

РАДИОТЕХНИКА, РАДИОЛОКАЦИЯ И СИСТЕМЫ СВЯЗИ

УДК 621.395

**ОПТИМИЗАЦИЯ ЭНЕРГЕТИЧЕСКИХ ПОРОГОВ
ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ ДЛЯ СЖАТИЯ РЕЧЕВЫХ СИГНАЛОВ
НА ОСНОВЕ АЛГОРИТМА РОЯ ЧАСТИЦ****В. Т. Дмитриев**, д.т.н., доцент, заведующий кафедрой РУС РГРТУ, Рязань, Россия;

orcid.org/0000-0001-5521-6886, e-mail: vol77@rambler.ru

Ву Хоанг Шон, аспирант кафедры РУС РГРТУ, Рязань, Россия;

orcid.org/0009-0004-7428-5296, e-mail: vuhoangson.adaf@gmail.com

Предложен и исследован метод адаптивной оптимизации энергетических порогов дискретного вейвлет-преобразования на основе алгоритма роя частиц (PSO) для сжатия речевых сигналов. Данный метод обеспечивает автоматический выбор оптимальных долей сохраняемой энергии на каждом уровне вейвлет-разложения. Оптимизация производится с целью максимизации степени сжатия при одновременном соблюдении заранее заданных требований к качеству восстановленной речи на приеме. Качество восстановленной речи на выходе системы оценивается с использованием двух объективных метрик: сегментного отношения сигнал-шум SegSNR и перцептивной метрики качества восстановленной речи ViSQOL. Экспериментальные исследования, проведенные на базе стандартных речевых сигналов записанных согласно ГОСТ Р 50840-95, показали, что предложенный метод позволяет достичь степени сжатия на уровне 87 % при сохранении высокого качества восстановленной речи SegSNR = 9,5 дБ, ViSQOL = 3,9 балла. По интегральному критерию эффективности предложенный подход превосходит классические методы на 7...14 %.

Ключевые слова: дискретное вейвлет-преобразование, алгоритм роя частиц, энергетическое пороговое сжатие, оптимизация параметров, SegSNR, ViSQOL, речевые сигналы.

DOI: 10.21667/1995-4565-2026-95-3-15**Введение**

Цифровая обработка речевых сигналов занимает центральное место в современных телекоммуникационных системах, где непрерывно возрастающие требования к эффективности передачи данных сталкиваются с ограниченностью частотно-временных ресурсов каналов связи. Речевые сигналы характеризуются высокой избыточностью, что открывает возможности для существенного сжатия без критической деградации воспринимаемого качества. Традиционные подходы к речевому кодированию, включающие параметрические методы первичного кодирования (LPC, CELP) и вейвлет-ориентированные техники, демонстрируют различные степени компромисса между скоростью передаваемой информации, вычислительной сложностью и качеством восстановления речи. Задача адаптивной оптимизации параметров сжатия под конкретные условия среды передачи остается актуальной, особенно в контексте гетерогенных сетей с динамически изменяющимися характеристиками каналов связи.

Дискретное вейвлет-преобразование (ДВП) обеспечивает эффективное спектрально-временное представление речи благодаря мультимасштабной декомпозиции с локализацией по времени и частоте. Фундаментальное свойство вейвлет-базисов – способность компактно представлять сигналы с ограниченной регулярностью – делает их естественным инструментом для сжатия речи, где энергия концентрируется в относительно небольшом числе коэффициентов разложения. Ключевым этапом вейвлет-сжатия является пороговая обработка ко-

эффициентов, определяющая баланс между степенью сжатия и качеством аппроксимации исходного сигнала. Классические методы используют глобальные или эмпирически адаптивные пороги, но не учитывают специфику энергетического распределения по уровням ДВП и не оптимизируют целевую функцию при заданных ограничениях на качество восстановленной речи.

Энергетический подход, при котором на каждом уровне сохраняется доля энергии, лучше учитывает структуру речевого сигнала и позволяет адаптивно управлять вкладом отдельных частотных поддиапазонов в общее качество речи. Однако оптимальный подбор вектора энергетических порогов – нетривиальная задача нелинейного программирования с неаналитическими целевой функцией и ограничениями, исключающая применение градиентных методов. Метаэвристические алгоритмы, в частности алгоритмы роя частиц (Particle Swarm Optimization, PSO), эффективны для подобных задач благодаря способности работать в мультимодальных пространствах поиска, простоте реализации и балансу между глобальным исследованием и локальной эксплуатацией решений.

Цель данной работы – разработка метода адаптивной оптимизации энергетических порогов многоуровневого дискретного вейвлет-преобразования на основе алгоритма роя частиц PSO для максимизации коэффициента сжатия при одновременном обеспечении требуемого качества восстановленной речи по метрикам ViSQOL и сегментного отношения сигнал-шум (SegSNR).

Теоретическая часть

В настоящем разделе представлены теоретические основы предлагаемого метода, включая дискретное вейвлет-преобразование с покadroвой обработкой, энергетическое пороговое сжатие коэффициентов, метрики оценки эффективности сжатия и постановку задачи оптимизации.

Дискретное вейвлет-преобразование речевых сигналов с покadroвой обработкой

Исходный дискретный речевой сигнал $x[n]$, где $n = 0, 1, \dots, N - 1$, с длительностью выборки N отсчётов, разбивается на M перекрывающихся кадров длиной $N_{fr} = 256$ отсчётов (что соответствует 32 мс при частоте дискретизации 8 кГц). Каждый кадр $y_m[n]$, где $m = 1, 2, \dots, M$, предварительно подвергается оконной обработке с использованием функции Ханна для минимизации краевых эффектов и далее рассматривается как локально стационарный фрагмент.

К каждому кадру независимо применяется J -уровневое ДВП, реализуемое посредством быстрого алгоритма Малла [1]. Преобразование выполняется итеративно: на каждом уровне j ($j = 1, 2, \dots, J$) входная последовательность фильтруется парой фильтров – низкочастотным $h[n]$ и высокочастотным $g[n]$, после чего подвергается децимации с коэффициентом 2. В результате получаются два набора коэффициентов: коэффициенты аппроксимации $cA_j^{(m)}$, представляющие низкочастотную составляющую, и коэффициенты деталей $cD_j^{(m)}$, соответствующие высокочастотной компоненте. После завершения J итераций формируется полное вейвлет-разложение кадра в виде набора коэффициентов: $\mathbf{C}_m = [cA_j^{(m)}; cD_j^{(m)}; cD_{j-1}^{(m)}; \dots; cD_1^{(m)}]^T$ общей длиной N_{fr} (при использовании периодических граничных условий).

