

И. С. Сергеев, старший инженер, аспирант, e-mail: Noctisik76@gmail.com,
ООО "НТП Трубопровод", Москва, Московский авиационный институт (НИУ)

Комбинированный метод оптимизации звукового потока на основе фрагментации с учетом минимального информационного содержания

Поступила в редакцию 10.05.2025

Принята к публикации 25.05.2025

Представлен метод оптимизации участков звукового потока с минимальным информационным содержанием (таких как тишина, паузы в речи и другие слабоинформативные звуки), основанный на фрагментации и структурном анализе сигнала. Реализация метода предполагает предварительную обработку аудиоданных с разбиением сигнала на фрагменты и присвоением каждому из них описателя, включающего сведения о количественных и качественных характеристиках (например, амплитудных колебаниях и предполагаемом уровне информационного содержания). Это позволяет автоматически выявлять участки с низкой информативностью и заменять их компактными описателями, тогда как как информативные фрагменты сохраняются без изменений.

Предлагаемая оптимизация дополняется алгоритмами сжатия без потерь, включая словарное кодирование (алгоритм Лемпеля—Зива—Уэлча) и энтропийные методы кодирования (алгоритм Хаффмана или арифметического кодирования), что позволяет существенно уменьшить объем звуковых данных при сохранении возможности восстановления их исходной структуры. Результаты экспериментов сопоставляются с эффективностью широко распространенных форматов сжатия с потерями (MP3 и AAC).

Ключевые слова: оптимизация звукового потока, минимальное информационное содержание, фрагментация сигнала, словарное кодирование, энтропийное кодирование, сжатие звуковой информации

Введение

Рост объемов аудиоданных, циркулирующих в информационных системах, ставит перед разработчиками задачу поиска эффективных методов их хранения, передачи и обработки. Особенно остро эта проблема проявляется в технических и телекоммуникационных системах, где передача речевой информации сопровождается ограничениями по пропускной способности каналов и объему хранимых данных. Примером таких систем являются виртуальные автоматические телефонные станции (ВАТС), широко применяемые в организациях для записи и последующего анализа телефонных переговоров, например, между клиентом и менеджером.

Одним из распространенных решений корпоративного уровня является программная ВАТС на базе Asterisk¹, где по умолчанию записи разговоров сохраняются в формате WAV — не-сжатом и ресурсоемком для хранения формате. Это обуславливает необходимость применения сторонних технических решений, направленных на сжатие аудиофайлов для долгосрочного хранения или передачи по сети. В связи с этим возрастает интерес к алгоритмам оптимизации речевого сигнала, позволяющим эффективно

уменьшать объем аудиоданных при сохранении их смыслового содержания.

Большинство существующих методов сжатия звуковых данных (например, MP3 и AAC) основаны на частотных преобразованиях и психоакустических моделях, которые эффективно сокращают объем информации, но не позволяют явно выделить и классифицировать фрагменты с точки зрения их содержательности. При этом потери данных происходят без возможности контролируемого восстановления по уровню смысловой важности [1].

В ряде прикладных задач, связанных с речевой информацией, значительная часть аудиосигнала представлена участками с минимальным информационным содержанием [2] — паузами, фоновыми шумами, дыханием и другими элементами, слабо влияющими на восприятие содержания. Это открывает возможность для целенаправленной оптимизации таких участков, при которой информативная часть сигнала сохраняется в исходном виде, а фрагменты с минимальным информационным содержанием заменяются на компактные описатели, обеспечивая существенное сокращение общего объема данных. Такой подход позволяет сохранить восприятие речевого содержания при минимальных потерях и обеспечивает контроль над структурой и локализацией этих потерь.

В данной работе предлагается метод оптимизации звукового потока, основанный на

¹ Asterisk Documentation: MixMonitor Application. Official Asterisk Project. URL: https://docs.asterisk.org/Latest_API/API_Documentation/Dialplan_Applications/MixMonitor

структурной фрагментации и назначении информационного содержания каждому фрагменту. Ключевая идея заключается в предварительном разбиении звукового сигнала на фрагменты и анализе параметра разброса значений амплитуды в пределах каждого из них. Участки, в которых наблюдается незначительное изменение значений амплитуды, интерпретируются как фрагменты с минимальным информационным содержанием и заменяются описателями, включающими параметры разброса и число отсчетов по фрагменту. При этом фрагменты с информативным содержанием сохраняются в исходном виде.

Дополнительно используется этап сжатия без потерь: применяется словарный алгоритм Лемпеля—Зива—Уэлча, а также энтропийное кодирование — с автоматическим выбором между алгоритмами Хаффмана и арифметического кодирования в зависимости от структуры данных.

Такой подход позволяет осуществлять управляемое уплотнение звуковых данных с возможностью восстановления фрагментов на основе контекста. Восстановление минимальных участков осуществляется путем генерации приближенных значений амплитуд, ограниченных диапазоном значений амплитуды, предшествующих и следующих за уплотненным фрагментом. Метод ориентирован на применение в технических системах, где важна не только степень сжатия, но и сохранение восстанавливаемой структуры сигнала.

