контексте эргономики

аудиосистем.

В данной работе рассматриваются пять распространённых форматов аудиокодирования: несжатое представление PCM (WAV), бездисципативное сжатие FLAC и три кодека с потерями — MP3, AAC и Opus. Исследование проводится на музыкальных и речевых сигналах с использованием объективных метрик, отражающих качество сигнала и эффективность сжатия. Анализируются сегментное отношение сигнал/шум, спектральные характеристики и коэффициент сжатия при различных целевых битрейтах. Такой подход позволяет количественно сравнить алгоритмы аудиосжатия и сформулировать обоснованные рекомендации по выбору параметров кодирования с учётом требований к передаче и восприятию аудиоинформации.

Понятийный базис

PCM (Pulse Code Modulation, импульсно-кодовая модуляция) — это базовый способ цифрового представления звукового сигнала, при котором непрерывная аналоговая волна преобразуется в последовательность дискретных отсчётов $x[n]$. Формирование PCM включает: дискретизацию с частотой f_s, определяющей временное разрешение и верхнюю границу представимого спектра, и квантование амплитуды с разрядностью N_{bit}, задающей точность представления уровня сигнала. Для линейного PCM эффективный битрейт однозначно определяется параметрами дискретизации и числом каналов: $R_{\text{PCM}} = f_s \cdot N_{\text{bit}} \cdot N_{\text{ch}}$.

Lossless-кодирование обеспечивает точное восстановление исходных отсчётов PCM: x[n] = x̂[n] для всех n. Типичный пример — FLAC. В таких системах потери качества отсутствуют, а выигрыш достигается за счёт статистической избыточности в сигнале и энтропийного кодирования.

Lossy-кодирование допускает отличие восстановленного сигнала x̂[n] от исходного x[n] ради снижения битрейта. Классические lossy-кодеки используют психоакустические принципы и распределение бит по спектральным компонентам. Теоретически компромисс “битрейт–искажение” описывается функцией rate–distortion R(D).

Частота дискретизации f_s показывает, сколько дискретных отсчётов берётся с аналоговой звуковой волны в секунду на каждый канал, и обычно выражается в кГц. Интуитивно более высокая частота дискретизации задаёт более «тонкую» временную сетку и позволяет цифровому сигналу точнее описывать быстрые изменения формы волны. В частотной области частота дискретизации определяет верхнюю границу частот, которые могут быть представлены без неоднозначности.

Разрядность квантования n_b задаёт точность представления амплитуды каждого отсчёта. Чем больше бит, тем больше дискретных уровней доступно для кодирования амплитуды, тем меньше ошибка округления и ниже шум, вносимый

квантованием. На практике разрядность тесно связана с достижимым динамическим диапазоном и запасом по уровню (headroom) при записи и обработке. В потребительских форматах исторически распространено 16-битное PCM как базовый уровень, тогда как 24-битное PCM широко применяется в производстве, чтобы лучше сохранять точность при монтаже, сведении и другой обработке. В контексте сжатия разрядность важна потому, что она определяет разрешение исходного PCM-представления и влияет на базовый объём данных, относительно которого оценивается эффективность сжатия.

Битрейт ***R*** показывает, сколько бит в секунду используется для представления аудио после кодирования, и напрямую определяет затраты на хранение и передачу. Для кодеков с потерями битрейт обычно является основным параметром управления: более низкие значения дают более сильное сжатие, но повышают риск слышимых артефактов, тогда как более высокие значения, как правило, улучшают качество ценой увеличения размера данных. Для кодеков без потерь достигнутый битрейт обычно не задаётся как жёсткая цель тем же образом; он зависит от свойств аудиоматериала и настроек кодера, при этом декодированные отсчёты совпадают с исходными.

С теоретической точки зрения Lossless-кодирование и Lossy-кодирование источникового кодирования подчиняются различным информационно-теоретическим ограничениям. Для бездисципативного (или noiseless) кодирования теорема Шеннона об источниковом кодировании утверждает, что минимально достижимая средняя длина кода на символ ограничена снизу энтропией источника. Для дискретного безпамятного источника Xс алфавитом $\mathcal{X}$и функцией распределения вероятностей $p(x)$энтропия определяется как

$$H(X) = -\sum_{x \in \mathcal{X}} p(x)\ \log_2 p(x). \qquad (1)$$

Здесь X— случайная величина, представляющая один символ источника, $\mathcal{X}$ — множество всех возможных значений символов, $p(x)$ — вероятность

наблюдения символа x, а $H(X)$ измеряет среднее количество информации на символ в битах. На практике энтропийные методы кодирования, такие как кодирование Хаффмана или арифметическое кодирование, позволяют приближаться к этой границе для достаточно длинных последовательностей.

В случае кодирования с потерями теория скорости–искажения Шеннона характеризует оптимальный компромисс между битовой скоростью и искажением посредством функции скорости–искажения $R(D)$, которая задаёт минимальную среднюю скорость передачи данных, необходимую для представления источника при условии, что математическое ожидание искажения не превышает заданного уровня D. Для заданной покомпонентной меры искажений $d(x, \hat{x})$функция скорости–искажения определяется как

$$R(D) = \min_{p(\hat{x}|x):\, \mathbb{E}[d(X,\hat{X})] \le D} I(X;\hat{X}), \qquad (2)$$

где $\hat{X}$— случайная величина, представляющая восстановленный символ, $p(\hat{x} \mid x)$ — условное распределение, моделирующее процесс восстановления, $d(x, \hat{x})$ — неотрицательная функция, характеризующая искажение при представлении xчерез $\hat{x}$, $\mathbb{E}[\cdot]$обозначает математическое ожидание, а $I(X;\hat{X})$— взаимная информация между X и $\hat{X}$. В отличие от бездисципативного кодирования, в этом случае допускается отклонение восстановленного сигнала от исходного, что принципиально расширяет возможности снижения битрейта.

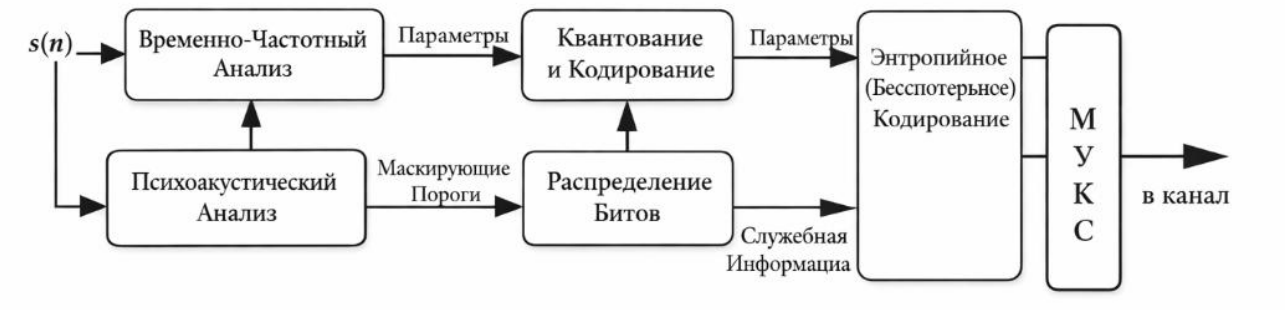

Рис.1. Типовая структура перцептивного аудиокодера

Современные аудиокодеки с потерями могут рассматриваться как практические реализации данного подхода, в которых снижение битовой скорости достигается за счёт контролируемого введения искажений. Конкретная интерпретация этих искажений и их влияние на воспринимаемое качество сигнала рассматриваются далее с позиций эргономики слухового восприятия.

Эргономичность

Теорема Найквиста–Шеннона о дискретизации гарантирует, что при условии, что частота дискретизации не менее чем вдвое превышает полосу сигнала, исходный непрерывный по времени аудиосигнал в принципе может быть идеально восстановлен по своим дискретным отсчётам. Если полосоограниченный сигнал имеет максимальную частоту B Гц и дискретизируется с частотой f_s, то идеальное восстановление возможно, когда

$$f_s \geq 2B, \qquad (3)$$

где f_s— частота дискретизации (в Гц), а B— максимальная частота (в Гц), присутствующая в спектре сигнала. Величина $f_s/2$ называется частотой Найквиста. В контексте аудио частота дискретизации 44,1 кГц обеспечивает теоретическую полосу частот примерно до 22 кГц, покрывая номинальный диапазон слышимости человека. При соблюдении этого условия сжатие выполняется над корректным дискретным представлением исходного аналогового сигнала, и различие между бездисципативным и с потерями кодированием относится исключительно к цифровой области.

Чтобы связать эти теоретические соображения с практическими аудиоформатами, удобно выразить битовую скорость несжатого линейного PCM через параметры дискретизации. Для PCM-сигнала с частотой дискретизации f_s(в Гц), разрядностью n_bбит на выборку и числом каналов Cбитовая скорость R_{PCM}(в бит/с) задаётся формулой

$$R_{\mathrm{PCM}} = f_s\, n_b\, C. \qquad (4)$$

Здесь n_bобозначает количество бит, используемых для квантования одной выборки на одном канале, а Cобычно равно 1 для моно и 2 для стереосигнала.

В качестве иллюстрации рассмотрим стереофонический сигнал «CD-качества», для которого $f_s = 44{,}100\text{Гц}$, $n_b = 16\text{бит}$, $C = 2$. Подстановка этих значений в (4) даёт

$$\begin{aligned} R_{\text{PCM}} &= 44{,}100 \times 16 \times 2 \\ &= 44{,}100 \times 32 \\ &= 1{,}411{,}200 \text{ бит/с} \\ &\approx 1{,}41 \text{ Мбит/с.} \end{aligned} \qquad (5)$$

Аналогично, для стереосигнала с частотой 48 кГц и той же разрядностью получаем

$$\begin{aligned} R_{\text{PCM}} &= 48{,}000 \times 16 \times 2 \\ &= 48{,}000 \times 32 \\ &= 1{,}536{,}000 \text{ бит/с} \\ &\approx 1{,}54 \text{ Мбит/с.} \end{aligned} \qquad (6)$$

Эти значения соответствуют типичным несжатым скоростям передачи данных, обычно приводимым для систем CD и DAT.

Условие теоремы Найквиста–Шеннона позволяет связать диапазон слышимости человека с подходящими частотами дискретизации аудио. Если номинальный диапазон слышимости моделировать как простирающийся до примерно $B = 20{,}000\text{Гц}$, то из (7) следует требование

$$f_s \geq 2B = 40{,}000 \text{ Гц.} \qquad (7)$$

Любая частота дискретизации не ниже 40 кГц, таким образом, удовлетворяет условию идеального восстановления 20-килогерцового полосоограниченного сигнала. Широко используемые аудиочастоты 44,1 кГц и 48 кГц обе удовлетворяют

$$44{,}100 \text{ Гц} \geq 40{,}000 \text{ Гц и } 48{,}000 \text{ Гц} \geq 40{,}000 \text{ Гц,} \qquad (8)$$

обеспечивая дополнительный запас для проектирования фильтров и практической реализации.

