

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ  
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»**  
( **Н И У « Б е л Г У »** )

ИНСТИТУТ ИНЖЕНЕРНЫХ ТЕХНОЛОГИЙ И ЕСТЕСТВЕННЫХ НАУК

КАФЕДРА ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫХ  
СИСТЕМ И ТЕХНОЛОГИЙ

**ВЫЯВЛЕНИЕ УЧАСТКОВ АКТИВНОСТИ РЕЧИ В ЗАДАЧАХ СТЕГАНОГРАФИИ**

Магистерская диссертация  
обучающегося по направлению подготовки 11.04.02 Инфокоммуникационные технологии  
и системы связи,  
магистерская программа «Системы и устройства радиотехники и связи»  
очной формы обучения, группы 07001532  
Бухало Дениса Сергеевича

Научный руководитель  
канд. техн. наук,  
доцент кафедры  
Информационно-  
телекоммуникационных  
систем и технологий  
НИУ «БелГУ» Прохоренко Е.И.

Рецензент  
Ведущий инженер электросвязи  
участка систем коммутации № 1  
г. Белгорода Белгородского  
филиала ПАО «Ростелеком»  
Уманец С.В.

БЕЛГОРОД 2017

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
ГЛАВА 1 ОСНОВНЫЕ ПОЛОЖЕНИЯ ТЕОРИИ ОБНАРУЖЕНИЯ АКТИВНОСТИ РЕЧИ	5
ГЛАВА 2 ОБЗОР МЕТОДОВ КОДИРОВАНИЯ РЕЧЕВОГО СИГНАЛА	10
2.1 VAD на основе сравнения с энергетическим порогом	10
2.2 VAD на основе коэффициентов линейного предсказания	12
2.3 Метод обнаружения пауз на основе субполосного анализа	14
ГЛАВА 3 ИССЛЕДОВАНИЕ АЛГОРИТМОВ ВЫЯВЛЕНИЯ АКТИВНОСТИ РЕЧИ	18
3.1 Сравнительная оценка работоспособности методов обнаружения пауз	18
3.2 Оценка выявления исключения пауз на ёмкость речевого материала	27
3.3 Оценка выявления точности определения границы пауза/речь на вероятность ошибки декодирования	29
ЗАКЛЮЧЕНИЕ	33
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	35
ПРИЛОЖЕНИЕ А	40
ПРИЛОЖЕНИЕ Б	42
ПРИЛОЖЕНИЕ В	43
ПРИЛОЖЕНИЕ Г	45
ПРИЛОЖЕНИЕ Д	47
ПРИЛОЖЕНИЕ Е	49
ПРИЛОЖЕНИЕ Ж	53
ПРИЛОЖЕНИЕ З	60

## **ВВЕДЕНИЕ**

Задача защиты информации от несанкционированного доступа решалась во все времена на протяжении истории человечества. Одним из решений таковой задачи является маскировка информации. Этим занимаются две науки – криптография и стеганография. Целью первой является скрывание информации за счёт шифрования, а второй – передача информации путём сохранения в тайне самого факта передачи. Доминирующую роль среди средств защиты информации играет шифрование сообщений с использованием методов криптографии, хоть они и не могут обеспечить необходимую защищенность каналов связи от несанкционированного доступа к информации со стороны третьих лиц. Поэтому в этих условиях защита информации, основанная на применении средств стеганографии, является целесообразной.

Развитие средств вычислительной техники дало новый толчок для развития стеганографии. Так одним из направлений современной стеганографии является цифровая стеганография [1,11,22], под которой подразумевают скрывание одних данных в другие посредством методов цифровой обработки сигналов. В качестве исходного материала используются цифровые данные, удобные для обмена, хранения и передачи. К таковым относятся речевые данные, которые являются наиболее удобной и естественной формой информационного обмена.

В голосовой связи речь может быть охарактеризована как прерывистый носитель информации из-за пауз, которые в режиме диалога могут занимать до 60% длительности исходных звукозаписей. В причину того, что внедрение информации при использовании стеганографических методов целесообразно осуществлять только в те отрезки речевых данных, где присутствует активная речь [1,11,22,31,36], то задача выявления участков активности речи является актуальной.

Поэтому целью данной работы является оценка эффективности методов выявления активности речи, позволяющих определять границы «активная речь/пауза» (т.е. осуществлять кодирование сигнала).