Теоретические основы оптимизации звукового потока

Информационное содержание [2] звукового потока может быть проанализировано с двух взаимодополняющих позиций. С точки зрения восприятия человеком оно определяется способностью извлекать смысл из звуковых образов — это могут быть как речевые фрагменты, так и любые акустические сигналы, распознаваемые как значимые (например, звук сирены, мелодия, звук шагов и др.). С технической точки зрения информационное содержание представляет собой совокупность амплитудных фрагментов сигнала, обладающих устойчивой структурой и несущих определенную семантическую нагрузку, которую можно описать с помощью количественных и качественных характеристик. Количественные характеристики связаны с диапазоном изменений амплитудных отсчетов, тогда как качественные характеристики отражают структуру сигнала.

Оптимизация звукового потока должна учитывать особенности распределения инфор-

мационного содержания внутри аудиосигнала. Смысловая нагрузка звукового потока неравномерна: помимо информативных фрагментов (например, содержащих речевую информацию) в сигналах присутствуют участки тишины, паузы, фрагменты с фоновыми шумами и другими слабоинформативными компонентами.

Фрагментация звукового сигнала с анализом амплитудного разброса позволяет выделять участки с низкой смысловой нагрузкой. *Минимальное информационное содержание* (МИС) определяется в данном методе через критерий ограниченного диапазона разброса значений амплитуды внутри фрагмента (небольшая равномерная разница между соседними значениями амплитуды в структурированном звуковом представлении по всему фрагменту). Такие фрагменты можно эффективно заменить компактным описанием, минимизируя тем самым объем хранимой информации без существенного влияния на качество восстановленного сигнала.

Таким образом, в теоретической основе предлагаемого подхода лежит переход от традиционного представления звукового сигнала после цифровой обработки [3] в виде непрерывной последовательности амплитудных отсчетов к структурированной модели, при котором каждый фрагмент анализируется с точки зрения его информационного содержания.

Особенности предлагаемого метода сжатия с потерями

В отличие от распространенных алгоритмов сжатия звука, основанных на потере данных (таких как MP3 и AAC) [4], предлагаемый метод сохраняет структурно значимую часть аудиосигнала в неизменном виде (вплоть до исходных значений амплитуды). Потери применяются только к фрагментам с МИС, что позволяет сохранить смысловую составляющую и выполнить частичное восстановление исходного .wav-файла.

Следует отметить, что эффективность предлагаемого метода в наибольшей степени проявляется при обработке аудиосигналов, содержащих чередующиеся участки высокой и низкой смысловой нагрузки. На практике это характерно для речевых записей, в которых присутствуют паузы, тишина, дыхание, повторяющиеся фрагменты и другие компоненты с пониженным информационным содержанием. Благодаря этому становится возможным формирование фрагментов с ограниченным амплитудным разбросом, которые поддаются замене компактными описателями.

В то же время экспериментальные исследования показали, что метод может быть приме-

ним и к другим типам аудиозаписей, включая технические и отдельные музыкальные фрагменты, при условии наличия в них структурно отделяемых участков с низкой амплитудной вариативностью. Однако при работе с насыщенными музыкальными записями, где разброс амплитуд остается высоким на протяжении всего сигнала, эффективность оптимизации фрагментов с МИС снижается.

Метод оптимизации звукового потока

Метод оптимизации звукового потока реализуется в несколько этапов предварительной обработки, позволяющих структурировать [5] исходные данные и классифицировать участки сигнала по уровню информационного содержания [6].

1. *Структуризация.* На данном этапе выполняется анализ временного ряда амплитудных значений в целях выделения характерных точек — локальных максимумов и минимумов, на основе которых строятся структуры. Эти структуры впоследствии служат базой для фрагментации.

2. *Фиксированная фрагментация.* Звуковой поток разбивается на фрагменты фиксированной длины с равномерным структурным распределением. Для каждого фрагмента рассчитываются количественные характеристики: диапазон разброса значений амплитуды, число отсчетов по фрагменту и др.

3. *Совокупная фрагментация.* Выполняется объединение фрагментов с близкими характеристиками в совокупности. Основным параметром здесь выступает диапазон разброса значений амплитуды. При этом проводится классификация фрагментов: если разброс амплитуд в пределах заданного порога (например, не более 2 единиц), фрагмент считается минимальным по информационному содержанию (МИС-фрагмент).

4. *Разметка по калибру диапазонов разброса значений амплитуд* — для каждого фрагмента определяется параметр разброса значений амплитуд, число отсчетов между фрагментами, число отсчетов "по прямой" и "по косой".

Результаты анализа фиксируют информацию о границах фрагментов, параметре разброса и числе отсчетов. На основе анализа проводится оптимизация фрагментов с МИС, по итогу которой:

- для фрагментов с МИС записывается только описатель, включающий параметр разброса и число отсчетов;
- для информативно насыщенных фрагментов сохраняются исходные амплитудные значения.