Для схем как бездисципативного, так и с потерями кодирования удобно использовать понятие «бит на выборку». Пусть задана кодовая скорость R (в бит/с), частота дискретизации f_s и число каналов C. Тогда среднее количество бит на выборку на канал, обозначаемое b_{ch}, равно

$$b_{\text{ch}} = \frac{R}{f_s\, C}. \qquad (9)$$

Здесь b_{ch} представляет среднее число бит, используемых для описания одной выборки в одном канале после сжатия. Для несжатого 16-битного PCM, используя (4) и (9) с $R = R_{\text{PCM}} = f_s n_b C$, немедленно получаем $b_{\text{ch}} = n_b = 16$ бит на выборку на канал, как и ожидается. Для стереофонического MP3-файла со скоростью 128 кбит/с при 44,1 кГц битовая скорость равна $R = 128{,}000$ бит/с, и

$$\begin{aligned} b_{\text{ch}} &= \frac{128{,}000}{44{,}100 \times 2} \\ &= \frac{128{,}000}{88{,}200} \\ &\approx 1{,}45 \text{ бит/выборку/канал,} \end{aligned} \qquad (10)$$

что более чем на порядок меньше, чем 16 бит на выборку, используемые в исходном PCM-представлении. Этот расчёт с помощью (9) и (10) наглядно показывает степень сжатия, достигаемую современными кодеками с потерями в терминах бит на выборку.

Наконец, энтропийный подход задаёт теоретическую верхнюю границу сжимаемости PCM-сигналов. Если 16-битную линейную PCM-выборку смоделировать как независимую равномерно распределённую случайную величину на множестве из 2^{16} квантов, то соответствующая энтропия будет

равна

$$H_{\max} = \log_2(2^{16}) = 16 \text{ бит/выборку.} \qquad (11)$$

Это значение представляет собой максимальную возможную энтропию на выборку при такой модели и, следовательно, жёсткую нижнюю границу для средней длины кода любого бездисципативного кодирования. Однако реальные аудиосигналы обладают выраженными временными и спектральными корреляциями, что объясняет принципиальную возможность существенного снижения средней битовой скорости по сравнению с линейным PCM при сохранении точного восстановления сигнала.

В оставшейся части данной работы мы сосредоточимся на пяти репрезентативных форматах аудиокодирования, актуальных для современных рабочих процессов в области музыки и вещания: линейном PCM в контейнере WAV, бездисципативном кодеке FLAC, а также перцептивных кодеках с потерями MP3, AAC и Opus.

Методы исследования

Для полноты методологического обоснования отметим, что в аудиоинженерии применяются перцептивно-ориентированные объективные методы оценки, в которых сравнение выполняется через промежуточные представления, моделирующие ранние стадии слухового восприятия. Типовая схема такого подхода включает модель периферического слуха, предобработку паттернов возбуждения и последующее агрегирование признаков для вычисления итоговой меры качества (рис.2). В настоящей работе данная схема используется как концептуальный контекст; далее применяются интерпретируемые объективные метрики, вычисляемые по сигналам после приведения всех вариантов к единому PCM-представлению.

Рис.2. Обобщённая схема перцептивно-ориентированной объективной оценки качества

В данном исследовании используется экспериментальная схема «единый исходный аудиосигнал — кодирование в несколько форматов — декодирование/воспроизведение — извлечение объективных параметров». Все

тестовые аудиофайлы в несжатом формате WAV применяются в качестве опорного сигнала, после чего выполняется перекодирование с получением пяти версий: FLAC (сжатие без потерь), MP3, AAC и Opus (сжатие с потерями). Далее все сжатые файлы приводятся к единому представлению в виде PCM-волны посредством декодирования, чтобы обеспечить сопоставимость спектральных и временных показателей в одной и той же области анализа. В работе используются только объективные количественно измеряемые метрики; субъективные прослушивания не проводятся. Для контроля переменных и обеспечения корректности сравнения все операции кодирования и анализа выполняются при одинаковых значениях частоты дискретизации, числа каналов и длительности. Возможные эффекты, вносимые кодером, такие как добавление заполнения и начальная задержка, учитываются с помощью процедур выравнивания сигналов и их обрезки.

Цель исследования — определить, какие форматы и режимы кодирования предпочтительны для различных классов аудиоинформации при ограничениях на размер файла или пропускную способность канала. При сопоставлении режимов кодирования учитывается базовый компромисс «битрейт–размер–искажения»: уменьшение целевого битрейта сопровождается сокращением размера файла и, как правило, ростом расхождения между декодированным и исходным сигналом по объективным метрикам, тогда как увеличение битрейта повышает близость к эталону ценой снижения эффективности сжатия. Для кодеков с потерями характерно замедление улучшений при переходе в область высоких битрейтов, когда дальнейшее увеличение пропускной способности приводит к относительно небольшим изменениям метрик при продолжающемся уменьшении коэффициента сжатия. Существенным фактором является класс исходного материала: речевые сигналы с выраженной фразовой сегментацией и более простой спектральной организацией обычно достигают стабильных значений метрик при меньших битрейтах, тогда как музыкальные сигналы со сложной спектрально-временной структурой, широкополосными компонентами и большим числом транзиентов требуют более высоких режимов для сохранения

спектрального баланса и ограничения искажений. Эти соображения определяют логику выбора диапазона битрейтов и состава метрик, а также основу для последующей интерпретации сравнительных результатов по форматам WAV/FLAC/MP3/AAC/Opus на различных типах аудиоматериала.

Описание алгоритма

Обработка, кодирование и вычисление метрик выполнялись на локальной рабочей станции под Microsoft Windows 10 в среде Anaconda/conda с отдельным окружением audio и интерпретатором Python 3.10. Основная логика эксперимента реализована на Python; преобразование форматов и декодирование выполнялись внешним набором утилит FFmpeg/ffprobe, установленным локально и вызываемым из Python через командную строку.

Фиксированные настройки кодирования для сравнения:

- целевые битрейты для MP3/AAC/Opus: 48, 64, 96, 128, 192, 256 kbps;
- фактический битрейт контролировался через ffprobe (поле bit_rate), а при отсутствии — оценивался по размеру файла и длительности.

Алгоритмический конвейер (pipeline):

1. Препроцессинг входных WAV: приведение всех записей к единому эталону WAV (PCM), 48 000 Гц, 16 бит, моно, с унификацией длительности (обрезка по центру до 25,0 с; при меньшей длительности обрезка не выполняется).
2. Кодирование эталона в FLAC и в три формата с потерями (MP3/AAC/Opus) при заданных целевых битрейтах.
3. Декодирование каждого сжатого файла обратно в WAV (48 кГц, моно, 16 бит) для корректного сопоставления с эталоном на уровне временного сигнала.
4. Вычисление метрик: коэффициент сжатия (по размеру файла), спектральные показатели (в том числе roll-off f95 и среднее значение спектрального центроида), высокочастотная доля энергии (HF_Ratio при пороге 8 кГц), а также сегментный SNR (SNRseg) после выравнивания по задержке и подгонки усиления.

5. Пакетная обработка нескольких файлов внутри каждой группы (n=5) и агрегация результатов: вычисление средних значений и стандартных отклонений по условиям (Codec × Bitrate), сохранение таблиц CSV и построение графиков «среднее ± 1 std».

Объекты анализа

В данной главе представлены пять аудиоформатов, исследуемых в работе: WAV (линейный PCM), FLAC, MP3, AAC и Opus. Эти форматы отражают три принципиально разные парадигмы кодирования: (i) несжатое представление формы сигнала (линейный PCM), (ii) сжатие без потерь с идеальным восстановлением отсчётов и (iii) перцептивное кодирование с потерями, в котором точность формы волны обменивается на снижение битрейта. Для каждого формата кратко рассматриваются принцип кодирования, типичные параметры, распространённые области применения, а также возможные артефакты и искажения. Это обеспечивает необходимую техническую основу для интерпретации объективных измерений, представленных в экспериментальной части.

WAV (линейный PCM)

WAV — широко используемый контейнер для несжатого линейного PCM-аудио. В линейном PCM звуковая волна представляется непосредственно дискретными отсчётами с фиксированной разрядностью квантования. Поэтому битрейт определяется частотой дискретизации, разрядностью и числом каналов, а не выбирается как целевой параметр кодера.

В данной работе WAV/PCM используется как эталонное представление, относительно которого сравниваются все сжатые форматы. Поскольку WAV/PCM сохраняет форму волны без потерь кодирования, любые отклонения в спектральной полосе, гармонической структуре или SNR для сжатых форматов интерпретируются относительно этого базового сигнала.

WAV/PCM чаще всего используется там, где требуется точное представление формы волны и ограничения по хранению/передаче не являются критичными: профессиональная запись и пост-продакшн, редактирование и судебно-акустический анализ, архивирование, а также исследовательские датасеты и вычислительные пайплайны обработки сигналов, где важны простота, совместимость и отсутствие зависимости от поведения кодека.

FLAC (сжатие без потерь)

FLAC (Free Lossless Audio Codec) — формат сжатия без потерь, предназначенный для уменьшения размера файлов при сохранении идеального восстановления отсчётов. В отличие от кодеков с потерями, FLAC не удаляет «менее важную» с точки зрения слуха информацию, а использует статистическую избыточность PCM-сигнала, достигая сжатия без ухудшения качества.

Поскольку FLAC является lossless, декодирование FLAC приводит к PCM-сигналу, побитово идентичному исходному. Следовательно, объективные метрики точности на декодированном сигнале в идеале не должны отражать искажения кодека, а лишь возможные эффекты вычислительного тракта анализа.

На практике достигаемый битрейт FLAC зависит от свойств сигнала. Поэтому FLAC полезен для оценки того, насколько сильно типичный музыкальный и вещательный контент содержит избыточность, при сохранении требования идеального восстановления.

FLAC широко используется для распространения музыки без потерь и хранения личных коллекций, когда важна экономия места без компромисса по точности сигнала. Он распространён в hi-fi коллекциях, долгосрочном архивировании, а также в обмене «мастер-копиями» между системами производства/лейбла/архива как компромисс между размером и строгим сохранением качества.

MP3 (MPEG-1/2 Layer III)

MP3 — один из наиболее распространённых кодеков с потерями, стандартизированный в семействе MPEG Audio. Снижение битрейта достигается за счёт перцептивных принципов: компоненты сигнала, которые, вероятно, будут замаскированы более сильными компонентами, кодируются с меньшей точностью или частично удаляются при минимизации субъективно воспринимаемого ухудшения.

MP3 обычно сочетает преобразовательное представление (для концентрации энергии по частотам), психоакустические модели маскирования и алгоритмы распределения бит. На практике MP3 задаётся целевым битрейтом (CBR) или параметром качества, который приводит к переменному битрейту (VBR). Несмотря на то, что эффективность MP3 в среднем ниже современных кодеков, он остаётся важным из-за широкой совместимости.

Типичные артефакты MP3 на низких битрейтах включают спад высоких частот, пре-эхо на транзиентах, «свистящие/вихревые» шумовые компоненты и ухудшение стереообраза. Выраженность эффектов зависит от контента и параметров кодирования.

MP3 до сих пор часто встречается благодаря максимальной совместимости с устройствами и программным обеспечением. Он распространён в старых музыкальных библиотеках, загруженных аудиофайлах и пользовательском контенте. Хотя современные сервисы нередко используют более новые кодеки, MP3 продолжает применяться там, где критична совместимость воспроизведения, а также для быстрого обмена файлами и переносного прослушивания при умеренных битрейтах.