После вейвлет-разложения сжатие достигается за счёт отбрасывания малозначимых коэффициентов. В работе используется энергетический подход: на каждом уровне l сохраняется заданная доля α_l его энергии. В отличие от классических амплитудных методов, такой подход оценивает информативность коэффициентов по их вкладу в энергию уровня, что лучше соответствует особенностям речевых сигналов. Для набора коэффициентов уровня l ($l = 0$ для аппроксимации, $l = 1, 2, \dots, J$ для деталей) определим полную энергию уровня:

$$E_l^{(m)} = \sum_{k=1}^{L_l} (c_k^{(l,m)})^2, \tag{1}$$

где L_l – число коэффициентов на уровне l кадра m ; $c_k^{(l,m)}$ – k -й коэффициент уровня l кадра m ($k = 1, 2, \dots, L_l$). Энергетический порог $\alpha_l \in [0, 1]$ задает целевую долю энергии, подлежащую сохранению на уровне l . Алгоритм сжатия последовательно применяется к каждому уровню разложения всех кадров речевого сигнала.

Алгоритм энергетического порогового сжатия для уровня l кадра m состоит из четырех этапов.

Шаг 1. Упорядочить коэффициенты по убыванию абсолютных значений, получив перестановку π такую, что:

$$\left| c_{\pi(1)}^{(l,m)} \right| \geq \left| c_{\pi(2)}^{(l,m)} \right| \geq \dots \geq \left| c_{\pi(L_l)}^{(l,m)} \right|, \tag{2}$$

Шаг 2. Вычислить кумулятивную энергию первых s коэффициентов в упорядоченном списке:

$$E_{\Sigma}^{(l,m)}(s) = \sum_{\xi=1}^s (c_{\pi(\xi)}^{(l,m)})^2, \tag{3}$$

где ξ – индекс суммирования, используемый локально только в данной формуле ($\xi = 1, \dots, s$).

Шаг 3. Найти минимальный индекс s^* , удовлетворяющий условию сохранения заданной доли энергии:

$$\frac{E_{\Sigma}^{(l,m)}(s^*)}{E_l^{(m)}} \geq \alpha_l. \tag{4}$$

Шаг 4. Сформировать вектор пороговых коэффициентов:

$$\tilde{c}_k^{(l,m)} = \begin{cases} c_k^{(l,m)}, & \text{если } \left| c_k^{(l,m)} \right| \geq \left| c_{\pi(s^*)}^{(l,m)} \right|, \\ 0, & \text{иначе.} \end{cases} \tag{5}$$

Число сохранённых коэффициентов на уровне l кадра m обозначается как $s_l^{*(m)}$ и равно найденному значению s^* . Эта величина определяет степень сжатия на уровне l и используется для количественной оценки разрежённости сигнала после обработки.

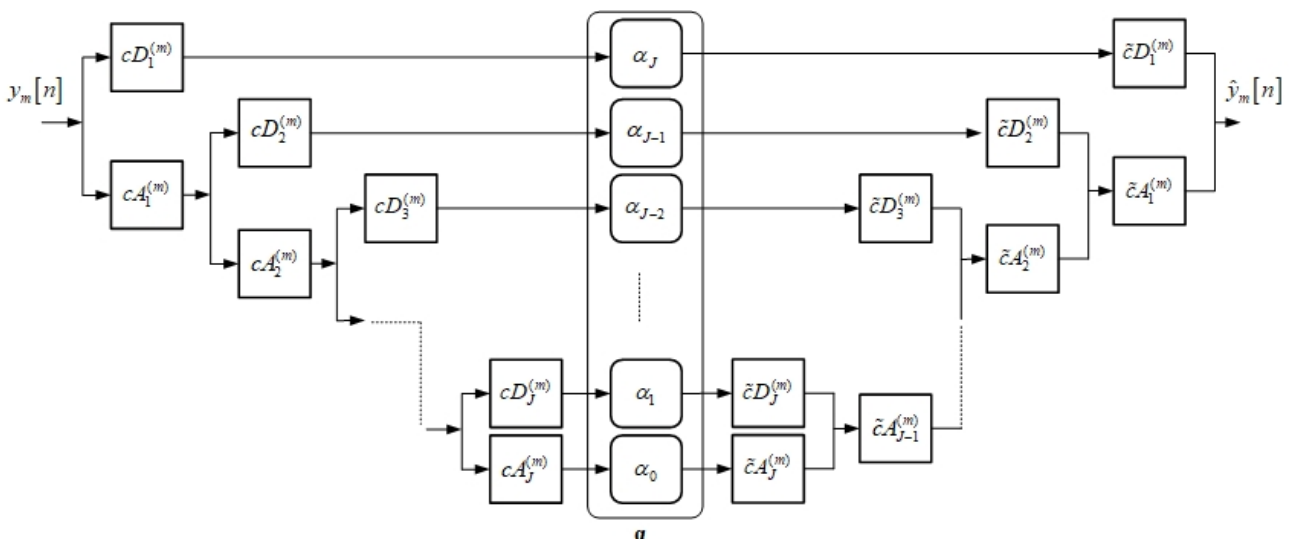


Рисунок 1 – Схема энергетического порогового сжатия сигнала на основе ДВП
Figure 1 – Signal compression scheme based on DWT with energy-based thresholding

Данная процедура применяется независимо к каждому из $J + 1$ уровней разложения (аппроксимация и J уровней деталей), и в результате для каждого кадра m формируется полный вектор пороговых коэффициентов \tilde{C}_m , содержащий все отобранные коэффициенты

$\tilde{c}_k^{(l,m)}$ по всем уровням l . Вектор энергетических порогов $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_J]^T \in [0, 1]^{J+1}$ полностью определяет стратегию сжатия сигнала: каждый порог α_l задаёт целевую долю энергии, сохраняемую на уровне l для всех кадров, что обеспечивает согласованное управление степенью сжатия по частотным полосам. Восстановление сигнала выполняется через обратное дискретное вейвлет-преобразование (ОДВП) для каждого кадра сигнала m :

$$\hat{y}_m[n] = \text{IDWT}_J(\tilde{C}_m). \quad (6)$$

с использованием дуальных фильтров синтеза $\tilde{h}[n]$ и $\tilde{g}[n]$.

Процедура покадровой обработки сигнала с энергетическим пороговым сжатием, описанная выше, показана на рисунке 1. Схема иллюстрирует полный цикл: от разложения сигнала до восстановления, с указанием вектора порогов α как ключевого элемента управления степенью сжатия.