Таким образом, итоговое представление звукового сигнала содержит точные значения амплитуды лишь для участков с предполагаемой смысловой нагрузкой, а участки с низким информационным содержанием заменяются компактной записью. Такой подход позволяет добиться эффективного уменьшения объема данных без потери значимых характеристик сигнала.

На рис. 1 представлено визуальное разбиение звукового потока, соответствующего произнесенной речевой фразе "Будем пробовать", на совокупные фрагменты с различным уровнем информационного содержания. Фрагменты с минимальным информационным содержанием (SF_min) выделены и масштабированы для наглядности. Разделение выполнено на основе анализа амплитудного разброса. В качестве инструмента анализа структуры звукового сигнала использовалась авторская система "Волнан"¹. Данный инструмент применялся на этапе оценки локальных изменений значений амплитуды и верификации результатов фрагментации.

В рамках проверки устойчивости метода к преднамеренному удалению малозначимой информации был проведен эксперимент с полным обнулением значений амплитуды в ряде фрагментов, классифицированных как участки с минимальным информационным содержанием. На рис. 2 представлено визуальное отображение измененного звукового сигнала. Несмо-

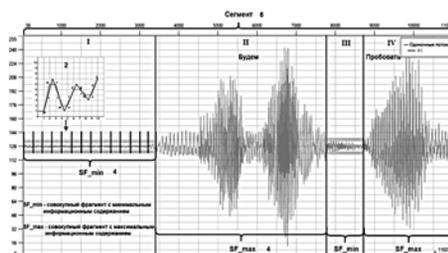


Рис. 1. Разбиение звукового потока на фрагменты с минимальным и выраженным информационным содержанием (фраза "Будем пробовать")

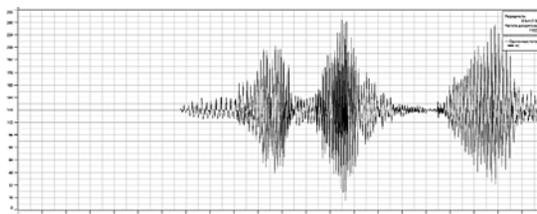


Рис. 2. Осциллограмма сигнала с нулевыми значениями амплитуды в фрагментах МИС

¹ Свидетельство о государственной регистрации программы для ЭВМ № 2023684011. Российская Федерация. Система обеспечения извлечения информационного содержания волн: заявл. 05.11.2023; опубл. 25.12.2023 / М. М. Фадеев, Н. Е. Балакирев. — Зарегистрировано в Федеральной службе по интеллектуальной собственности.

тря на искусственное удаление данных в ряде участков, основные информативные фрагменты сигнала остаются неизменными. Это подтверждает, что предложенный метод позволяет реализовать управляемые потери без потери восприятия информационного содержания.

Расширение метода оптимизации звукового потока

Даже после замены фрагментов с минимальным информационным содержанием на компактные описатели итоговый объем данных может оставаться значительным, особенно при большом числе фрагментов. Для дальнейшего уменьшения размера оптимизированного представления предлагается применение методов без потерь: словарного и энтропийного сжатия [7].

На первом этапе используется алгоритм Лемпеля—Зива—Уэлча (LZW) [8], формирующий динамический словарь повторяющихся фрагментов описателей. Это позволяет заменить часто встречающиеся комбинации разбросов и чисел отсчетов на более компактные ссылки в словаре.

На завершающем этапе применяется энтропийное кодирование. В зависимости от характеристик данных может использоваться:

- алгоритм Хаффмана [7], эффективный при наличии ярко выраженного распределения вероятностей описателей;
- арифметическое кодирование, позволяющее достичь более высокой степени сжатия за счет дробной длины кодовых слов (кодирование последовательностей символов) [9].

Дополнение базовой процедуры оптимизации фрагментов с минимальным информационным содержанием методами без потерь позволяет увеличить общий коэффициент сжатия. Однако реализация этих методов требует учета специфики полученного потока данных.

После этапа оптимизации структура данных включает чередование компактных описателей для фрагментов с минимальным информационным содержанием и последовательностей из описателя и исходных значений амплитуды для информативных фрагментов. Это приводит к смешанному и не всегда регулярному потоку, что может снижать эффективность словарных и энтропийных методов при прямом применении.

На практике эффективность дальнейшего сжатия зависит от степени повторяемости описателей и однородности амплитудных фрагментов. В случаях, когда поток демонстрирует устойчивые закономерности, применение алгоритма Лемпеля—Зива—Уэлча дает заметное сокращение объема, а последующее энтропийное кодирование [10] (Хаффмана или арифметическое) допол-

нительно уплотняет структуру. В других случаях эти методы могут оказываться избыточными.