AAC (Advanced Audio Coding)

AAC — семейство перцептивных кодеков с потерями, разработанное для повышения эффективности по сравнению с более ранними стандартами. В потребительских сценариях чаще всего используется профиль AAC-LC (Low

Complexity) как универсальный вариант. Подобно MP3, AAC опирается на преобразовательное кодирование и психоакустическое моделирование, но обладает более гибким набором инструментов и обычно обеспечивает лучшую эффективность при сопоставимом качестве.

Во многих современных системах AAC выбирают из-за баланса качества, степени сжатия и поддержки аппаратными/программными декодерами. На умеренных битрейтах AAC нередко обеспечивает более широкую эффективную полосу и меньше артефактов, чем MP3, хотя результат зависит от реализации кодера и выбранных настроек.

AAC широко используется в современных цепочках распространения контента, особенно в стриминге и мобильной экосистеме. Он часто применяется в онлайн-видео и музыкальных сервисах благодаря хорошему качеству при сравнительно низких битрейтах и широкой аппаратной поддержке на смартфонах, планшетах и приставках. AAC также распространён в медиаконтейнерах (MP4/M4A) и встречается в вещательных и околовещательных рабочих процессах.

Эффект спектрального маскирования лежит в основе перцептивных аудиокодеков: сильные компоненты спектра повышают порог слышимости для соседних частот.

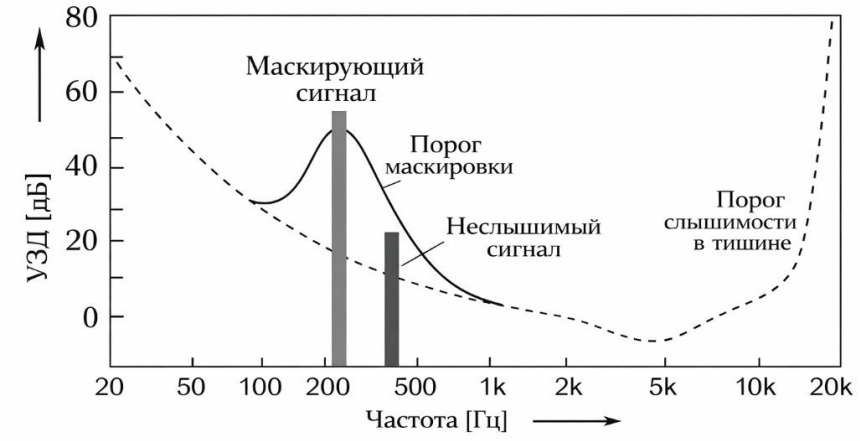

Рис.3. Иллюстрация спектрального маскирования: порог слышимости в тишине и порог маскирования при наличии маскирующего сигнала.

По рис.3 видно, что при наличии “маскирующего” сигнала порог слышимости локально повышается (masking threshold), и слабый соседний сигнал становится неразличимым. В кодеках MP3/AAC это используется для распределения бит и допускаемой квантовочной ошибки по частотным полосам.

Opus (сетевой кодек с потерями)

Opus — современный кодек с потерями, стандартизированный IETF и ориентированный на интерактивную, пакетную передачу по IP-сетям. Отличительная особенность Opus — способность эффективно кодировать как речь, так и музыку в широком диапазоне режимов, обеспечивая малую алгоритмическую задержку и устойчивость при меняющихся сетевых условиях. Он широко применяется в реальном времени: VoIP, видеоконференции, игровой голосовой чат и WebRTC-коммуникации в браузере.

С точки зрения кодирования Opus объединяет режимы, оптимизированные под разные типы сигналов (речеподобные vs. музыкальные), и способен адаптивно распределять биты. В отличие от «файловых» кодеков прошлого (например, MP3), Opus часто рассматривают в контексте требований к задержке и устойчивости доставки наряду с эффективностью сжатия.

Для музыкального и вещательного материала Opus может работать на средних и высоких битрейтах, приближаясь к «практически прозрачному» качеству, тогда как на низких битрейтах приоритетом становится разборчивость и непрерывность. Включение Opus в исследование позволяет напрямую сравнить «наследуемые» кодеки распространения (MP3), широко используемые современные кодеки распространения (AAC) и современный сетевой кодек (Opus) в рамках единого набора объективных метрик.

Opus особенно распространён в интерактивном интернет-аудио, где критичны малая задержка и устойчивость: VoIP, видеозвонки и онлайн-встречи, голосовой чат в играх, а также коммуникации на базе WebRTC. Он создавался именно для пакетной передачи и эффективно работает в широком диапазоне

битрейтов, поддерживая и речевые, и музыкальные сценарии. Благодаря своей распространённости в интерактивных сервисах Opus является репрезентативным примером современной «сетевой» аудиокомпрессии и дополняет MP3 и AAC, более характерные для традиционного файлового распространения и стриминга.

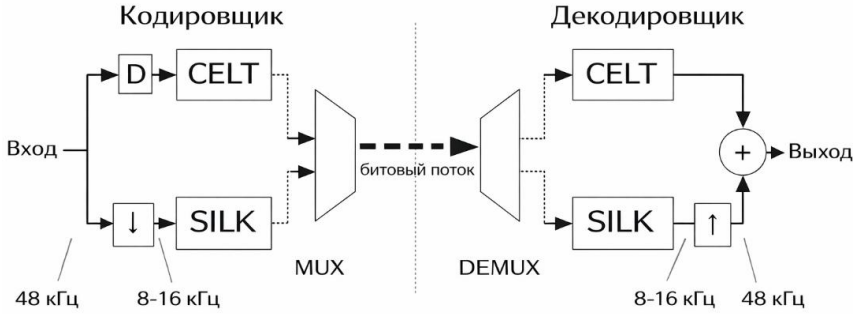

Рис.4 Краткий обзор кодека Opus

Ниже приведено обобщённое сопоставление рассматриваемых форматов/кодеков по характерной зависимости от режима кодирования, типичным сильным и слабым сторонам, а также областям применения.

Таб.1 Сравнение форматов WAV/FLAC/MP3/AAC/Opus: свойства и применение.

Формат /кодек	Зависимость от битрейта/режима	Преимущества	Ограничения	Типичные применения
WAV	Режим кодирования отсутствует; битрейт задаётся параметрами PCM	Эталонная точность, простота анализа, высокая совместимость	Большой размер, высокая требуемая полоса	Запись и монтаж, DSP-обработка, датасеты, архив исходников
FLAC	Режим битрейта отсутствует; степень сжатия зависит от структуры сигнала	Восстановление без потерь, экономия места относительно WAV	Сжатие ограничено избыточностью; больше файлов с потерями	Хранение без потерь, обмен «мастер»-копиями, архив
MP3	Снижение битрейта усиливает потери; рост битрейта улучшает близость к эталону	Максимальная совместимость, предсказуемые режимы	Менее эффективен, артефакты на низких битрейтах	Массовое распространение, старые библиотеки, обмен при требовании совместимости
AAC	Аналогично MP3; эффективность обычно выше при тех же битрейтах	Хорошее качество при умеренных битрейтах, широкая поддержка	Зависит от реализации; фактический битрейт может отличаться	Стриминг, мобильные устройства, MP4/M4A, вещательные цепочки
Opus	Адаптивные режимы, рассчитан на широкий диапазон и интерактивность	Универсален для речи/музыки, эффективен, низкая задержка	Чувствителен к настройкам; менее типичен для офлайн-архивов	VoIP, видеоконференции, WebRTC, игровой чат, real-time аудио

Основные характеристические параметры

Для обеспечения воспроизводимого и интерпретируемого сравнения пяти форматов (WAV/FLAC/MP3/AAC/Opus) без проведения субъективных прослушиваний в работе используются пять объективных показателей: коэффициент сжатия (CR), эффективная полоса f_{95}, доля высокочастотной энергии (HF Ratio), сегментный SNR (SNR_{seg}) и спектральный центроид (Spectral Centroid). Все показатели вычисляются по сигналам, полученным по схеме «исходный WAV → кодирование → декодирование обратно в PCM», после чего результаты агрегируются по типам аудио.

Пять выбранных показателей формируют взаимодополняемую систему: CR характеризует эффективность сжатия; f_{95}— энергетическую ширину спектра; HF Ratio — относительное сохранение высоких частот; SNR_{seg} — величину внесённых искажений в сравнении с эталоном; SC_{mean} — смещение спектрального «центра тяжести».

Коэффициент сжатия (Compression Ratio, CR)

CR характеризует эффективность сжатия и напрямую связан с затратами на хранение и передачу данных. Для одного и того же исходного аудио фиксируются размеры файлов $Size(WAV)$и $Size(Codec)$(в байтах), после чего вычисляется:

$$CR = \frac{Size(WAV)}{Size(Codec)} \qquad (12)$$

где $Codec \in$ {FLAC, MP3, AAC, Opus} . Чем выше CR, тем меньше итоговый файл и тем выше эффективность сжатия. Для кодеков с переменным битрейтом (особенно Opus) дополнительно фиксируется фактический средний битрейт.

Эффективная полоса по энергии (f_{95})

Эффективная полоса f_{95} количественно описывает, до какой частоты

сосредоточено 95% спектральной энергии. На основе среднего распределения мощности $P(f)$строится функция накопленной энергии:

$$C(f) = \frac{\sum_{0 \le \xi \le f} P(\xi)}{\sum_{0 \le \xi \le f_{max}} P(\xi)} \qquad (13)$$

Частота f_{95}определяется условием $C(f_{95}) = 0.95$. Уменьшение f_{95}обычно указывает на ослабление/обрезание высоких частот после сжатия.

Доля высокочастотной энергии (HF Ratio)

HF Ratio отражает, какая часть общей энергии приходится на высокочастотный диапазон.

Выбор 8 кГцпозволяет одновременно учитывать особенности речевого и музыкального материала. Доля высокочастотной энергии вычисляется как:

$$HF = \frac{\sum_{f_c \le f \le f_{max}} P(f)}{\sum_{0 \le f \le f_{max}} P(f)} \qquad (14)$$

где $f_{max} = 24$ кГц. Снижение HF обычно соответствует более сильному подавлению высоких частот в результате кодирования.

Сегментное отношение сигнал/шум (SNR_{seg})

Сегментный SNR оценивает уровень искажений, внесённых кодированием, через ошибку между опорным и восстановленным сигналом. После выравнивания по времени и амплитудного согласования определяется ошибка:

$$e(t) = x(t) - \alpha y(t) \qquad (15)$$

Далее сигнал разбивается на кадры, и для каждого кадра вычисляется SNR, после чего берётся среднее:

$$SNR_{seg} = \frac{1}{N}\sum_{n=1}^{N} 10\log_{10}\left(\frac{\sum x_n^2}{\sum e_n^2 + \epsilon}\right) \qquad (16)$$

где ϵ — малая константа для предотвращения деления на ноль. По сравнению с глобальным SNR сегментный вариант лучше отражает временную нестационарность аудио и чувствительнее к локальным искажениям.