Метрики эффективности сжатия

– Степень сжатия количественно оценивается через коэффициент разреженности:

$$\rho(\alpha) = \left(1 - \frac{K_{total}}{M_{coeff}}\right) \cdot 100\%, \quad (7)$$

где $M_{coeff} = M \cdot N_{fr}$ – общее число вейвлет-коэффициентов всех M кадров; $K_{total} = \sum_{m=1}^M \sum_{l=0}^J S_l^{*(m)}$ – общее число ненулевых коэффициентов после порогового сжатия. Высокие значения коэффициента разреженности $\rho(\alpha)$, близкие к 100 %, свидетельствуют об эффективном сжатии речевого сигнала.

– SegSNR – сегментное отношение сигнал-шум. Сигнал разбивается на R непересекающихся сегментов длиной L_{seg} . Для каждого сегмента $r = 1, 2, \dots, R$ вычисляется локальное отношение сигнал-шум:

$$\text{SNR}_r = 10 \log_{10} \left(\frac{\sum_{n \in S_r} x^2[n]}{\sum_{n \in S_r} (x[n] - \hat{x}[n])^2} \right) [\text{дБ}], \quad (8)$$

где S_r – множество индексов отсчётов r сегмента. Итоговое значение сегментного отношения сигнал-шум SegSNR вычисляется как среднее арифметическое по всем допустимым сегментам:

$$Q_1(\alpha) = \frac{1}{R} \sum_{r=1}^R \text{SNR}_r [\text{дБ}]. \quad (9)$$

Сегменты с энергией сигнала менее 10^{-10} исключаются, или им присваивается $\text{SNR}_r = 60 [\text{дБ}]$. Метрика SegSNR обеспечивает лучшую корреляцию с субъективным восприятием качества речи по сравнению глобальным отношением сигнал-шум, поскольку учитывает локальные искажения, критичные для разборчивости.

– ViSQOL – перцептуальная метрика качества речи, имитирующая субъективное восприятие:

$$Q_2(\alpha) = \text{ViSQOL}(x, \hat{x}) \in [1, 5], \quad (10)$$

где выходное значение $Q_2(\alpha)$ соответствует шкале MOS (Mean Opinion Score).

Методика оптимизации энергетических порогов с использованием алгоритма роя частиц

Алгоритм роя частиц PSO относится к классу метаэвристических методов глобальной оптимизации и основан на моделировании коллективного поведения агентов при поиске оптимальных решений в многомерном пространстве поиска [2]. Каждая частица роя представ-

ляет собой потенциальное решение задачи и характеризуется текущим положением и скоростью, определяющими её перемещение в пространстве параметров.

В целях достижения максимального коэффициента сжатия речи требуется балансировать между разрежённостью вейвлет-коэффициентов и качеством восстановления сигнала. Традиционные подходы с фиксированными порогами не адаптируются к особенностям сигнала. В данной работе вектор энергетических порогов $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_J]^T \in [0, 1]^{J+1}$ настраивается адаптивно с помощью PSO, что позволяет решать многокритериальную задачу без аналитического моделирования сложных зависимостей. Задача оптимизации формулируется как поиск оптимального вектора энергетических порогов α^{opt} , обеспечивающего максимизацию интегрального критерия эффективности при выполнении ограничений на качество:

$$\alpha^{opt} = \max_{\alpha \in [0, 1]^{J+1}} F(\alpha), \quad (11)$$

где целевая функция имеет следующий вид:

$$F(\alpha) = w_1 f_p(\alpha) + w_2 f_{Q_1}(\alpha) + w_3 f_{Q_2}(\alpha), \quad (12)$$

где: $f_p(\alpha) = \rho(\alpha)$ – коэффициент разрежённости, определённый в (7), выраженный в процентах; $f_{Q_1}(\alpha), f_{Q_2}(\alpha)$ – нормализованные метрики качества SegSNR и ViSQOL соответственно; весовые коэффициенты $w_1, w_2, w_3 \geq 0$ удовлетворяют условию $w_1 + w_2 + w_3 = 1$.

При этом должны выполняться ограничения: $Q_1(\alpha) \geq Q_1^{min}, Q_2(\alpha) \geq Q_2^{min}$, где $Q_1(\alpha)$ и $Q_2(\alpha)$ – метрики качества SegSNR и ViSQOL соответственно.

Алгоритм PSO работает с роем из N_p частиц. Каждая частица i ($i = 1, 2, \dots, N_p$) в итерации t характеризуется следующими параметрами:

– положение частицы $\alpha_i^{(t)} = [\alpha_{i,1}^{(t)}, \dots, \alpha_{i,D}^{(t)}]^T \in \mathbb{R}^D$, где $D = J + 1$ — размерность пространства поиска. Вектор $\alpha_i^{(t)}$ представляет текущий набор энергетических порогов;

– скорость частицы $\mathbf{v}_i^{(t)} = [v_{i,1}^{(t)}, \dots, v_{i,D}^{(t)}]^T \in \mathbb{R}^D$, определяющая направление и величину изменения положения;

– индивидуально лучшее положение частицы $\mathbf{p}_i^{(t)} \in \mathbb{R}^D$, соответствующее наилучшему значению целевой функции $F(\alpha)$, данной частицей за всё время работы алгоритма вплоть до итерации t ;

– глобально лучшее положение роя: $\mathbf{g}^{(t)} \in \mathbb{R}^D$, при котором значение целевой функции $F(\alpha)$ является наилучшим среди всех частиц роя на итерации t ;

Структура предложенного алгоритма оптимизации энергетических порогов на основе алгоритма PSO представлена на рисунке 2. Алгоритм PSO состоит из следующих шагов.

Шаг 1. Инициализация. На этапе инициализации алгоритма $t = 0$ положения и скорости всех частиц инициализируются случайным образом в пределах заранее заданных границ пространства поиска:

$$\alpha_i^{(0)} \sim U([0, 1]^{J+1}), \mathbf{v}_i^{(0)} \sim U([-v_{max}, v_{max}]^{J+1}), \quad (13)$$

где $U([a, b]^D)$ обозначает равномерное распределение в D -мерном гиперкубе $[a, b]^D$.