Проверка применимости и выбор энтропийного кодирования

Для повышения эффективности комбинированной оптимизации реализована проверка выходного потока после работы словарного алгоритма Лемпеля—Зива—Уэлча (LZW), определяющая целесообразность применения энтропийных алгоритмов и предпочтительный тип энтропийного кодирования. Проверка включает два этапа:

1. Проверка целесообразности:

— оценка степени сжатия по словарю: если коэффициент словарного сжатия менее 5 % (т. е. $K_{LZW} < 1,05$), то дальнейшее применение энтропийных методов предлагается считать нецелесообразным, поскольку это значение указывает на тот факт, что данные обладают высокой энтропией или низкой избыточностью. В таких случаях последующее применение энтропийных методов может привести к незначительному дополнительному сжатию или даже увеличению размера данных из-за накладных расходов, связанных с построением и хранением кодовых таблиц;

— анализ плотности алфавита: вычисляется доля уникальных символов в потоке. Если более 40 % байтов уникальны, поток считается неструктурированным, и применение кодирования Хаффмана и арифметического кодирования отклоняется;

— расчет энтропии: если энтропия выходного потока превышает 7,5 бит/символ, оба метода считаются неэффективными. В диапазоне 5,5...7,5 бит/символ допускается использование только арифметического кодирования. Если энтропия ниже 5,5, разрешены оба метода.

2. Выбор энтропийного метода сжатия:

— если энтропия ниже 5,5 бит/символ, и распределение вероятностей неравномерно, применяется кодирование Хаффмана;

— если энтропия в диапазоне 5,5...7,5 бит/символ, но поток допускает дополнительное уплотнение, используется арифметическое кодирование;

— при необходимости максимального коэффициента сжатия предпочтение также отдается арифметическому методу ввиду его способности оперировать с дробными длинами кодов (поскольку кодирование выполняется для последовательности символов, а не посимвольно).

Таким образом, система автоматически принимает решение как о целесообразности, так и о выборе конкретного алгоритма, обеспечивая адаптивную стратегию сжатия с учетом статистических свойств оптимизированного потока.

Математическая формализация метода оптимизации фрагментов с минимальным информационным содержанием

Пусть исходный звуковой сигнал представлен в виде дискретной последовательности S значений амплитуды a , где каждое значение амплитуды принадлежит множеству целых чисел \mathbb{Z} :

$$S = \{a_1, a_2, \dots, a_n\}, \quad a_i \in \mathbb{Z}, \quad i = 1, 2, \dots, n.$$

Сигнал разбивается на фрагменты F_j длины L_j :

$$F_j = \{a_{n_j}, a_{n_j+1}, \dots, a_{n_j + L_j - 1}\}, \quad n_j + 1 = n_{j-1} + L_{j-1}.$$

Для каждого фрагмента вычисляется диапазон разброса значений амплитуд R :

$$R_j = \max(F_j) - \min(F_j).$$

Фрагмент считается МИС-фрагментом, если

$$R_j \leq \delta,$$

где δ — заданный порог, например 2.

Формат описателя фрагментов:

- для МИС:
если $R_j \leq \delta$ (МИС-фрагмент), он заменяется описателем D :

$$D_j = (R_j, L_j);$$

- для информативного фрагмента:
если $R_j \geq \delta$ (информативный фрагмент), описатель D содержит

$$D_j = (R_j, L_j, \{a_{n_j}, a_{n_j+1}, \dots, a_{n_j + L_j - 1}\}).$$

Математическая формализация метода восстановления фрагментов с минимальным информационным содержанием

Пусть a_p — последнее значение амплитуды перед началом фрагмента МИС, a_q — первое значение амплитуды после фрагмента МИС. Тогда восстановление значений амплитуды данного фрагмента осуществляется аппроксимацией значений амплитуды в диапазоне $[a_p, a_q]$ с помощью генерации L_j случайных значений амплитуды:

$$\widehat{F}_j = \{r_1, \dots, r_{L_j}\},$$

где $r_i \in \text{Rand}[a_p, a_q]$.

Математическая формализация проверки применимости энтропийного кодирования

Пусть после словарного кодирования (алгоритм Лемпеля—Зива—Уэлча) получен поток символов $C = \{c_1, c_2, \dots, c_n\}$, где каждый $c_i \in \mathbb{B}$, где \mathbb{B} — байт (от 0 до 255).

Обозначим:

- $n = |C|$ — длина потока;
- $D = |\text{Unique}C|$ — число уникальных символов в C ;
- $K_{LZW} = \frac{S_{\text{до LZW}}}{S_{\text{после LZW}}}$ — коэффициент словарного сжатия;
- $H(C)$ — энтропия потока C , рассчитываемая по формуле

$$H(C) = -\sum_{i=1}^D p_i \cdot \log_2(p_i),$$

где $p_i = \frac{f_i}{n}$, а f_i — частота появления i -го символа в C .