Спектральный центроид (Spectral Centroid)

Спектральный центроид описывает «центр тяжести» спектра и часто интерпретируется как объективный показатель общей «яркости» (высокочастотного веса) сигнала. Для кадра n со спектральной мощностью $P_n(f)$центроид определяется как:

$$SC_n = \frac{\sum_f f \cdot P_n(f)}{\sum_f P_n(f)} \qquad (17)$$

В настоящей работе в отчёт включается **только среднее значение** спектрального центроида по всем кадрам:

$$SC_{mean} = \frac{1}{N}\sum_{n=1}^{N} S\,C_n \qquad (18)$$

Как правило, при подавлении высоких частот после сжатия SC_{mean} уменьшается; при появлении высокочастотных артефактов возможны отклонения, поэтому интерпретация проводится совместно с показателями f_{95}и HF Ratio.

В рамках исследования первоначально был сформирован расширенный набор из пяти параметров, отражающих различные аспекты изменения аудиосигнала при сжатии: коэффициент сжатия (CR), сегментное отношение сигнал/шум (SNRseg), среднее значение спектрального центроида, спектральный roll-off f95 и доля высокочастотной энергии (HF_Ratio).

В ходе экспериментального анализа было установлено, что метрики f95 и

HF_Ratio демонстрируют поведение, согласующееся с результатами по спектральному центроиду, однако обладают меньшей чувствительностью и более высокой вариативностью при сравнении различных типов сигналов. В связи с этим в разделе экспериментальных результатов основное внимание уделяется трем ключевым показателям (CR, SNRseg и спектральный центроид), тогда как f95 и HF_Ratio используются в качестве вспомогательных характеристик для подтверждения общих тенденций спектральных изменений.

Процесс обработки

Чтобы обеспечить согласованность внутренней обработки у разных кодеров, все входные WAV-файлы перед кодированием приводятся к единой частоте дискретизации и числу каналов. Этот шаг позволяет избежать дополнительных различий, возникающих из-за неявного ресэмплинга или смешивания каналов.

Для кодирования с потерями (MP3/AAC/Opus) применяется единая стратегия задания целевого битрейта (фиксированный битрейт или целевой средний битрейт), при этом для каждого выходного файла фиксируется фактический средний битрейт. Поскольку Opus часто использует VBR (переменный битрейт), фактическое значение может отличаться от заданного, поэтому «фактический средний битрейт» рассматривается как одна из ключевых контролируемых переменных при последующем сравнении.

Все сжатые файлы (FLAC/MP3/AAC/Opus) перед анализом декодируются в PCM-волновую форму. Поскольку некоторые кодеры могут вносить начальную задержку или добавлять заполнение в конце (padding), в работе выполняется выравнивание опорного и тестового сигналов (например, временное выравнивание на основе взаимной корреляции либо последующее вырезание общего эффективного интервала) для того, чтобы метрики, основанные на ошибке, вычислялись на одной и той же временной оси.

Единая предобработка

Унификация частоты дискретизации и каналов. Чтобы исключить влияние скрытого ресэмплинга и обеспечить сопоставимость между кодеками (в том числе Opus), все входные аудиофайлы приводятся к единой частоте дискретизации 48 кГц. Число каналов фиксируется единообразно для всего эксперимента. Для частоты 48 кГц верхняя граница анализа равна частоте Найквиста: $f_{max} = 24$ кГц.

Фиксированные параметры STFT. Для расчёта частотных признаков используется кратковременное преобразование Фурье (STFT) со следующими параметрами:

- оконная функция: Hann;
- длина окна: 2048 отсчётов (≈ 42.7 мс при 48 кГц);
- шаг (hop length): 512 отсчётов (≈ 10.7 мс);
- число точек БПФ: 2048.

Эти параметры обеспечивают разумный компромисс между временным и частотным разрешением и подходят как для музыкальных сигналов, так и для речевых/радиовещательных фрагментов. Для устойчивой оценки спектральных распределений мощность спектра усредняется по всем временным кадрам, формируя среднее распределение $P(f)$.

Выравнивание по времени и выделение общего эффективного интервала. Кодирование/декодирование может добавлять задержку и/или «паддинг» в начале/конце. Чтобы корректно вычислять показатели, основанные на разности сигналов (в частности SNR_{seg}), опорный сигнал $x(t)$ и тестовый сигнал $y(t)$выравниваются по времени: задержка оценивается по взаимной корреляции, после чего один из сигналов сдвигается. Далее используется только общий перекрывающийся интервал, исключая неинформативные добавленные области.

Подгонка усиления (амплитудное согласование) для SNR. Небольшие различия общего масштаба амплитуды после декодирования могут искусственно ухудшить SNR. Поэтому перед вычислением ошибки для SNR выполняется

глобальная подгонка коэффициента усиления α по методу наименьших квадратов:

$$\alpha = \arg\min_{\alpha} \| x - \alpha y \|^2 = \frac{\langle x, y \rangle}{\langle y, y \rangle}. \qquad (18)$$

В вычислениях SNR_{seg} используется $\alpha y(t)$. Для спектральных показателей (f_{95}, HF Ratio, спектральный центроид) эта подгонка не требуется.

Исходные данные и стандартизованный выход препроцессинга задавались следующим образом.

Исходные WAV-файлы:

- контейнер/формат: WAV
- частота дискретизации: 44,1 кГц / 48 кГц
- число каналов: 1 (моно) или 2 (стерео)
- разрядность: 16 бит или 24 бит
- длительность: не менее 30 с

Стандартизованный файл после препроцессинга:

- контейнер/формат: WAV (PCM)
- частота дискретизации: 48 000 Гц
- число каналов: 1 (моно); при стерео-входе выполняется сведение каналов в моно
- разрядность: 16 бит (PCM s16)
- временная унификация: обрезка по центру до 25,0 с;

Экспериментальные данные и описание выборки

Входной набор данных организован по стилям в каталоге <source_by_style> и включает пять групп: PopVocal, Classical, EDM, Rock и Broadcast. Для каждой музыкальной группы отобрано по пять WAV-файлов (n=5), что позволяет далее рассчитывать средние значения метрик и стандартные отклонения по однородным по типу материалам. Речевой набор Broadcast планируется получить посредством самостоятельной записи .

Для визуального контроля внутригрупповой сопоставимости и демонстрации различий между типами сигналов в временной области для каждой группы приводятся: (i) осциллограмма одного репрезентативного фрагмента и (ii) наложение осциллограмм всех пяти фрагментов данного стиля. Наложение используется для качественной оценки стабильности огибающей, плотности транзиентов и выраженности пауз внутри одной категории; количественные выводы в работе делаются на основе рассчитанных метрик и их статистической агрегации.

1. Pop Vocal (поп-музыка с вокалом)

Группа PopVocal сформирована из пяти китайскоязычных вокально-ориентированных поп-композиций, в которых ведущий вокал является доминирующим компонентом на фоне музыкального сопровождения. Для данного типа материала характерно сочетание квазипериодических участков (устойчивая основная частота и гармоники на протяжённых вокальных тонах) и непериодических шумоподобных компонентов (сибилянты/фрикативные согласные, шум дыхания), а также локальных транзиентов, связанных с ударными и согласными высокой энергии. Такая структура репрезентативна для анализа устойчивости кодирования в условиях смешанного сигнала «вокал + аккомпанемент», где важны как среднечастотные детали вокала, так и высокочастотные шумовые компоненты.

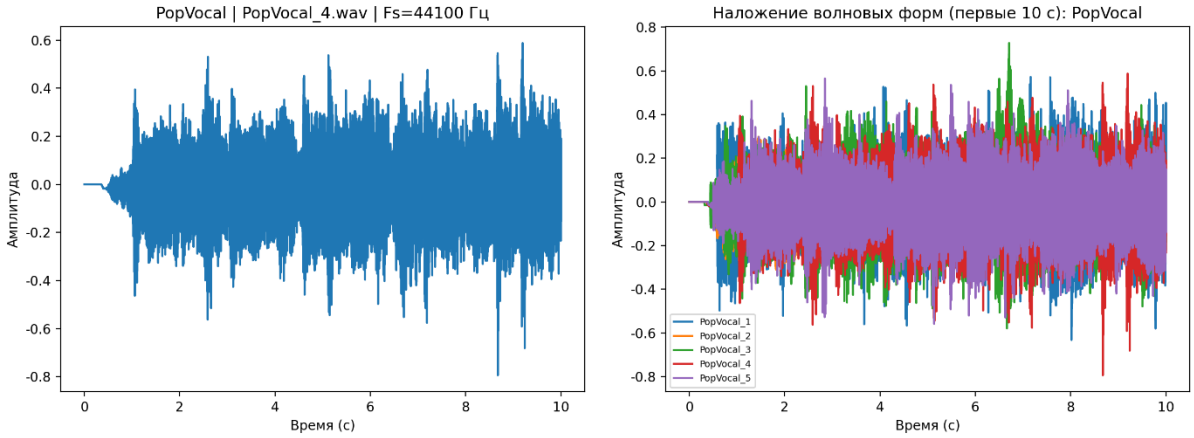

Рис.5 Осциллограмма отдельного образца и осциллограммы пяти образцов типа PopVocal, наложенные друг на друга

По рис.5 можно отметить выраженные фразовые изменения огибающей и наличие повторяющихся транзиентных событий; визуально это согласуется с ожидаемой структурой вокально-ориентированных поп-миксов.

2. Classical (классическая)

Группа Classical представлена пятью акустическими инструментальными композициями джазово-акустического характера (инструментальные стандарты/композиции без доминирующего электронного саунд-дизайна). Такие записи обычно обладают выраженной гармонической структурой, естественными атаками и затуханиями, а также более широким динамическим диапазоном по сравнению с плотными современными миксами. Во временной области наблюдается чередование участков низкой и высокой энергии, что делает данный тип сигналов чувствительным к артефактам кодирования на слабых уровнях и при сохранении тонких деталей атак.

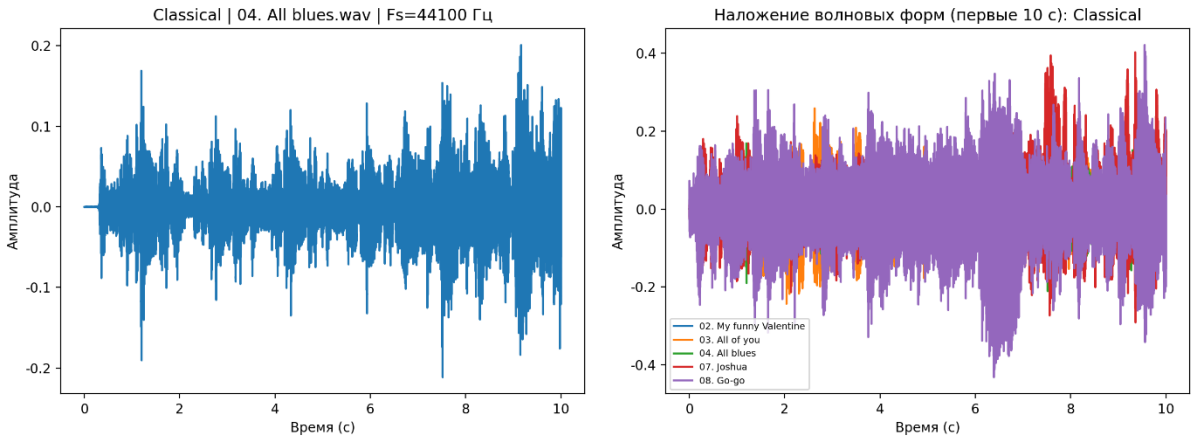

Рис.6. Осциллограмма отдельного образца и осциллограммы пяти образцов типа Classical, наложенные друг на друга

На рис.6 ожидаемо наблюдается более высокая вариативность огибающей между образцами при сохранении общей характеристики группы — выраженной динамической контрастности и инструментальных атак.