Индивидуально лучшие положения частиц на начальном этапе принимаются равными их начальным положениям:

$$\mathbf{p}_i^{(0)} = \alpha_i^{(0)}, i = 1, 2, \dots, N_p, \quad (14)$$

а глобально лучшее положение роя определяется по следующей формуле:

$$\mathbf{g}^{(0)} = \arg \max_{j=1, \dots, N_p} F(\mathbf{p}_j^{(0)}). \quad (15)$$

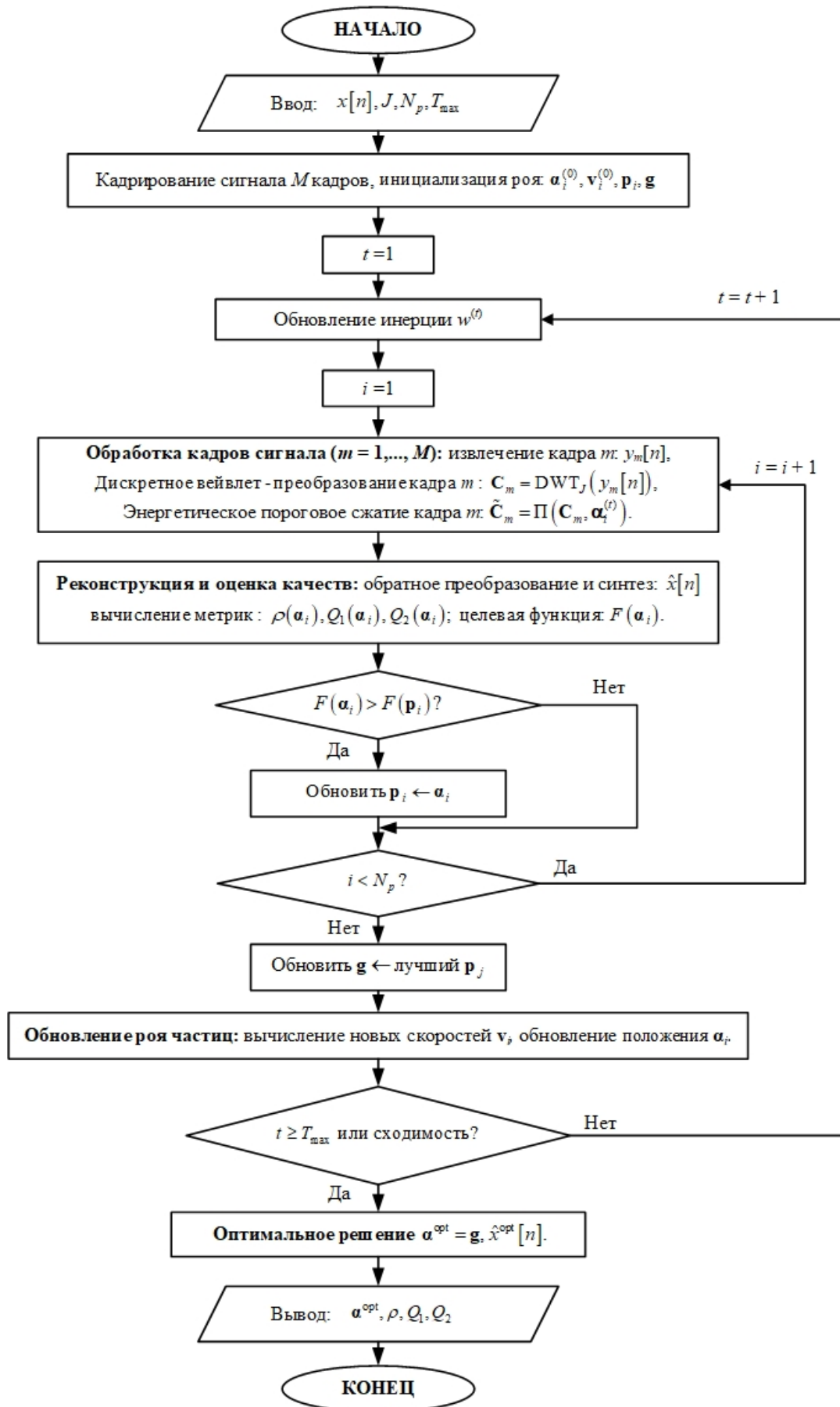


Рисунок 2 – Блок-схема алгоритма оптимизации энергетических порогов на основе PSO
 Figure 2 – Flowchart of PSO-based energy threshold optimization algorithm

Шаг 2. Итеративное обновление. На каждой последующей итерации $t = 1, 2, \dots, T_{\max}$ выполняется обновление коэффициента инерции, скоростей и положений частиц.

– Коэффициент инерции, определяющий баланс между глобальным и локальным поиском, изменяется линейно по закону:

$$w^{(t)} = w_{\max} - \frac{(w_{\max} - w_{\min})t}{T_{\max}}. \quad (16)$$

– Обновление скорости. Скорость i -й частицы в d -й координате обновляется согласно стандартному уравнению PSO:

$$v_{i,d}^{(t+1)} = w^{(t)} \cdot v_{i,d}^{(t)} + c_1 \cdot r_{1,d}^{(t)} \cdot (p_{i,d}^{(t)} - \alpha_{i,d}^{(t)}) + c_2 \cdot r_{2,d}^{(t)} \cdot (g_d^{(t)} - \alpha_{i,d}^{(t)}), \quad (17)$$

где c_1 и c_2 – когнитивный и социальный коэффициенты, определяющие вес личного и коллективного опыта; $r_{1,d}^{(t)}$, $r_{2,d}^{(t)}$ – независимые случайные числа в интервале $[0,1]$, генерируемые на каждой итерации для каждой координаты.

– Обновление положения. Новое положение частицы i в измерении d определяется как:

$$\alpha_{i,d}^{(t+1)} = \alpha_{i,d}^{(t)} + v_{i,d}^{(t+1)}. \quad (18)$$

Для обеспечения соблюдения ограничений $[0,1]$ применяется гибридная стратегия обработки границ. Если после обновления $\alpha_{i,d}^{(t+1)} < 0$ или $\theta_{i,d}^{(t+1)} > 1$, то:

$$\alpha_{i,d}^{(t+1)} \leftarrow \begin{cases} 0, & \text{если } \theta_{i,d}^{(t+1)} < 0, \\ 1, & \text{если } \theta_{i,d}^{(t+1)} > 1, \\ \alpha_{i,d}^{(t+1)}, & \text{иначе,} \end{cases} \quad (19)$$

и соответствующая компонента скорости корректируется по правилу $v_{i,d}^{(t+1)} \leftarrow -\gamma \cdot v_{i,d}^{(t+1)}$, где $\gamma = 0,5$.

Обновление индивидуально и глобально лучших положений:

– Если новое положение частицы приводит к улучшению значения целевой функции, её индивидуально лучшее положение обновляется:

$$\mathbf{p}_i^{(t+1)} = \begin{cases} \mathbf{\alpha}_i^{(t+1)}, & \text{если } f(\mathbf{\alpha}_i^{(t+1)}) < f(\mathbf{p}_i^{(t)}), \\ \mathbf{p}_i^{(t)}, & \text{иначе.} \end{cases} \quad (20)$$

– Глобально лучшее положение роя обновляется как наилучшее среди всех индивидуально лучших положений:

$$\mathbf{g}^{(t+1)} = \arg \max_{j=1, \dots, N_p} F(\mathbf{p}_j^{(t+1)}). \quad (21)$$

Шаг 3. Завершение. Алгоритм завершается по достижении максимального числа итераций T_{\max} либо при выполнении критерия сходимости. Решением задачи считается вектор $\mathbf{\alpha}^{\text{opt}} = \mathbf{g}^{(T_{\max})}$.