Условия применимости энтропийного кодирования

Энтропийное кодирование применяется только в случае, если одновременно выполняются условия

$$\begin{cases} K_{LZW} \geq 1,05; \\ \frac{D}{n} \leq 0,4; \\ H(C) \leq 7,5. \end{cases}$$

Выбор алгоритма:

$$\text{Метод} = \begin{cases} \text{алгоритм Хаффмана,} \\ \text{если } H(C) < 5,5; \\ \text{арифметическое кодирование,} \\ \text{если } 5,5 < H(C) \leq 7,5; \\ \text{не применяется,} \\ \text{если } H(C) > 7,5. \end{cases}$$

Программная реализация

Разработанная программная реализация алгоритма "Акустический оптимизатор" представляет собой десктопное приложение на языке C# с графическим пользовательским интерфейсом. Программа обеспечивает обработку входных звуковых файлов формата .wav, реализацию алгоритма оптимизации с учетом минимального информационного содержания, а также применение словарного и энтропийного сжатия по выбору пользователя или автоматически по результату встроенных проверок на целесообразность применения.

Пример интерфейса ручного режима для задачи сжатия/восстановления с примером выполнения продемонстрирован на рис. 3.

Аналогичный интерфейс присутствует для алгоритмов сжатия без потерь (словарные алгоритмы, энтропийные алгоритмы). Пользователю доступен выбор файла для сжатия/восстановления, выбор диапазонов калибра для процедуры совокупной фрагментации звукового потока перед оптимизацией объема фрагментов МИС и расчет коэффициентов сжатия.

Для выполнения комбинации из методов уплотнения (сценариев сжатия) пользователю доступен другой интерфейс, где диапазоны калибровки рассчитываются автоматически, и проводятся проверки на целесообразность энтропийного сжатия для предлагаемого метода. На рис. 4 визуальное представление реализации алгоритмов в приложении.

В интерфейсе пользователь может выбрать включение словарного сжатия (LZW), а также задать использование энтропийного кодирования. В случае выбора автоматического режима система применяет проверку статистических характеристик потока (коэффициент сжатия, доля уникальных символов, энтропия) и определяет, применять ли энтропийное кодирование и какой именно алгоритм — Хаффмана или арифметический.

Для сопоставления эффективности предлагаемого метода оптимизации в программе пользователю также предоставлена возможность сжатия с потерями в форматы MP3 и AAC для сравнения результатов. Все аудиофайлы представлены в виде одноканального (моно) сигнала с глубиной квантования 8 бит и частотой дискретизации 11 025 Гц. В ходе экспериментов битрейт фиксирован на уровне 88 кбит/с. Для преобразования в формат MP3 использовалась библиотека "NAudio.Lame"¹, а для преобразования в AAC применялся модуль "FFmpeg"².

На рис. 5 представлен пример выполнения полного сценария сжатия звукового файла alphabet8.wav с использованием всех этапов предложенного алгоритма. В ходе обработки были последовательно применены:

- оптимизация фрагментов с минимальным информационным содержанием (МИС);
- словарное сжатие (алгоритм Лемпеля—Зива—Уэлча);
- арифметическое кодирование.

¹ NAudio.Lame: Wrapper for libmp3lame.dll to add MP3 encoding support to NAudio / Corey Murtagh. — GitHub. URL: <https://github.com/Corey-M/NAudio.Lame>

² FFmpeg: A complete, cross-platform solution to record, convert and stream audio and video. — Официальный сайт проекта. URL: <https://www.ffmpeg.org/>

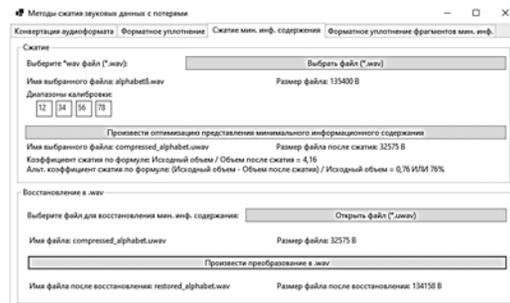


Рис. 3. Интерфейс сжатия фрагментов с минимальным информационным содержанием

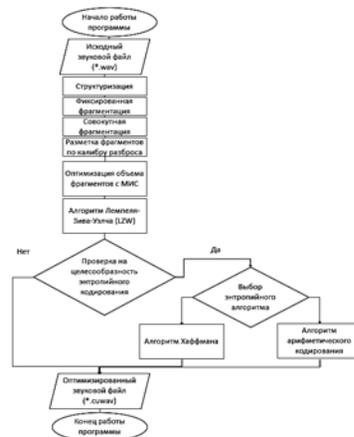


Рис. 4. Визуальное представление реализации комбинированного метода оптимизации звукового потока



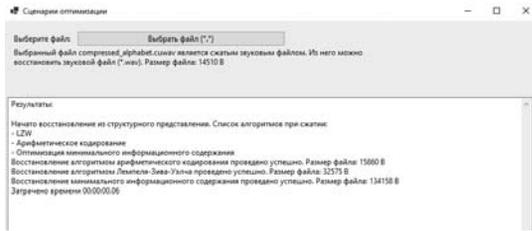


Рис. 6. Пример восстановления *.wav файла из сжатого представления

ненных этапах сжатия сохраняется в заголовке файла в виде битовой маски, указывающей последовательность алгоритмов.