3. EDM (электронная музыка)

Группа EDM включает пять современных электронных/танцевальных композиций (с преобладанием синтезированных тембров и выраженным ритмическим рисунком). Для этого типа характерны длительные квазистационарные участки (пэды, басовые линии, синтезаторы) и регулярные транзиенты ударных, что в временной области проявляется устойчивой средней энергией и периодическими пиками огибающей. Данная группа используется как репрезентативный материал для анализа поведения кодеков на транзиентах и в условиях «плотного» спектрального заполнения, в том числе в высокочастотной области.

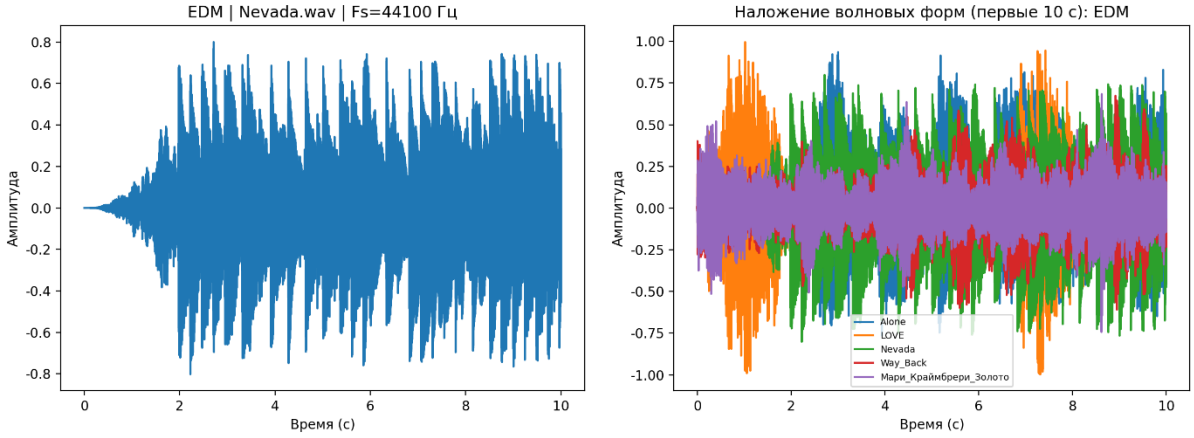

Рис.7. Осциллограмма отдельного образца и осциллограммы пяти образцов типа EDM, наложенные друг на друга

На рис.7 обычно отчётливо выделяются повторяющиеся транзиентные пики и сравнительно стабильная огибающая между ними, что соответствует типичной структуре EDM-миксов.

4. Rock (рок)

Группа Rock сформирована из пяти англоязычных рок/поп-рок композиций с плотным многодорожечным миксом. В подобных материалах спектр, как правило, насыщен (ударные, бас, гитары, вокал), а динамическая обработка (компрессия/лимитирование) приводит к более «заполненной» временной структуре. В осциллограмме это проявляется длительными интервалами повышенного уровня и высокой плотностью транзиентов. Группа включена как репрезентативный пример «нагруженного» сигнала, где компромиссы кодеков по распределению битов и маскированию проявляются наиболее заметно.

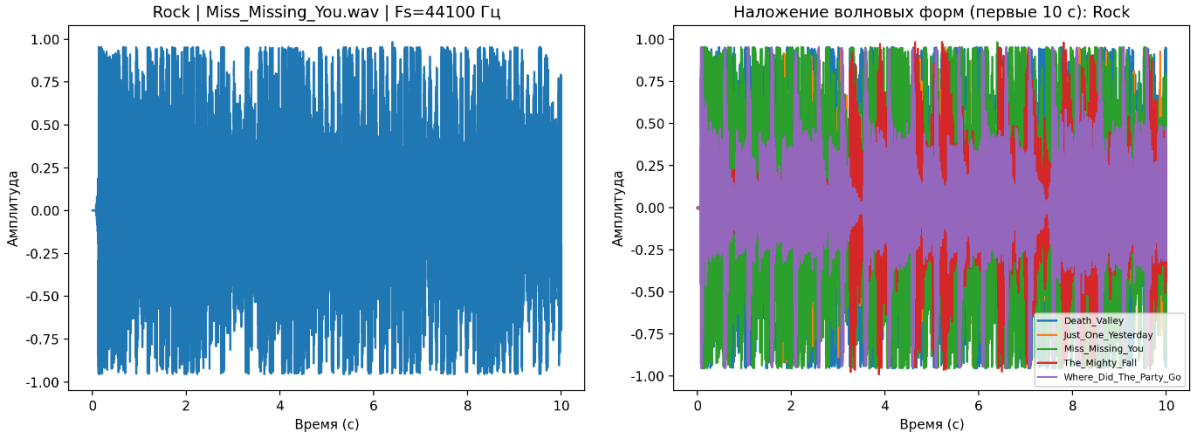

Рис.8. Осциллограмма отдельного образца и осциллограммы пяти образцов типа Rock, наложенные друг на друга

По рис.8 можно отметить высокую временную насыщенность: сравнительно небольшой разброс уровней огибающей при большом числе транзиентов, что типично для плотных рок.

5. Broadcast/News Speech (новости/радио)

Речевой материал получен путем самостоятельной записи с использованием профессионального микрофона Ulanzi V100, что обеспечивает контролируемость условий и воспроизводимость набора. Записи выполнены с параметрами дискретизации 48 000 Гц, 16 бит, стерео (2 канала). Набор Broadcast включает пять речевых фрагментов: два фрагмента новостного дикторского чтения на китайском языке, два фрагмента чтения на русском языке и один фрагмент англоязычного радиоформата. Речь включает чередование речевых сегментов и пауз, содержит участки с квазистационарной структурой (гласные) и кратковременные транзиенты (взрывные согласные), а также шумоподобные компоненты (фрикативные согласные). Осциллограмма обычно демонстрирует выраженную сегментацию на фразы и паузы, что важно для анализа устойчивости метрик и для интерпретации изменений высокочастотной энергии, связанной с согласными и шумовыми компонентами.

Данный тип выбран как контраст к музыкальным сигналам и как приближение к реальному сценарию передачи речи при ограниченной пропускной способности.

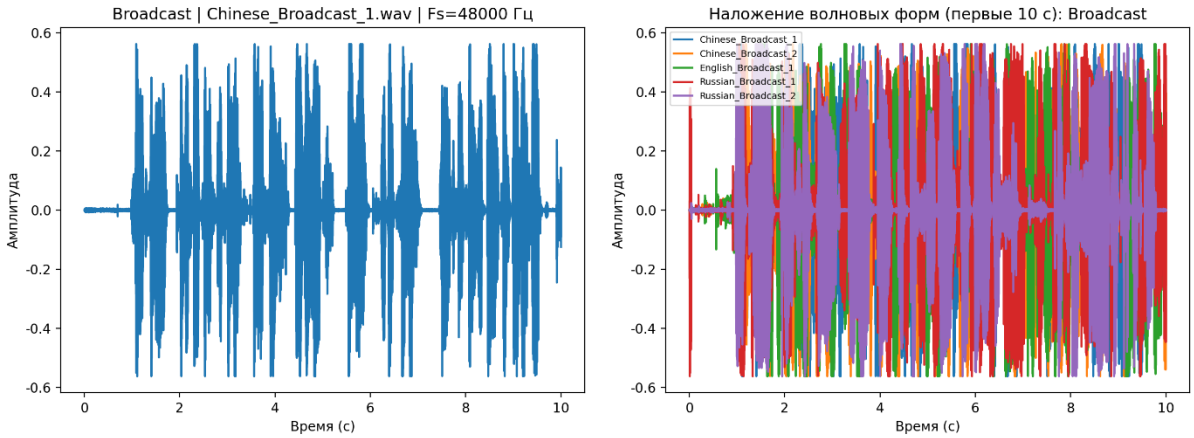

Рис.9. Осциллограмма отдельного образца и осциллограммы пяти образцов типа Broadcast, наложенные друг на друга

По рис.9. можно отметить выраженную речевую сегментацию и высокую транзиентную насыщенность: на фоне относительно низкоамплитудных участков (паузы и межфразовые интервалы) регулярно возникают короткие высокоэнергетические всплески, соответствующие артикуляционным атакам и согласным. При наложении пяти записей общий характер огибающей сохраняется (чередование активной речи и кратких пауз), тогда как различия между кривыми обусловлены индивидуальными фразовыми границами и темпом чтения, что является типичным для дикторской речи и важно для интерпретации метрик, чувствительных к транзиентам и высокочастотным шумоподобным компонентам.

В завершение описания выборки приведено краткое сопоставление классов аудиоматериала, использованных в эксперименте. Классы различаются по спектрально-временной организации . Эти различия важны для дальнейшего сопоставления поведения метрик при изменении битрейта. Таб.2 суммирует типичные особенности каждого класса и ожидаемую чувствительность к снижению битрейта

Таб.2 Обобщённая характеристика классов аудиоматериала и ожидаемая чувствительность к снижению битрейта

Класс	**Типичная спектрально-временная структура**	**Ключевые элементы сигнала**	**Ожидаемая чувствительность к снижению битрейта**
Pop Vocal	Умеренная плотность, вокал и аранжировка, среднее число транзиентов	Форманты/тональность вокала, умеренные ВЧ компоненты	Средняя
Classical	Высокая динамика, тонкие детали, длительные затухания	Слабые гармоники, хвосты реверберации	Средняя–повышенная
EDM	Высокая спектральная насыщенность, регулярные транзиенты, широкополосность	Ударные атаки, синтезированные ВЧ, плотный микс	Повышенная
Rock	Плотный микс, много транзиентов и маскирующих компонентов	Атаки ударных/струнных, широкий спектр	Повышенная
Broadcast	Сегментация на фразы/паузы, чередование стационарных и шумоподобных участков	Гласные (квазистационарные), согласные (транзиенты/шум)	Низкая–средняя

Экспериментальные результаты

В данной главе представлены результаты экспериментального исследования влияния алгоритмов аудиосжатия с потерями на объективные характеристики восстановленного сигнала и эффективность сжатия. Анализ выполнен для трёх кодеков с потерями — MP3, AAC и Opus при шести целевых битрейтах: 48, 64, 96, 128, 192 и 256 кбит/с.

Эксперименты проведены на пяти типах входных сигналов: PopVocal, Classical, EDM, Rock и Broadcast. Для каждого типа использовалось по пять аудиофайлов (n = 5). Все результаты представлены в виде средних значений и стандартных отклонений, что позволяет учитывать внутригрупповую вариативность материала.