Этот вектор содержит оптимальные значения энергетических порогов $\alpha_l^{\text{opt}} = g_l^{(T_{\max})}$ для каждого уровня разложения $l = 1, 2, \dots, J$, обеспечивающие максимальную степень сжатия при соблюдении заданных минимальных порогов качества Q_1^{\min} и Q_2^{\min} .

Экспериментальные исследования предложенного алгоритма

Экспериментальное исследование выполнено в два этапа: предварительный этап по выбору типа вейвлета и числа уровней разложения, и основной этап по верификации предложенного метода оптимизации энергетических порогов.

Предварительный этап: выбор типа вейвлета и числа уровней разложения

Эффективность вейвлет-сжатия существенно зависит от выбора вейвлет-базиса и числа уровней разложения. Для исследования выбрано 17 вейвлетов из четырёх семейств, наиболее

часто используемых при обработке речевых сигналов: Добеши (db2, db4, db6, db8, db10); Симлеты (sym4, sym6, sym8, sym10); Койфлеты (coif2, coif3, coif4, coif5); биортогональные вейвлеты (bior2.2, bior3.3, bior4.4, bior5.5).

Предварительное исследование проведено на выборке из 100 записей речи частотой 8 кГц, отобранных в соответствии с требованиями стандарта ГОСТ Р 50840-95 [8]. На этом этапе для каждой комбинации (вейвлет, J), где $J \in \{2, 3, 4, 5, 6\}$, выполнялось J -уровневое дискретное вейвлет-преобразование с последующим энергетическим пороговым сжатием при фиксированных порогах $\alpha_l = 0,9$ для всех уровней. Эффективность каждого вейвлета оценивалась по интегральному критерию (11) с весовыми коэффициентами $w_1 = 0,5, w_2 = 0,25, w_3 = 0,25$, реализующими сбалансированный подход между степенью разреженности и качеством восстановления речевого сигнала:

$$F = 0,5f_p + 0,25f_{Q_1} + 0,25f_{Q_2}. \quad (22)$$

Для каждого вейвлета выбирается то число уровней разложения J (от 2 до 6), при котором достигается наивысшее значение интегрального критерия F . Затем вейвлеты сравниваются между собой по этим наилучшим значениям F , полученным при их собственных оптимальных J . Для каждого вейвлета оптимальное число уровней определяется согласно следующей формуле:

$$J_{opt} = \arg \max_{J \in \{2, 3, \dots, 6\}} F. \quad (23)$$

Итоговое сравнение вейвлетов выполняется по максимальным значениям F (вейвлет, J_{opt}).

На рисунке 3 представлены зависимости интегрального критерия эффективности F от числа уровней вейвлет-разложения J для четырёх репрезентативных вейвлетов, принадлежащих различным семействам: Добеши (кривая 1 – db4), Симлеты (кривая 2 – sym6), Койфлеты (кривая 3 – coif3) и биортогональных вейвлетов (кривая 4 – bior3.3).

Из анализа приведенных на рисунке 3 зависимостей следует, что для всех исследованных вейвлетов оптимальное число уровней разложения составляет $J_{opt} = 4$, при котором достигается максимум интегрального критерия ($F = 69,11$ для db4). Недостаточное разложение ($J = 2, J = 3$) снижает эффективность на 3...8 % из-за низкого частотного разрешения. Избыточное разложение ($J > 4$) приводит к деградации на 2...4 % вследствие критически малого числа аппроксимирующих коэффициентов и пренебрежимо малой амплитуды высокочастотных деталей. Инвариантность оптимума $J = 4$ к типу вейвлета указывает на определяющую роль спектральной структуры речевого сигнала.

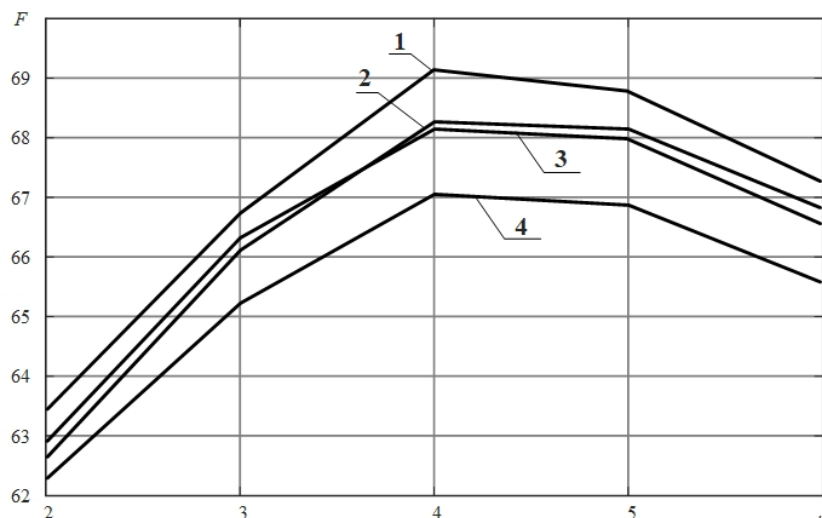


Рисунок 3 – Зависимости интегрального критерия эффективности F от числа уровней вейвлет-разложения J для различных типов вейвлетов
Figure 3 – Dependences of integral efficiency criterion F on the number of wavelet decomposition levels J for different wavelet types

В таблице 1 приведено сравнение эффективности сжатия речи при использовании различных вейвлетов при $J=4$. Как следует из таблицы 1, при одинаковых уровнях разложения $J=4$ вейвлет Добеши db4 демонстрирует наилучшие значения интегрального критерия ($F = 69,11$). Преимущество db4 наблюдается по всем использованным показателям: степень сжатия составляет $\rho = 81,23\%$ (максимальное значение), SegSNR достигает $Q_1 = 11,34$ дБ, а показатель ViSQOL $Q_2 = 3,81$ балла, что также является наилучшим результатом среди сравниваемых вейвлетов. Таким образом, выбор вейвлета Добеши db4 и $J=4$ обоснован результатами систематического экспериментального исследования и обеспечивает оптимальные характеристики разложения и компрессии речевого сигнала для последующих этапов обработки.