Процесс восстановления начинается с чтения служебной информации из заголовочной части файла, на основании которой формируется обратная последовательность алгоритмов декомпрессии. Процесс восстановления состоит в применении алгоритмов декодирования в порядке, обратном тому, что был в процессе сжатия. Последним этапом восстановления является замена описателей во фрагментах МИС на аппроксимированные значения амплитуды.

На рис. 6 показан пример автоматического восстановления звукового файла из выбранного пользователем файла *.cuwav, полученного после применения всех этапов комбинированной оптимизации (см. рис. 5). Программа автоматически определяет цепочку примененных алгоритмов и выполняет восстановление в обратной последовательности: сначала проводится восстановление с использованием арифметического кодирования, затем декомпрессия потока LZW и, наконец, восстановление значений амплитуды во фрагментах с минимальным информационным содержанием по описателям. На каждом этапе восстановления фиксируется промежуточный размер выходного файла. В результате восстановления был получен .wav-файл объемом 134 158 байт, близкий к исходному сигналу по структуре и содержательной части.

Экспериментальные результаты

Для оценки эффективности предложенного метода была проведена серия экспериментов на звуковых файлах формата .wav с различной длительностью и насыщенностью речевого сигнала. Для каждого файла поэтапно применялись:

- оптимизация фрагментов с минимальным информационным содержанием;
- словарное сжатие (алгоритм Лемпеля—Зива—Уэлча);
- энтропийное кодирование (алгоритм Хаффмана или арифметическое кодирование — по результату встроенных проверок);
- восстановление *.wav-файла.

По результатам автоматических проверок выявлено, что для всех обрабатываемых звуковых файлов поток данных после применения словарного сжатия (LZW) характеризуется высокой степенью равномерности распределения символов и высокой энтропией (7,1...7,7 бит/символ). В таких условиях энтропийное кодирование по алгоритму Хаффмана либо приводит к избыточности (увеличению размера файла за счет словаря), либо уступает по степени сжатия арифметическому кодированию. В ряде случаев алгоритм Хаффмана даже не проходит проверку целесообразности. Таким образом, по результатам анализа можно сделать вывод о нецелесообразности использования кодирования Хаффмана в рамках предложенного метода и предпочтительности применения исключительно арифметического кодирования на завершающем этапе.

В таблице представлены экспериментальные результаты сжатия. Для каждого исследования представлены полученные объемы сжатия в байтах, коэффициент (K) сжатия по формуле "Исходный объем / Объем данных после сжатия" и время сжатия (t) в секундах. Для сравнения показателей также проведены эксперименты сжатия в форматы MP3 и AAC в статическом битрейте (CBR). Звуковые файлы являются одноканальными (моно), число бит на сэмпл (значение амплитуды) — 8 бит, частота дискретизации — 11 025. Битрейт — 88 Кбит/с. Поскольку для сжатия в MP3 и AAC использовались сторонние библиотеки, то время выполнения алгоритмов не сравнивается со временем предлагаемой оптимизации (оно всегда выше, и это сравнение некорректно).

Примечание. Для обеспечения сопоставимости результатов с предложенным методом при

Экспериментальные результаты сжатия

Имя файла (*.wav)	Длительность, с	Исходный объем, байт	Оптимизация МИС + LZW + арифметическое кодирование	MP3 (CBR)	AAC (CBR)
			Коэффициент сжатия		
monologue	45	505 840	4,17	1,37	1,61
alphabet	12	135 400	9,33	1,36	1,60
micNoise	33	335 945	6,51	1,37	1,61
femMono	25	283 135	3,53	1,37	1,63
femDialog	25	279 828	2,66	1,37	1,61
phrase1	10	115 776	4,4	1,36	1,59
phrase2	2	21 548	1,55	1,26	1,49
lectionPart	20	228 562	3,13	1,37	1,61
guitarE6	12	133 636	3,06	1,36	1,57
fullSilence	30	335 173	10,2	1,37	1,63

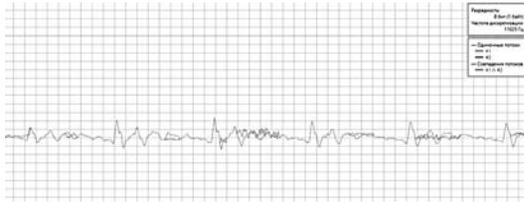


Рис. 7. Визуальное сравнение исходного и восстановленного сигналов файла "alphabet" (произнесение фонем алфавита русского языка "а-я")

кодировании в форматы MP3 и AAC использовался режим постоянного битрейта (CBR). Это исключает влияние адаптивного управления качеством (как в VBR/ABR) и позволяет корректно сравнивать объемы выходных файлов при одинаковых условиях.

В дополнение к количественной оценке на рис. 7 представлено визуальное сравнение исходного и восстановленного звуковых сигналов в виде осциллограмм. Четко видно, что основное расхождение приходится на участки с минимальным содержанием, которые восстанавливаются приблизительно, тогда как информативные фрагменты воспроизводятся без искажений. Это подтверждает управляемый характер потерь в предложенном методе оптимизации. Осциллограмма составлена с использованием программы "ВОЛНАН".