Для достижения цели научно-исследовательской работы поставлены следующие задачи:

- 1) обзор основных методов кодирования речевого сигнала;
- 2) сравнительная оценка работоспособности методов кодирования речевого сигнала;
- 3) оценка выявления исключения пауз на ёмкость речевого материала;
- 4) оценка выявления точности определения границ активной речь/пауза на вероятность ошибки декодирования.

Объектом исследования являются методы кодирования речевого сигнала, а предметом исследования – речевые данные.

В работе использованы методы статистического анализа, цифровой обработки речевых данных, линейной алгебры, субполосного анализа/синтеза и вычислительных экспериментов.

Данная работа состоит из Введения, трёх глав, Заключения и Приложений. Работа изложена на 64 страницах машинописного текста, включая 31 рисунок, 7 таблиц и список литературных источников из 47 наименований.

## ГЛАВА 1 ОСНОВНЫЕ ПОЛОЖЕНИЯ ТЕОРИИ ОБНАРУЖЕНИЯ АКТИВНОСТИ РЕЧИ

В голосовой связи речь может быть охарактеризована как прерывистый носитель информации из-за пауз, которые являются характерными особенностями данного вида сигналов по сравнению с другими мультимедийными сигналами, такими как видео, аудио и данные [6,47]. Интервалы, где голосовая информация присутствует, называют участками активной речи, а паузы между речевыми «всплесками» – участками пауз.

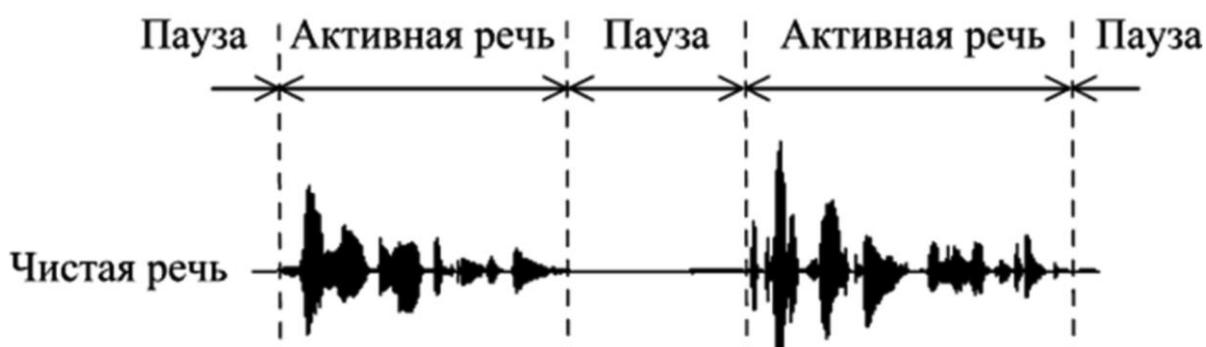


Рисунок 1.1 – Участки активной речи и паузы

Речевой сигнал, как и другие виды сложных сигналов произвольной формы, при их изучении и анализе, удобно представлять в виде некоторой комбинации более простых сигналов известной формы. Такой приём, когда произвольная функция раскладывается по различным системам детерминированных базисных функций, называется обобщённым спектральным анализом.

Основным инструментом, используемым в спектральном анализе, является пара преобразований Фурье, которые позволяют связать временной сигнал с его представлением в частотной области.

Прямое преобразование Фурье:

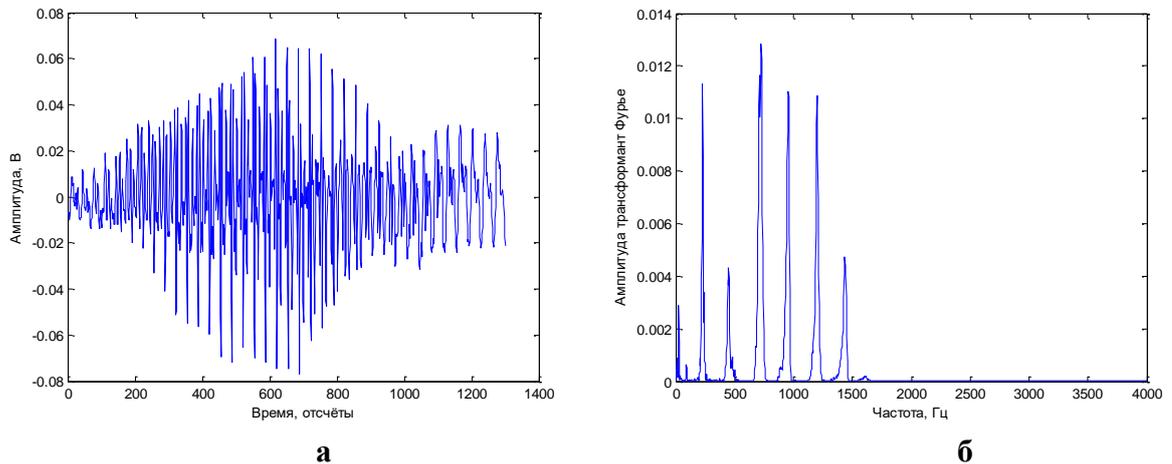
$$x(t) = \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} dt, \quad (1.1)$$