Таблица 1 – Сравнение эффективности сжатия речи при использовании различных вейвлетов при $J=4$ уровнях разложения

Table 1 – Comparison of speech compression efficiency using different wavelets at $J=4$ decomposition levels

Вейвлет	Степень разреженности (ρ , %)	SegSNR (Q_1 , дБ)	ViSQOL (Q_2 , балл)	F
db2	79,84	11,02	3,72	67,70
db4	81,23	11,34	3,81	69,11
db6	80,95	11,28	3,79	68,83
db8	80,45	11,15	3,75	68,27
db10	79,92	11,05	3,71	67,72
sym4	80,89	11,25	3,78	68,72
sym6	80,52	11,19	3,76	68,38
sym8	80,21	11,12	3,74	68,07
sym10	79,78	11,01	3,70	67,56
coif2	80,68	11,22	3,77	68,54
coif3	80,28	11,14	3,75	68,17
coif4	79,85	11,06	3,72	67,74
coif5	79,42	10,95	3,68	67,23
bior2.2	79,52	10,98	3,69	67,36
bior3.3	79,18	10,89	3,67	67,02
bior4.4	78,95	10,82	3,65	66,74
bior5.5	78,62	10,75	3,63	66,42

Основной этап: верификация метода оптимизации

На основном этапе выполнена экспериментальная верификация предложенного метода адаптивной оптимизации энергетических порогов на основе алгоритма PSO с использованием выбранных параметров вейвлета Добеши db4 и $J=4$ уровней разложения. Для экспериментальной верификации предложенного метода проведены численные исследования на речевых данных с целью определения оптимальных параметров алгоритма PSO, анализа его сходимости и сравнения с существующими методами порогового сжатия на основе дискретного вейвлет-преобразования.

Эксперименты выполнены на той же выборке, что и в предварительном исследовании, – 100 речевых файлов, отобранных в соответствии с требованиями стандарта ГОСТ Р 50840-95.

Параметры алгоритма PSO установлены следующим образом: размер роя $N_p = 200$; максимальное число итераций $T_{\max} = 100$; коэффициент инерции изменяется линейно от $w^{(0)} = 0,9$ до $w^{(T_{\max})} = 0,3$; когнитивный и социальный коэффициенты установлены равными $c_1 = c_2 = 2,0$; максимальная скорость ограничена значением $v_{\max} = 0,15$; область поиска для всех компонент вектора порогов установлена как $\alpha_l \in [0,1]$, где $l = 0,1,\dots,J = 4$; веса целевой функции $w_1 = 0,7$, $w_2 = 0,15$, $w_3 = 0,15$; ограничения качества: $Q_1(\mathbf{a}) \geq Q_1^{\min} = 9,5$ дБ (по

SegSNR) и $Q_2(\alpha) \geq Q_2^{\min} = 3,5$ балла (по шкале MOS) в соответствии с рекомендациями ITU-T P.800 [6]. Критерий ранней остановки: алгоритм завершается при отсутствии улучшения глобального наилучшего решения $\mathbf{g}^{(t)}$ в течение 20 последовательных итераций при изменении целевой функции менее $\varepsilon = 10^{-6}$.

На рисунке 4 представлены результаты многокритериальной оптимизации энергетических порогов вейвлет-преобразования с использованием алгоритма PSO, отражающие изменение ключевых показателей эффективности по итерациям.

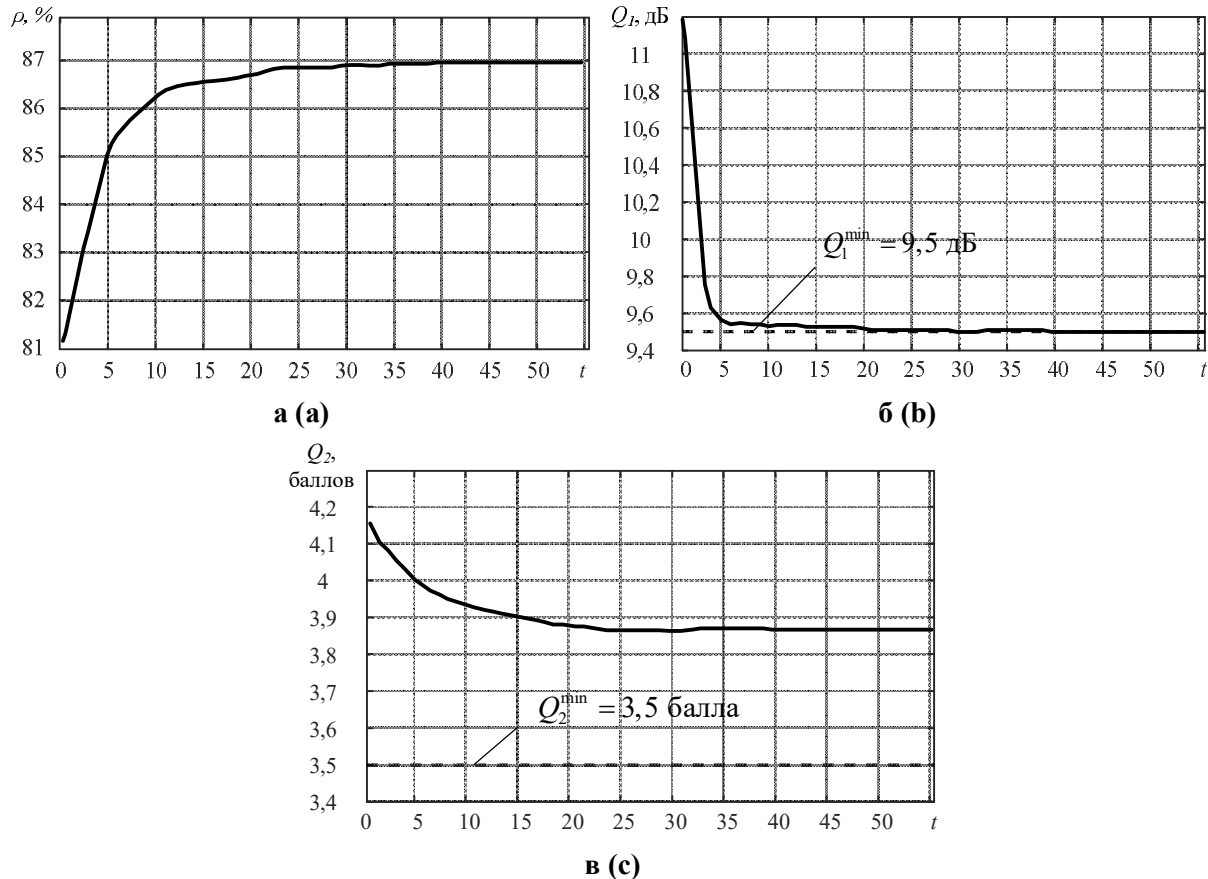


Рисунок 4 – Результаты многокритериальной оптимизации методом PSO: зависимости от номера итерации t (а) степени разреженности; (б) SegSNR; (в) ViSQOL
Figure 4 – Multi-objective optimization results obtained using PSO method: dependencies on iteration number t (a) sparsity level; (b) SegSNR; (c) ViSQOL