1.Сегментный SNR (SNRseg)

Сегментное отношение сигнал/шум (SNRseg) использовалось как основная объективная метрика искажения сигнала после кодирования и декодирования. Метрика вычислялась после временного выравнивания и амплитудного согласования относительно эталонного WAV-сигнала.

Зависимости SNRseg от битрейта для различных типов аудиосигналов представлены на рис.10. Для всех рассмотренных типов наблюдается устойчивый рост SNRseg при увеличении битрейта. Наиболее интенсивное увеличение происходит в диапазоне 48–128 кбит/с, тогда как при переходе к 192–256 кбит/с рост замедляется, что указывает на приближение к области насыщения.

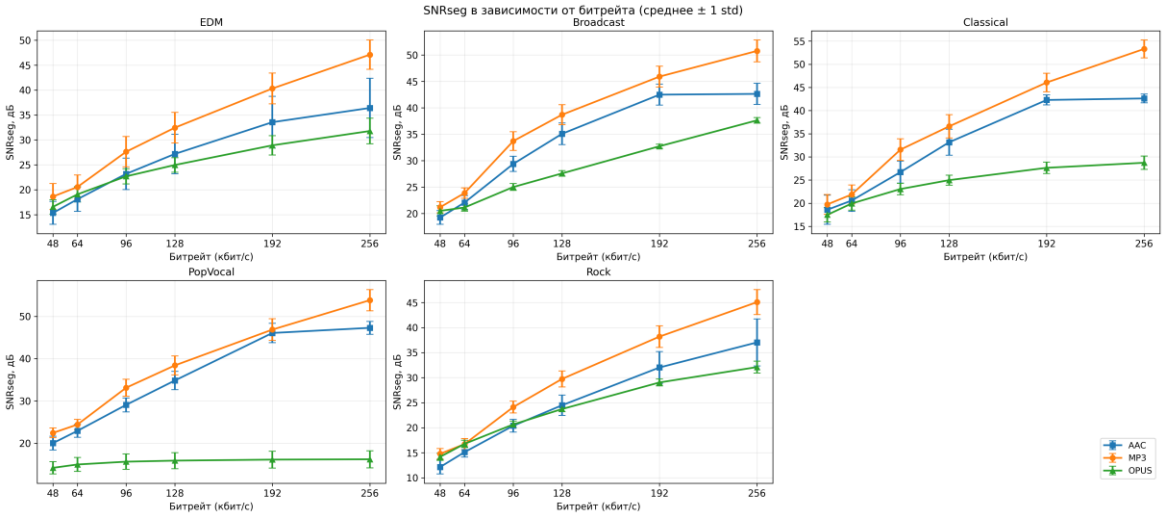

Рис.10. Зависимость SNRseg от битрейта для разных типов

Сравнение кодеков показывает устойчивую тенденцию: во всех типах сигналов MP3 демонстрирует наибольшие значения SNRseg, AAC занимает промежуточное положение, а Opus характеризуется более низкими значениями данной метрики. Различия особенно выражены для музыкальных сигналов с развитой гармонической структурой (PopVocal, Classical), где при 256 кбит/с значения SNRseg для MP3 превышают 50 дБ, тогда как для Opus остаются существенно ниже.

Следует отметить, что данное сравнение относится исключительно к метрике SNRseg, ориентированной на отклонение формы волны, и не отражает напрямую субъективную разборчивость или перцептивное качество, для оптимизации которых кодек Opus изначально проектировался.

Для речевого материала (Broadcast) рост SNRseg с увеличением битрейта также сохраняется, однако абсолютные значения ниже, чем для музыкальных сигналов, что связано с высокой долей транзиентов и шумоподобных компонентов речи.

Следует подчеркнуть, что SNRseg является сигнал-ориентированной объективной метрикой и не всегда напрямую коррелирует с субъективным восприятием качества, особенно для кодеков с различными психоакустическими моделями. Полученные различия отражают прежде всего степень отклонения формы волны восстановленного сигнала от эталона.

Для кодека Opus более низкие значения SNRseg не следует интерпретировать как однозначное ухудшение субъективного качества, поскольку данный кодек оптимизирован под психоакустические критерии и сетевые условия, а не под минимизацию временной ошибки формы волны.

2.Спектральный центроид

Среднее значение спектрального центроида использовалось для оценки изменения спектрального баланса и «яркости» сигнала после кодирования. Результаты представлены на рис. 11.

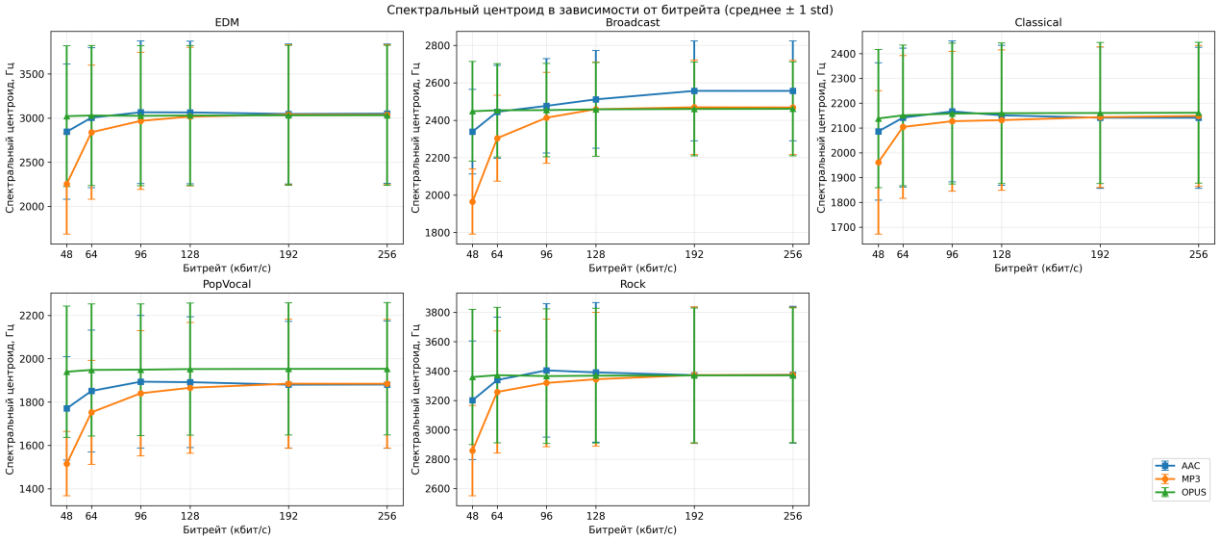

Рис. 11 Зависимость среднего спектрального центроида от битрейта для разных типов сигналов (среднее ± 1 std, n=5).

Для всех типов аудиосигналов наблюдается рост спектрального центроида при увеличении битрейта, наиболее выраженный в диапазоне 48–96 кбит/с. Начиная приблизительно с 128 кбит/с значения спектрального центроида стабилизируются, а различия между кодеками уменьшаются.

Абсолютные уровни спектрального центроида существенно различаются между типами сигналов. Минимальные значения характерны для PopVocal, более высокие — для Classical, а максимальные — для EDM и Rock, что отражает различия в спектральной плотности и насыщенности высокочастотными компонентами. Речевой материал (Broadcast) занимает промежуточное

положение и демонстрирует чёткую зависимость от битрейта.

На низких битрейтах в ряде случаев Opus демонстрирует несколько более высокие значения спектрального центроида по сравнению с MP3 и AAC, что может указывать на лучшее сохранение или перераспределение высокочастотной энергии. При увеличении битрейта различия между кодеками сглаживаются, и при 128–256 кбит/с все три кодека обеспечивают близкие значения спектрального центроида.

Наблюдаемые изменения спектрального центроида согласуются с поведением доли высокочастотной энергии (HF Ratio), рассчитанной на этапе метрик, что подтверждает корректность интерпретации спектральных изменений.

3.Коэффициент сжатия (CR)

Коэффициент сжатия CR определялся как отношение размера эталонного WAV-файла к размеру закодированного файла.

Кодеки с потерями (MP3/AAC/Opus):

Зависимости CR от целевого битрейта представлены на рис. 12.

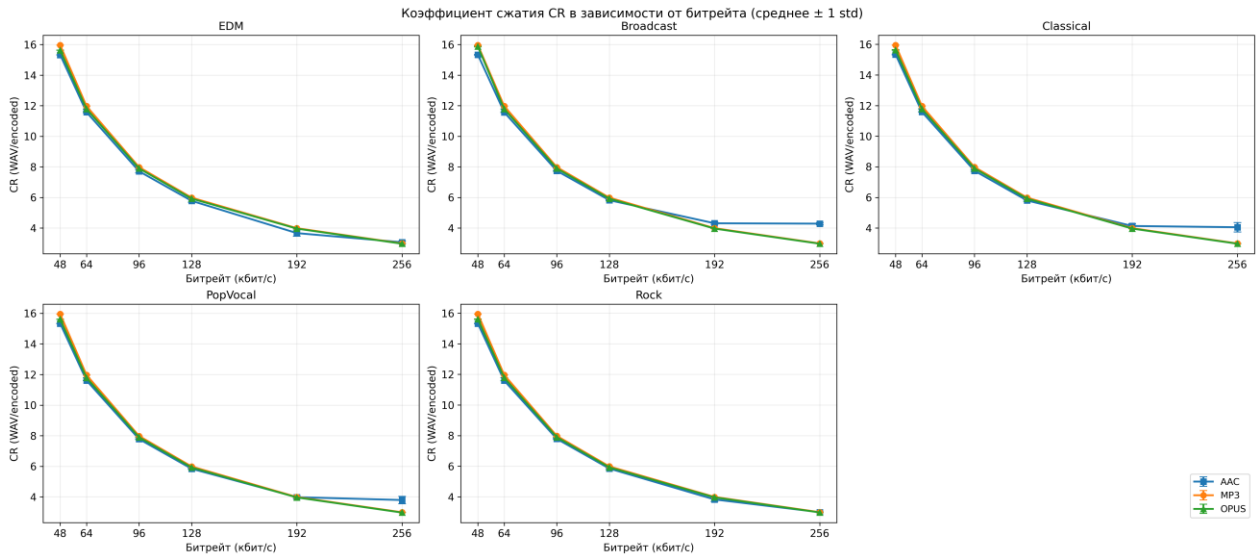

Рис.12 Зависимость коэффициента сжатия CR от битрейта для разных типов сигналов (среднее ± 1 std, n=5).

Для всех типов сигналов коэффициент сжатия монотонно уменьшается при увеличении битрейта. На низком битрейте 48 кбит/с CR достигает значений порядка 15–16, что соответствует высокой степени сжатия. При 96–128 кбит/с CR снижается до 6–8, а в диапазоне 192–256 кбит/с — до 3–4. Зависимость носит нелинейный характер с выраженным замедлением изменения на высоких битрейтах.