Для исследования выбраны два звуковых потока исходного звукового файла alphabet.wav и восстановленного файла.

Выводы по экспериментальным результатам

Разработанный метод демонстрирует устойчивое сжатие звуковых данных с коэффициентом от 4,17 до 10,2 при низком времени сжатия (до 0,3 с) и может количественно сравниваться с общеизвестными форматами сжатия MP3 и AAC, превосходя MP3 при статическом битрейте.

Предлагаемый метод наилучшим образом проявляет себя при сжатии речевой информации, однако может быть адаптирован и для других типов звуковых данных, включая музыкальные фрагменты. Так, при обработке сигнала "guitar_E6" (нота Ми, сыгранная на 6-й струне гитары) были успешно сжаты участки, не содержащие основную ноту, т. е. обладающие низкой информационной насыщенностью. Вместе с тем, эффективность оптимизации по МИС снижается при обработке звуков с высокой амплитудной вариативностью. На примере гитарного перебора с длительностью 1 мин установлено, что в подобных случаях основную роль играют универсальные алгоритмы без потерь (LZW и арифметическое кодирование), поскольку в музыкальном сигнале практически

отсутствуют малозначимые фрагменты, подлежащие описательной компрессии.

При сравнении с MP3 и AAC используется статический битрейт для соблюдения схожих условий сжатия, поскольку в предлагаемом методе битрейт не изменяется. Предлагаемый метод показывает лучшие результаты, чем MP3 и AAC.

Заключение

Предложенный метод оптимизации звукового потока с учетом минимального информационного содержания демонстрирует высокую эффективность при обработке звуковых данных, в первую очередь речевого характера, где чередуются участки с высокой и низкой смысловой нагрузкой. Ключевая особенность подхода заключается в структурной фрагментации сигнала и дифференцированной обработке: информативные фрагменты сохраняются в исходном виде, тогда как слабоинформативные (паузы, шумы, монотонные колебания) заменяются компактными описателями.

Экспериментально подтверждена возможность достижения коэффициента сжатия до 10,2 при сохранении восприятия и структуры звукового образа. Восстановление оптимизированных участков выполняется методом аппроксимации, ограниченной соседними амплитудами, что обеспечивает управляемость потерь. Метод показал устойчивую применимость и к неречевым сигналам, включая музыкальные фрагменты с эпизодами низкой амплитудной изменчивости. Однако при высокой плотности информационного содержания (например, в живом музыкальном исполнении) эффективность оптимизации снижается, и основное сжатие достигается за счет использования словарных и энтропийных алгоритмов.

По результатам проверок можно отметить, что наибольшую степень уплотнения обеспечивает связка словарного метода (алгоритм Лемпеля—Зива—Уэлча) с арифметическим кодированием. При этом применение алгоритма Хаффмана в таких условиях оказывается нецелесообразным из-за недостаточной эффективности при высокоэнтропийных распределениях.

Таким образом, разработанный метод является эффективным инструментом для сжатия звуковой информации в задачах, ориентированных на баланс между скоростью, компактностью и допустимыми искажениями. Перспективы дальнейшего развития включают адаптацию алгоритма к потоковой передаче, а также сочетание с нейросетевыми механизмами для улучшения качества реконструкции в условиях агрессивного сжатия.

Дальнейшее развитие работы связано с адаптацией метода для потоковой обработки звука в реальном времени, исследованием устойчивости к акустическим помехам и возможной интеграцией с моделями машинного обучения для улучшения процедуры восстановления фрагментов и автоматической классификации содержательности.

Список литературы

1. **Бранденбург К.** MP3 и AAC: объяснение [Электронный ресурс] // Fraunhofer IIS. 1999. 12 с. URL: https://www.iis.fraunhofer.de/content/dam/iis/de/doc/ame/conference/AES-17-Conference_mp3-and-AAC-explained_AES17.pdf
2. **Балакирев Н. Е., Фадеев М. М., Родионов В. С.** Качественный подход в раскрытии информационного содержания волновых данных // Труды МАИ. 2024. № 136. URL: https://trudymai.ru/upload/iblock/e63/sctyasotfzi3nngx1m3z86vc-5gagvi23b19_Balakirev_Fadeev_Rodionov.683.pdf
3. **Солонина А. И., Д. А. Улахович, Арбузов С. М., Соловьева Е. Б.** Основы цифровой обработки сигналов: Курс лекций. СПб: БХВ-Петербург, 2005. 768 с. С. 7—17.
4. **Чижов И. И., Балабанова Т. Н.** Сжатие аудиоданных на основе психоакустических принципов восприятия звука человеком // Современные аспекты информационных технологий. 2021. № 2 (30). С. 106—110. DOI: 10.17308/sait/1995-5499/2024/3/127-137
5. **Балакирев Н. Е., Нгуен Х. З., Малков М. А., Фадеев М. М.** Структуризация и качественное рассмотрение звукового потока в системе синтеза и анализа речи // Программные продукты и системы. 2018. Т. 31, № 4. С. 768—776.
6. **Фадеев М. М., Балакирев Н. Е., Родионов В. С., Сергеев И. С., Умрюхин Е. А.** Количественная и качественная фрагментация потока данных // Материалы XXIV Международной конференции "Информатика: проблемы, методология, технологии". (Воронеж, 14—15 февраля 2024). Воронеж: Изд. дом ВГУ, 2024. С. 807—813.
7. **Скляр Б.** Цифровая связь: теоретические основы и практическое применение. М.: Вильямс, 2003. 1104 с.;
8. **Welch T. A.** A Technique for High-Performance Data Compression // IEEE Computer. 1984. Vol. 17, N. 6. P. 8—19.
9. **Уиттен Дж.** Арифметическое кодирование и его практическое применение // Коммуникации ACM. 1987. Т. 30, № 6. С. 520—540.
10. **Сергеев И. С., Балакирев Н. Е.** Сравнение алгоритмов сжатия звуковой информации алгоритмом Хаффмана и арифметическим кодированием // Наукосфера. 2022. № 8 (2). С. 31—35.