На рисунке 4, *а* показана зависимость степени разреженности ρ от номера итерации t . На начальном этапе (итерации 0...10) наблюдается интенсивный рост ρ от 81 % до 86 %, затем темп замедляется (итерации 11...35), и с итерации 36 устанавливается стационарный режим $\rho = 87$ %; алгоритм останавливается на 48-й итерации. На рисунке 4, *б* представлена зависимость SegSNR от t с горизонтальной линией минимального требования $Q_1^{\min} = 9,5$ дБ. После начального превышения ограничения (11,2 дБ) метрика снижается и стабилизируется на уровне $9,50 \pm 0,05$ дБ (итерации 13...55), точно соответствуя минимальному требованию. На рисунке 4, *в* показана эволюция перцептуального качества ViSQOL. Начальное значение ViSQOL = 4,18 (субъективно «хорошо» по шкале MOS) снижается к итерации 5 до 4,0 под действием целевой функции, которая приоритизирует рост ρ при допустимом падении качества. После итерации 25 ViSQOL стабилизируется на уровне $3,87 \pm 0,01$, что соответствует оценке «удовлетворительно–хорошо», и обеспечивает запас $\Delta Q_2 = 0,37$ относительно минимального требования $Q_2^{\min} = 3,5$ балла. Алгоритм PSO сходится к оптимальному вектору

$\alpha^{\text{opt}} = [0,989 \ 0,954 \ 0,584 \ 0,447 \ 0,171]^T$. Высокие пороги для низкочастотных уровней (сохранение энергии) и низкие для высокочастотных (сильное сжатие) отражают специфику энергетического распределения речи.

Для оценки эффективности проведено сравнение с классическими методами. В таблице 2 приведено сравнение предложенного метода PSO с тремя базовыми подходами к пороговой обработке вейвлет-коэффициентов, усреднённое по выборке из 100 речевых сигналов. Для всех методов использовались идентичные параметры вейвлет-преобразования.

Таблица 2 – Сравнительная эффективность методов порогового сжатия речи

Table 2 – Comparative efficiency of wavelet thresholding methods for speech compression

Метод	Степень разреженности (ρ , %)	SegSNR (Q_1 , дБ)	ViSQOL (Q_2 , балл)
Универсальный порог Донохо	76,32	9,18	3,45
Адаптивный SUREShrink	79,47	10,81	3,68
Энергетический порог с фиксированными значениями $\alpha_i = 0,9$	81,23	11,34	3,81
Энергетический порог, оптимизированный с помощью алгоритма PSO	87,00	9,5	3,87

Анализ результатов, представленных в таблице 2, позволяет сделать следующие выводы. Классические методы – универсальный порог Донохо и адаптивный SUREShrink – демонстрируют ограниченную эффективность в задаче сжатия речи. Универсальный порог обеспечивает наименьшую разреженность ($\rho = 76,32\%$) и не удовлетворяет заданному ограничению по SegSNR ($Q_1 = 9,18 < Q_1^{\text{min}} = 9,5$ дБ). Адаптивный SUREShrink улучшает показатели (SegSNR = 10,81 дБ), но разреженность уступает энергетическим методам на 1,76...7,53 %. Это подтверждает, что амплитудные пороговые стратегии неэффективны для речевых сигналов, где информативность компонент определяется их энергетическим вкладом, а не только амплитудой. Энергетический метод с фиксированным порогом $\alpha_i = 0,9$ превосходит амплитудные подходы (SegSNR = 11,34 дБ, ViSQOL = 3,81 балла), однако фиксированный порог не учитывает частотно-зависимое распределение энергии речи, что снижает эффективность сжатия. Предложенный метод PSO обеспечивает максимальную степень сжатия (87 %) при точном соблюдении минимально допустимых уровней качества (SegSNR = 9,5 дБ, ViSQOL = 3,87 балла), что подтверждает его преимущество по критерию эффективности сжатия в условиях заданных ограничений.

Заключение

В работе предложен метод адаптивной оптимизации энергетических порогов вейвлет-преобразования для сжатия речи на основе алгоритма PSO. Метод обеспечивает автоматический выбор доли сохраняемой энергии на каждом уровне разложения, максимизируя степень сжатия при соблюдении заданных ограничений на SegSNR и ViSQOL.

Экспериментальные исследования, выполненные на речевых сигналах, записанных в соответствии с ГОСТ Р 50840-95, показали, что при использовании вейвлета Добеши db4 и четырёх уровней разложения предложенный подход обеспечивает степень сжатия 87 %, при этом достигая точно заданного минимального уровня качества: SegSNR = 9,5 дБ и ViSQOL = 3,9 балла. Предложенный подход превосходит универсальный порог Донохо на 14 %, SUREShrink – на 9 %, энергетический метод с фиксированными порогами – на 7 % при одновременном повышении перцептуального качества ViSQOL до 3,9 балла.

Полученные значения энергетических порогов могут служить основой для сжатия речи с частотой 8 кГц. Результаты открывают перспективы для расширения на широкополосную речь, покадровую оптимизацию и совместную оптимизацию с последующими этапами кодирования.