где  $X(\omega)$  - частотное представление сигнала временного сигнала  $x(t)$ .

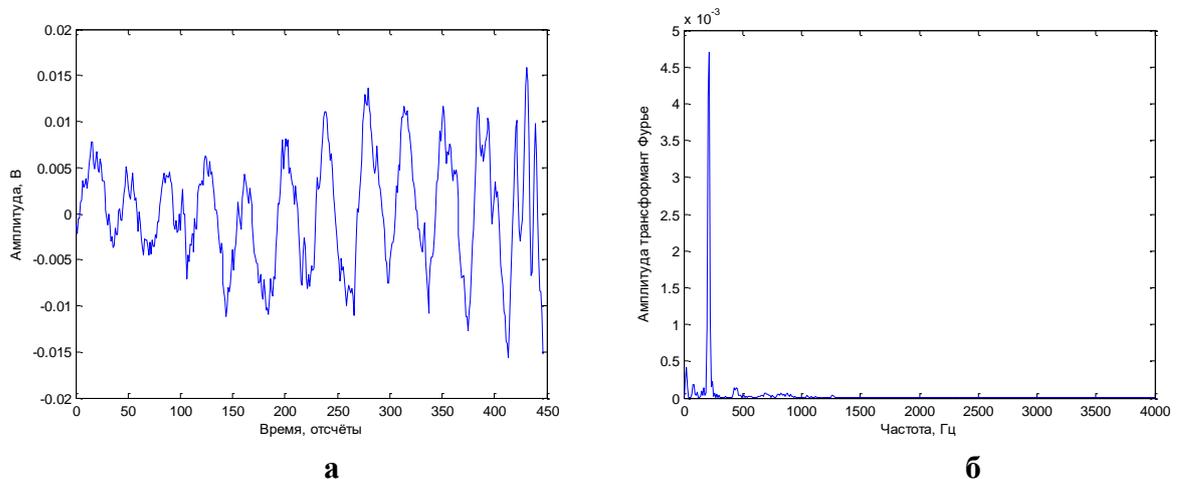
Обратное преобразование Фурье:

$$X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \quad (1.2)$$

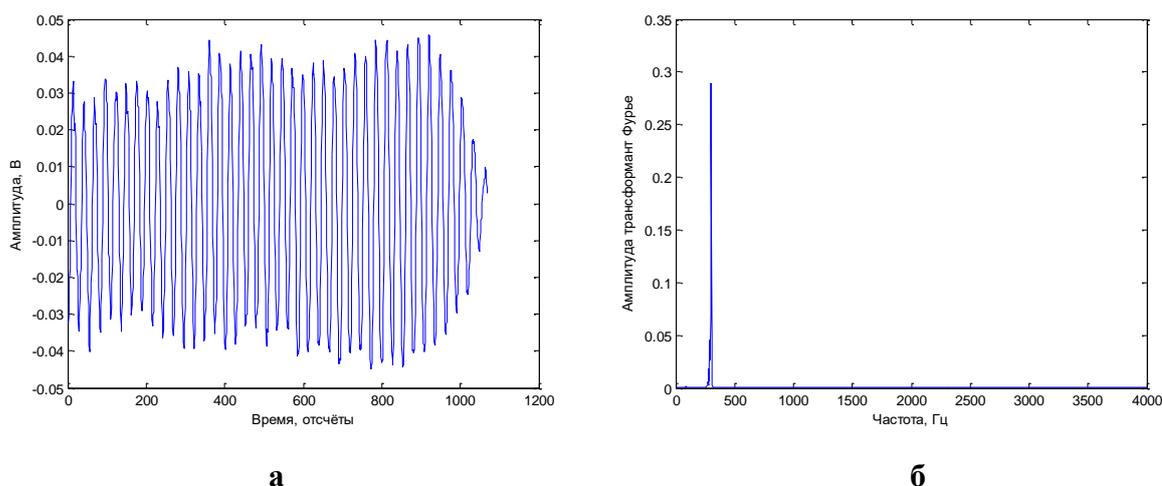
Соотношения 1.1 и 1.2 являются эквивалентными, поскольку могут быть получены друг из друга линейным преобразованием.