Сравнение кодеков показывает, что при низких и средних битрейтах различия между MP3, AAC и Opus по коэффициенту сжатия минимальны. Однако на высоких битрейтах (192–256 кбит/с) в ряде групп (PopVocal, Classical, Broadcast) кодек AAC демонстрирует несколько более высокие значения CR, что указывает на формирование меньших по размеру файлов при сопоставимом качестве кодирования. Для EDM и Rock различия между кодеками в этом диапазоне выражены слабо.

Следует отметить, что для кодека AAC при высоких целевых битрейтах фактический средний битрейт в ряде случаев отличается от номинального, что приводит к увеличению коэффициента сжатия по сравнению с MP3 и Opus при одинаково заданном целевом значении.

Сжатие без потерь (FLAC относительно WAV)

В отличие от кодеков с потерями, формат FLAC не предполагает задания целевого битрейта, поскольку реализует бездисципативное сжатие. В этом случае эффективность сжатия определяется исключительно статистическими свойствами сигнала. Для оценки была вычислена средняя величина коэффициента сжатия CR для FLAC относительно WAV отдельно для каждого типа сигналов.

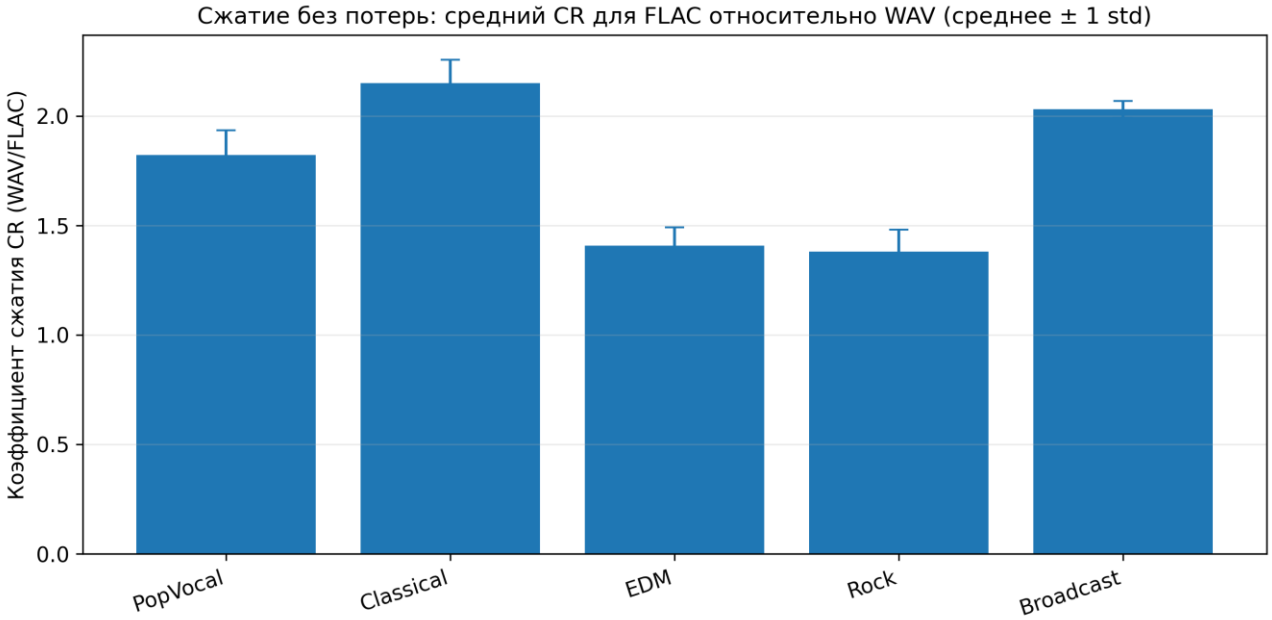

Рис.13 Средний коэффициент сжатия CR для FLAC относительно WAV по типам сигналов (среднее ± 1 std)

Таб.3 Коэффициент сжатия CR (WAV/FLAC) для различных типов сигналов (среднее ± std, n = 5)

Тип	n	CR (WAV/FLAC) (среднее)	CR (WAV/FLAC) (std)
PopVocal	5	1.821	0.1123
Classical	5	2.1491	0.1074
EDM	5	1.4069	0.0833
Rock	5	1.3797	0.1
Broadcast	5	2.0298	0.0391

Для бездисципативного сжатия коэффициент CR не зависит от целевого битрейта и определяется исключительно статистическими свойствами исходного сигнала. Поэтому для кодека FLAC коэффициент сжатия был усреднён по всем файлам внутри каждой группы аудиоматериала.

В таб.3 представлены средние значения коэффициента сжатия CR = size(WAV) / size(FLAC) и соответствующие стандартные отклонения, рассчитанные по выборке n = 5 файлов для каждого типа сигналов. Полученные результаты показывают, что эффективность бездисципативного сжатия существенно зависит от характера аудиоматериала. Наибольшие значения CR наблюдаются для сигналов типа Classical и Broadcast, что связано с высокой степенью коррелированности и сравнительно простой спектральной структурой. Минимальные значения CR характерны для сигналов EDM и Rock,

отличающихся высокой спектральной плотностью, насыщенностью высокочастотных компонентов и выраженной динамической сложностью. Таким образом, в отличие от кодеков с потерями, для FLAC коэффициент сжатия определяется не параметрами кодирования, а внутренними статистическими свойствами исходного сигнала. Малые значения стандартного отклонения во всех группах подтверждают устойчивость результатов и позволяют рассматривать средние значения CR как репрезентативные характеристики эффективности бездисципативного сжатия для соответствующих типов сигналов.

В сравнении с бездисципативным сжатием, перцептивные кодеки с потерями (MP3/AAC/Opus) достигают существенно больших коэффициентов сжатия за счёт допускаемых отклонений восстановленного сигнала от эталона. Это видно при сопоставлении характерных диапазонов CR: для FLAC значения CR определяются исключительно статистической избыточностью материала и остаются ограниченными (как правило, порядка 1.3–2.2 в зависимости от типа сигнала), тогда как при кодировании с потерями CR существенно варьирует и может достигать значений порядка 15–16 на низких битрейтах. Следовательно, прирост эффективности сжатия в случаях MP3/AAC/Opus обеспечивается не «лучшим» устранением избыточности, а переходом к компромиссу скорость–искажение, в рамках которого часть информации удаляется или квантуется более грубо. При этом различия CR между типами сигналов для FLAC отражают преимущественно энтропийные свойства исходного материала, тогда как для кодеков с потерями зависимость CR в первую очередь задаётся выбранным битрейтом и лишь вторично — особенностями спектрально-временной структуры сигнала.

4. Обобщение результатов

В настоящем разделе результаты экспериментального сравнения интерпретируются в контексте теоретических положений о компромиссе «качество–объём» при аудиокодировании и о зависимости чувствительности от

класса материала.

Полученные экспериментальные зависимости в целом подтверждают сформулированные в разделе «Методы исследования» предположения и позволяют уточнить их по классам материала и диапазонам битрейта (48–256 кбит/с)

В таб.4–6 приведены усреднённые значения сегментного отношения сигнал/шум (SNRseg), спектрального центроида и коэффициента сжатия CR для различных типов аудиосигналов и кодеков.

Таб.4 Сводные результаты при 64 кбит/с (Тип × Кодек): SNRseg, спектральный центроид, CR (среднее ± std).

Тип	**Кодек**	**SNRseg, дБ (среднее)**	**SNRseg, дБ (std)**	**Спектральный центроид, Гц (среднее)**	**Спектральный центроид, Гц (std)**	**Коэффициент сжатия CR (WAV/encoded) (среднее)**	**Коэффициент сжатия CR (WAV/encoded) (std)**
PopVocal	MP3	24.414	1.23	1752.4	240	11.9673	0
PopVocal	AAC	22.922	1.537	1851.2	281.4	11.6101	0.0082
PopVocal	OPUS	14.976	1.654	1948.2	304.7	11.7827	0.0019
Classical	MP3	21.879	2.036	2104.1	288.4	11.9617	0.0003
Classical	AAC	20.57	2.346	2141.6	280.1	11.5848	0.0107
Classical	OPUS	19.933	1.447	2150.9	283.6	11.7784	0.0026
EDM	MP3	20.548	2.424	2838.9	758.8	11.9709	0
EDM	AAC	18.094	2.455	3004.6	792.1	11.5788	0.0381
EDM	OPUS	19.069	1.545	3028.9	792.7	11.7864	0.0019
Rock	MP3	16.722	1.098	3257.8	416	11.9627	0.0004
Rock	AAC	15.118	0.988	3339.2	427.7	11.5915	0.0123
Rock	OPUS	16.774	0.738	3372.1	461.2	11.781	0.0007
Broadcast	MP3	23.879	0.98	2303.7	229.7	11.9709	0
Broadcast	AAC	22.082	1.319	2444.9	248.3	11.5712	0.0184
Broadcast	OPUS	21.15	0.681	2453.1	249.4	11.7854	0.0027

Таб.5 Сводные результаты при 128 кбит/с (Тип × Кодек): SNRseg, спектральный центроид, CR (среднее ± std).

Тип	Кодек	SNRseg, дБ (среднее)	SNRseg, дБ (std)	Спектральный центроид, Гц (среднее)	Спектральный центроид, Гц (std)	Коэффициент сжатия CR (WAV/encoded) (среднее)	Коэффициент сжатия CR (WAV/encoded) (std)
PopVocal	MP3	38.402	2.241	1865.8	302	5.9853	0
PopVocal	AAC	34.833	2.164	1891.7	301.8	5.8371	0.0331
PopVocal	OPUS	15.887	1.901	1952.1	304.6	5.9227	0.001
Classical	MP3	36.591	2.528	2131.7	283.4	5.984	0.0001
Classical	AAC	33.158	2.818	2151	282.7	5.7968	0.0505
Classical	OPUS	24.975	1.088	2159.5	284	5.9216	0.0012
EDM	MP3	32.459	3.059	3016.5	785.5	5.9861	0
EDM	AAC	27.188	3.974	3063.8	807.5	5.7777	0.0813
EDM	OPUS	24.969	1.446	3029	792	5.9238	0.001
Rock	MP3	29.738	1.611	3344.4	454.7	5.9842	0.0001
Rock	AAC	24.48	2.051	3390.9	475.1	5.8425	0.0531
Rock	OPUS	23.726	0.515	3368.6	459	5.9227	0.0002
Broadcast	MP3	38.702	1.886	2458.9	251.6	5.9861	0
Broadcast	AAC	35.091	2.062	2511.4	260.9	5.8197	0.0273
Broadcast	OPUS	27.598	0.546	2458	250.4	5.9232	0.0012

Таб.6 Сводные результаты при 256 кбит/с (Тип × Кодек): SNRseg, спектральный центроид, CR (среднее ± std).