Библиографический список

1. **Mallat S.G.** A theory for multiresolution signal decomposition: the wavelet representation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1989. Vol. 11, No. 7. P. 674-693.
2. **Kennedy J., Eberhart R.** Particle swarm optimization // Proceedings of ICNN'95 – International Conference on Neural Networks. Perth, WA, Australia, 1995. Vol. 4. P. 1942-1948.
3. **Donoho D.L., Johnstone I.M.** Ideal spatial adaptation by wavelet shrinkage // Biometrika. 1994. Vol. 81. No. 3. P. 425-455.
4. **Daubechies I.** Ten lectures on wavelets. Philadelphia: Society for Industrial and Applied Mathematics, 1992. 357 p.
5. ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. М.: Издательство стандартов. 1995.
6. ITU-T Recommendation P.800. Methods for subjective determination of transmission quality. Geneva: International Telecommunication Union, 1996.
7. **Hines A., Gillen E., Skoglund J., Kokaram A., Harte N.** ViSQOL: An objective audio quality metric for low bitrate codecs // The Journal of the Acoustical Society of America. 2015. Vol. 137. No. 6. P. EL449-EL455.
8. **Дмитриев В.Т., Ву Хоанг Шон** Применение трехканальной модификации алгоритма Хургина – Яковлева в алгоритмах первичного кодирования речевых сигналов // Вестник Рязанского государственного радиотехнического университета. 2024. № 88. С. 3-14.
9. **Дмитриев В.Т.** Первичные кодеки речевых сигналов на основе вейвлет-пакетного разложения и модификации алгоритма Хургина – Яковлева // Цифровая обработка сигналов. 2024. № 90. С. 46-51.
10. **Дмитриев В.Т., Филипенко О.С.** Анализ ошибки восстановления речевых сигналов при использовании различных вейвлет-пакетных разложений // Современные технологии в науке и образовании – СТНО-2017 [текст]: сб. тр. междунар. науч.-техн. форума: в 10 т. Т.1 / под общ. ред. О.В. Милловзорова. – Рязань: Рязан. гос. радиотехн. ун-т, 2018. С. 176-178.
11. **Дмитриев В.Т., Харланова Е.А.** Алгоритм кодирования речевых сигналов на основе представления Хургина – Яковлева и вейвлет-пакетного разложения в системах компьютерной телефонии // Вестник Рязанского государственного радиотехнического университета. 2010. № 31. С. 98-101.
12. **Дмитриев В.Т., Смирнов М.С.** Исследование помехоустойчивой и защищенной системы передачи речевых сигналов на основе представления Хургина – Яковлева // Вестник Рязанского государственного радиотехнического университета. 2022. № 82. С. 27-37.
13. **Дмитриев В.Т., Лантратов С.Ю.** Адаптивный алгоритм кодирования на основе кодека CELP и модификации алгоритма Хургина-Яковлева // Вестник Рязанского государственного радиотехнического университета. 2024. № 88. С. 21-30.

UDK 621.395

OPTIMIZATION OF ENERGY THRESHOLDS IN WAVELET TRANSFORM FOR SPEECH SIGNAL COMPRESSION BASED ON PARTICLE SWARM OPTIMIZATION

V. T. Dmitriev, Dr. in technical sciences, department of radio control and communication, Head of the department, RSREU, Ryazan, Russia;

orcid.org/0000-0001-5521-6886, e-mail: vol77@rambler.ru

Vu Hoang Son, post-graduate student, RSREU, Ryazan, Russia;

orcid.org/0009-0004-7428-5296, e-mail: vuhoangson.adaf@gmail.com

An adaptive method for optimizing energy thresholds in discrete wavelet transform (DWT) based on Particle Swarm Optimization (PSO) algorithm for speech signal compression has been proposed and investigated. The method enables automatic selection of optimal energy retention ratios at each level of wavelet decomposition. Optimization is performed to maximize the compression ratio while simultaneously satisfying predefined constraints on the quality of reconstructed speech at a receiver. The quality of reconstructed speech is evaluated using two objective metrics: segmental signal-to-noise ratio (SegSNR) and perceptual

speech quality metric ViSQOL. Experimental studies conducted on standard speech signals recorded in accordance with GOST R 50840-95 demonstrate that the method proposed achieves a compression ratio of 87 % while maintaining high reconstruction quality: SegSNR = 9,5 dB and ViSQOL = 3,9 points. According to integral efficiency criterion, the proposed approach outperforms classical methods by 7...14 %.

Keywords: discrete wavelet transform, particle swarm optimization, energy-based thresholding, parameter optimization, SegSNR, ViSQOL, speech signals.

DOI: 10.21667/1995-4565-2026-95-3-15

References

1. **Mallat S. G.** A theory for multiresolution signal decomposition: the wavelet representation // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1989, vol. 11, no. 7, pp. 674-693.
2. **Kennedy J., Eberhart R.** Particle swarm optimization. *Proceedings of ICNN'95 – International Conference on Neural Networks*. Perth, WA, Australia. 1995, vol. 4, pp. 1942-1948.
3. **Donoho D.L., Johnstone I.M.** Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. 1994, vol. 81, no. 3, pp. 425-455.
4. **Daubechies I.** Ten lectures on wavelets. Philadelphia: Society for Industrial and Applied Mathematics, 1992. 357 p.
5. ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. М.: Издательство стандартов, 1995.
6. ITU-T Recommendation P.800. Methods for subjective determination of transmission quality. Geneva: International Telecommunication Union, 1996.
7. **Hines A., Gillen E., Skoglund J., Kokaram A., Harte N.** ViSQOL: An objective audio quality metric for low bitrate codecs. *The Journal of the Acoustical Society of America*. 2015, vol. 137, no. 6, pp. EL449–EL455.
8. **Dmitriev V.T., Vu Hoang Son** Primenenie trekhkanal'noi modifikatsii algoritma Khurgina–Yakovleva v algoritmakh pervichnogo kodirovaniya rechevykh signalov. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2024, no. 88, pp. 3-14. (in Russian).
9. **Dmitriev V.T.** Pervichnye kodeki rechevykh signalov na osnove veivlet-paketnogo razlozheniya i modifikatsii algoritma Khurgina–Yakovleva. *Tsifrovaya obrabotka signalov*. 2024, no. 4, pp. 46-51. (in Russian).
10. **Dmitriev V.T., Filipenko O.S.** Analiz oshibki vosstanovleniya rechevykh signalov pri ispol'zovanii razlichnykh veivlet-paketnykh razlozhenii. *Sovremennye tekhnologii v nauke i obrazovanii – STNO-2017: sbornik trudov mezhdunarodnogo nauchno-tekhnicheskogo foruma: v 10 t. Vol. 1 / pod obshch. red. O.V. Milovzorova*. Ryazan': Ryazanskii gosudarstvennyi radiotekhnicheskii universitet, 2018, pp. 176-178. (in Russian).
11. **Dmitriev V.T., Kharlanova E.A.** Algoritm kodirovaniya rechevykh signalov na osnove predstavleniya Khurgina – Yakovleva i veivlet-paketnogo razlozheniya v sistemakh komp'yuternoii telefonii. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2010, no. 31, pp. 98-101. (in Russian).
12. **Dmitriev V.T., Smirnov M.S.** Issledovanie pomekhoustoychivoi i zashchishchennoi sistemy peredachi rechevykh signalov na osnove predstavleniya Khurgina – Yakovleva. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2022, no. 82, pp. 27-37. (in Russian).
13. **Dmitriev V.T., Lantratov S.Yu.** Adaptivnyi algoritm kodirovaniya na osnove kodeka CELP i modifikatsii algoritma Khurgina – Yakovleva. *Vestnik Ryazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2024, no. 88, pp. 21-30. (in Russian).