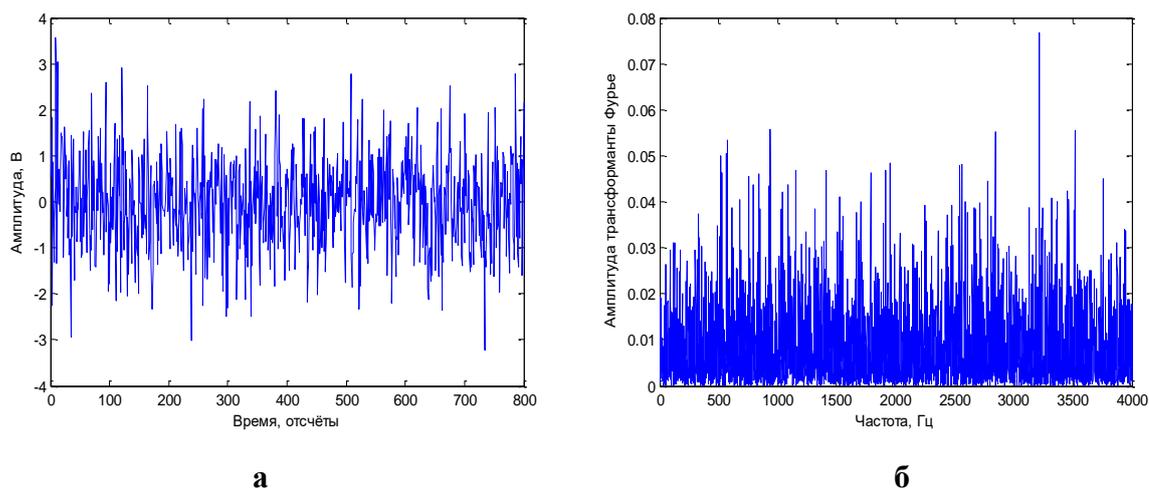
**Рисунок 1.2 – Сигнал, соответствующий безударному звуку «А», во временной (а) и частотной (б) областях ( $f_d = 8$  кГц)**



**Рисунок 1.3 – Сигнал, соответствующий звуку «В», во временной (а) и частотной (б) областях ( $f_d = 8$  кГц)**



**Рисунок 1.4 – Сигнал, соответствующий звуку «Н», во временной (а) и частотной (б) областях ( $f_d=8$  кГц)**



**Рисунок 1.5 – Сигнал, соответствующий паузе, во временной (а) и частотной (б) областях ( $f_d=8$  кГц)**

С связи с тем, что звуки речи меняют друг друга, то основным свойством речевых сигналов следует принять нестационарность, т.е. характеристики речевых сигналов меняются во времени. Так одной из основных характеристик, меняющейся в зависимости от анализируемого отрезка сигнала, является энергия сигнала, которая может быть выражена через равенство Парсеваля:

$$\|\vec{x}\|_t^2 = \int_t^{t+T} x^2(\tau) d\tau = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X_t(\omega, T)|^2 d\omega, \quad (1.3)$$

$$\|\vec{x}\|_t^2 = \sum_{r=0}^{\infty} P_r(\vec{x}), \quad (1.4)$$

где  $\|\vec{x}\|_t^2$  - энергия отрезка сигнала;  $P_r(\vec{x})$  - энергия сигнала в  $\Omega_r$  - том частотном интервале.

Исследования тонкой структуры энергетического спектра речевого сигнала в частотной области позволили установить, что энергия звуков речи распределена неравномерно и сосредоточена в достаточно узких частотных интервалах (см. рисунки 1.2 Б-1.4 Б), в то время как энергия отрезка сигнала, принадлежащего паузе, распределена равномерно во всём анализируемом частотном диапазоне (см. рисунок 1.5 Б).