I. S. Sergeev, Senior Engineer¹, PhD Student², e-mail: Noctisik76@gmail.com,

¹Piping Systems Research & Engineering Co "TRUBOPROVOD" PSRE Co., Moscow, 11141, Russian Federation,

²Moscow Aviation Institute (NRU), Moscow, 125993, Russian Federation

The Combined Method for Audio Stream Optimization Based on Fragmentation and Minimal Information Content Assessment

Received on 10.05.2025

Accepted on 25.05.2025

This paper presents a method for optimizing segments of an audio stream with minimal informational content (such as silence, speech pauses, and other weakly informative sounds) based on fragmentation and structural analysis of the signal. The method involves preliminary processing of audio data by dividing the signal into fragments and assigning each a descriptor containing information about quantitative and qualitative characteristics (e.g., amplitude fluctuations and the assumed level of informational content). This approach enables automatic identification of low-informative segments, which are replaced with compact descriptors, while informative fragments are preserved unchanged. The proposed optimization is complemented by lossless compression algorithms, including dictionary-based coding (Lempel—Ziv—Welch algorithm) and entropy-based methods (Huffman or arithmetic coding), allowing a significant reduction in audio data size while retaining the ability to reconstruct the original structure. The experimental results are compared with the performance of widely used lossy compression formats (MP3 and AAC).

Keywords: audio stream optimization, minimal informational content, signal fragmentation, dictionary-based coding, entropy coding, audio compression

DOI: 10.17587/it.31.578-586

References

1. **Brandenburg K.** MP3 and AAC Explained, *Fraunhofer IIS*, 1999, 12 p., available at: https://www.iis.fraunhofer.de/content/dam/iis/de/doc/ame/conference/AES-17-Conference_mp3-and-AAC-explained_AES17.pdf
2. **Balakirev N. E., Fadeev M. M., Rodionov V. S.** Qualitative approach in extracting the information content of wave data, *Trudy MAI*, 2024, no. 136, available at: <https://trudymai.ru/eng/published.php?ID=180683>
3. **Solonina A. I., Ulakhovich D. A., Arbuzov S. M., Solovyeva E. B.** Fundamentals of Digital Signal Processing: Lecture Course. St. Petersburg, BHV-Petersburg, 2005, pp. 7—17.
4. **Chizhov I. I., Balabanova T. N.** Compression of Audio Data Based on Psychoacoustic Principles of Human Sound Perception, *Modern Aspects of Information Technologies*, 2021, no. 2 (30), pp. 106—110, DOI: 10.17308/sait/1995-5499/2024/3/127-137
5. **Balakirev N. E., Nguen Kh.Z., Malkov M. A., Fadeev M. M.** *Programmnye produkty i sistemy*, 2018, vol. 31, no. 4, pp. 768—776.
6. **Fadeev M. M., Balakirev N. E., Rodionov V. S., Sergeev I. S., Umryukhin E. A.** *Materialy XXIV Mezhdunarodnoi konferentsii "Informatika: problemy, metodologiya, tekhnologii"*, Voronezh, Izd. dom VGU, 2024, pp. 807—813.
7. **Sklyar B.** Digital Communications: Fundamentals and Applications, Moscow, Williams, 2003, 1104 p.
8. **Welch T. A.** A Technique for High-Performance Data Compression, *IEEE Computer*, 1984, vol. 17, no. 6, pp. 8—19.
9. **Witten I. H., Neal R. M., Cleary J. G.** Arithmetic coding for data compression, *Communications of the ACM*, June 1987, vol. 30, no. 6, pp. 520—540.
10. **Sergeev I. S., Balakirev N. E.** Comparison of Audio Compression Algorithms: Huffman vs. Arithmetic Coding, *Naucosfera*, 2022, no. 8 (2), pp. 31—35.