Тип	Кодек	SNRseg, дБ (среднее)	SNRseg, дБ (std)	Спектральный центроид, Гц (среднее)	Спектральный центроид, Гц (std)	Коэффициент сжатия CR (WAV/encoded) (среднее)	Коэффициент сжатия CR (WAV/encoded) (std)
PopVocal	MP3	53.816	2.491	1884.3	296.9	2.993	0
PopVocal	AAC	47.281	1.541	1880.8	293.6	3.7949	0.2427
PopVocal	OPUS	16.207	2.012	1953.4	305.2	2.9739	0.0008
Classical	MP3	53.303	1.919	2148.5	283.8	2.9928	0
Classical	AAC	42.614	0.93	2141.2	284.5	4.0517	0.3052
Classical	OPUS	28.729	1.431	2161.5	284.4	2.9741	0.0005
EDM	MP3	47.116	2.933	3037.9	791.9	2.9932	0
EDM	AAC	36.417	5.944	3050.1	789	3.0714	0.0855
EDM	OPUS	31.787	2.569	3031.8	792.4	2.9743	0.0007
Rock	MP3	45.075	2.496	3374.3	460.9	2.9928	0
Rock	AAC	37.035	4.685	3375.4	466.2	2.9912	0.0246
Rock	OPUS	32.106	1.189	3371	459.2	2.9746	0.0002
Broadcast	MP3	50.769	2.062	2468.2	251.8	2.9932	0
Broadcast	AAC	42.642	1.995	2556.6	266.4	4.287	0.0763
Broadcast	OPUS	37.635	0.545	2460.8	250.8	2.9742	0.0011

По данным графиков и сводных таблиц можно сделать следующие выводы, подтверждающие компромисс «качество–объём» и зависимость результатов от типа материала:

1. Для кодеков с потерями коэффициент сжатия в первую очередь определяется целевым битрейтом и лишь в меньшей степени зависит от типа сигнала и конкретного алгоритма кодирования. При этом на высоких битрейтах (192–256 кбит/с) кодек AAC в ряде случаев формирует файлы меньшего размера, что приводит к несколько более высоким значениям CR по сравнению с MP3 и Opus. В случае бездисципативного кодирования FLAC коэффициент сжатия определяется статистическими и спектрально-временными свойствами исходного сигнала и демонстрирует выраженную зависимость от его типа.

2. Сегментное отношение сигнал/шум (SNRseg) устойчиво возрастает с увеличением битрейта для всех типов сигналов и всех рассмотренных кодеков с потерями. Во всех экспериментальных условиях MP3 демонстрирует наибольшие значения SNRseg, AAC занимает промежуточное положение, а Opus характеризуется более низкими значениями данной метрики, что связано с различиями в используемых психоакустических моделях и принципах кодирования.

3. Среднее значение спектрального центроида быстро приближается к эталонному уровню при увеличении битрейта до 96–128 кбит/с и далее изменяется незначительно. Это указывает на восстановление спектрального баланса сигнала при достижении достаточной пропускной способности канала и подтверждает целесообразность выбора данного диапазона битрейтов как практического компромисса между качеством и эффективностью сжатия.

4. Подтверждается зависимость чувствительности к снижению битрейта от класса аудиоинформации. Материалы с плотной спектральной структурой и большим числом транзиентов (EDM, рок) демонстрируют более выраженную деградацию объективных показателей в низких режимах 48–64 кбит/с, тогда как классика/акустика и речевой материал характеризуются более стабильной динамикой метрик при изменении битрейта.

Выводы

В работе выполнено экспериментальное сравнение пяти форматов представления аудио (WAV/FLAC/MP3/AAC/Opus) на пяти классах материала по объективным метрикам CR, SNRseg и среднему спектральному центроиду; для кодеков с потерями рассмотрен диапазон целевых битрейтов 48–256 кбит/с. Полученные зависимости подтверждают компромисс между объёмом и точностью восстановления: при снижении битрейта уменьшается размер файла, но возрастают отклонения от эталона по SNRseg и спектральным показателям; при увеличении битрейта наблюдается обратная тенденция. Наиболее характерная область изменения качества по спектральному центроиду приходится на 96–128 кбит/с: до этого диапазона центроид быстро приближается к эталону, а при дальнейшем росте битрейта изменения становятся существенно слабее, тогда как CR продолжает монотонно снижаться.

Для кодеков с потерями коэффициент сжатия в основном определяется битрейтом; различия между MP3/AAC/Opus по CR в низких и средних режимах невелики, а при 192–256 кбит/с в ряде групп AAC формирует несколько меньшие по размеру файлы. Для FLAC битрейт как режим не задаётся, и эффективность сжатия определяется статистической избыточностью исходного сигнала. Дополнительно выявлена зависимость результатов от класса аудиоинформации: при 48–64 кбит/с более выраженная деградация метрик наблюдается для EDM и рока, тогда как для классики/акустики и речевого материала изменения по битрейту имеют более плавный характер, что соответствует различиям их временно-спектральной структуры

На основании полученных данных можно сформулировать предпочтительность форматов по условиям применения:

1. При отсутствии жёстких ограничений на объём и при необходимости сохранения точной формы сигнала базовым представлением остаётся WAV/PCM;

2. При требовании строгого сохранения отсчётов при уменьшении объёма хранения без потери качества предпочтителен FLAC, причём ожидаемая

экономия зависит от типа материала;

3. В сценариях передачи и массового распространения, где критичны размер файла и пропускная способность, выбор осуществляется среди кодеков с потерями MP3/AAC/Opus в зависимости от класса материала и режима. Для речевого контента, ориентированного на разборчивость (новостная подача, интервью, разговорные фрагменты), экспериментальные зависимости соответствуют практическому выбору режимов 64–96 кбит/с; при необходимости дополнительного снижения искажений в согласных и шумоподобных компонентах речи целесообразен переход к 128 кбит/с. Для музыкального материала общего назначения (потоковое прослушивание и обмен записями) диапазон 96–128 кбит/с является наиболее рациональным по совокупности уменьшения объёма и стабилизации спектральных показателей.Для музыкальных сигналов со сложной спектрально-временной структурой (EDM, рок) в низких режимах 48–64 кбит/с фиксируется более выраженная деградация показателей, поэтому предпочтительны режимы от 128 кбит/с; при наличии ресурса по каналу переход к 192 кбит/с обеспечивает более высокие значения объективных метрик.

Экспериментальные данные показывают, что в рамках выбранных объективных метрик и диапазона 48–256 кбит/с основной вклад в изменение показателей вносит битрейт, различия между MP3/AAC/Opus в большинстве режимов по масштабу уступают влиянию изменения битрейта. Вместе с тем, в области высоких битрейтов (примерно 192–256 кбит/с) для ряда типов AAC демонстрирует несколько более высокий CR, а для Opus нередко фиксируются более низкие значения SNRseg по сравнению с MP3/AAC. В прикладном плане выбор формата и кодека соответствует их теоретически ожидаемой области применения: AAC типичен для распространения в современных медиаконтейнерах и стриминге, MP3 — для максимальной совместимости с широким спектром устройств и программного обеспечения, Opus — для интерактивных сетевых сценариев, где приоритетны малая задержка и устойчивость передачи.

Список источников

1. T. Painter and A. Spanias, "Perceptual coding of digital audio," in Proceedings of the IEEE, vol. 88, no. 4, pp. 451-515, April 2000, doi: 10.1109/5.842996.
2. Compact Disc Digital Audio System, (IEC/ANSI) CEI-IEC-908,1987.
3. K. Brandenburg and J. D. Johnston, “Second generation perceptual audio coding: The hybrid coder,” in Proc. 88th Conv. Aud. Eng. Soc., Mar. 1990, preprint 2937.
4. Y. F. Dehery, M. Lever, and P. Urcun, “A MUSICAM source code for digital audio broadcasting and storage,” in Proc. ICASSP-91, May 1991, pp. 3605–3608.
5. J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," in IEEE Journal on Selected Areas in Communications, vol. 6, no. 2, pp. 314-323, Feb. 1988, doi: 10.1109/49.608.
6. M. Bosi and R. E. Goldberg, “Introduction to Digital Audio Coding and Standards,” Kluwer Academic Publishers
7. T. Thiede et al., “PEAQ—The ITU Standard for Objective Measurement of Perceived Audio Quality,” Journal of the Audio Engineering Society, vol. 48, no. 1/2, pp. 3–29, 2000.
8. M. Torcoli, T. Kastner and J. Herre, "Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 1530-1541, 2021, doi: 10.1109/TASLP.2021.3069302.
9. Method for the subjective assessment of intermediate quality level of audio systems,” ITU-R Rec. BS.1534.3, 2015.
10. Y. Hu and P. C. Loizou, “Evaluation of objective quality measures for speech enhancement,” IEEE Trans. Audio, Speech, Lang. Process., vol. 16, no. 1, pp. 229–238, Jan. 2008.
11. M. Torcoli and S. Dick, “Comparing the effect of audio coding artifacts on objective quality measures and on subjective ratings,” in Proc. Audio Eng. Soc. 144th Conv., 2018, Art. no. 9951.
12. M. Hansen and B. Kollmeier, “Objective modeling of speech quality with a psychoacoustically validated auditory model,” J. Audio, Eng. Soc, vol. 48, no. 5, pp. 395–409, 2000.
13. C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay,” IEEE Transactions on Acoustics, Speech, and Signal Processing, 1976.
14. G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293–302, Jul. 2002, doi: 10.1109/TSA.2002.800560.
15. Herre, J., Quackenbush, S.R., Kim, M., & Skoglund, J. (2025). Perceptual Audio Coding: A 40-Year Historical Perspective. ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1-5.
16. Herre, J.; Dick, S. Psychoacoustic Models for Perceptual Audio Coding—A Tutorial Review. Appl. Sci. 2019, 9, 2854. https://doi.org/10.3390/app9142854

17. S. Quackenbush, "MPEG Unified Speech and Audio Coding," in IEEE MultiMedia, vol. 20, no. 2, pp. 72-78, April-June 2013, doi: 10.1109/MMUL.2013.24.

18. R. I. Kargin and L. G. Statsenko, “Formats of Audio Data Compression: Analysis and Comparison,” Izvestiya SPbGETU “LETI”, no. 9, pp. 31–37, 2019

19. V. N. Nosulenko and I. V. Starikova, Comparison of Sound Quality of Musical Fragments Differing in Encoding Method, Experimental Psychology, vol. 2, no. 3, pp. 19–34, 2009

20. N. Zeghidour, et al., “SoundStream: An End-to-End Neural Audio Codec,” IEEE/ACM Trans. Audio Speech Lang. Process., vol. 30, pp. 495–507, 2022.

21. J. Princen and A. Bradley, "Analysis/Synthesis filter bank design based on time domain aliasing cancellation," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 34, no. 5, pp. 1153-1161, October 1986, doi: 10.1109/TASSP.1986.1164954.

22. J. Princen, A. Johnson and A. Bradley, "Subband/Transform coding using filter bank designs based on time domain aliasing cancellation," ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, TX, USA, 1987, pp. 2161-2164, doi: 10.1109/ICASSP.1987.1169405.

23. J.-M. Valin, K. Vos, and T. Terriberry, “High-Quality, Low-Delay Music Coding in the Opus Codec,” in Proc. 135th AES Convention, 2013.

24. C. H. Taal, R. C. Hendriks, R. Heusdens and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 2010, pp. 4214-4217, doi: 10.1109/ICASSP.2010.5495701.

25. A. Hines, J. Skoglund, A. Kokaram and N. Harte, "ViSQOL: The Virtual Speech Quality Objective Listener," IWAENC 2012; International Workshop on Acoustic Signal Enhancement, Aachen, Germany, 2012, pp. 1-4.