Для анализа особенностей речевых сигналов можно использовать метод вычисления точных значений частей энергии, сосредоточенных в заданных частотных интервалах [14,15,16,17,18,19].

Полный набор частей энергии отрезка сигнала можно определить как:

$$P_r = \vec{x}^T A_r \vec{x}, \quad (1.5)$$

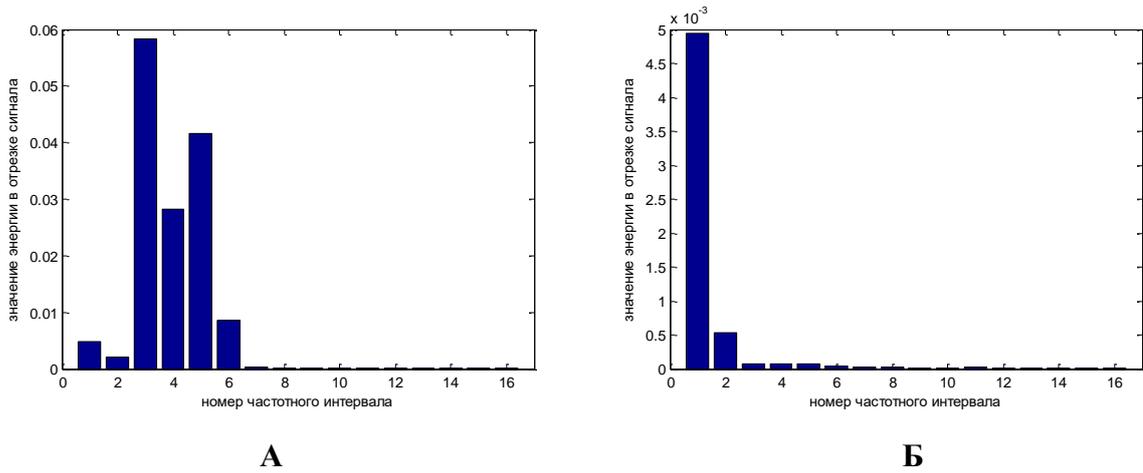
где  $\vec{x}$  - отрезок анализируемого сигнала;  $r$  - номер анализируемого частотного интервала,  $r = \overline{1, R}$ , где  $R$  - количество частотных интервалов, на которые разбивается ось частот;  $A_r$  - субполосная матрица, рассчитанная для  $r$ -того частотного интервала [14,15]:

$$A_r = \{a_{ik}\} = \begin{cases} \frac{\sin[\nu_2(i-k)] - \sin[\nu_1(i-k)]}{\pi(i-k)}, & i \neq k \\ \frac{\nu_2 - \nu_1}{\pi}, & i = k \end{cases}, \quad (1.6)$$

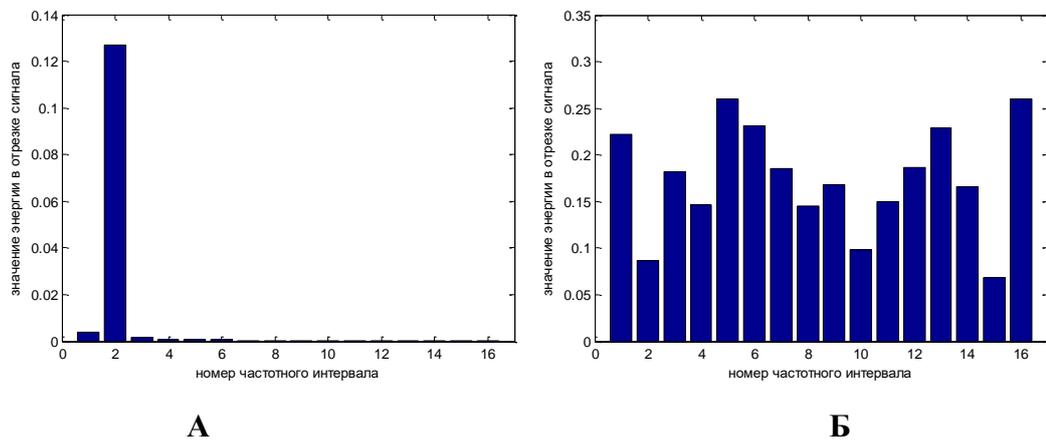
где  $\nu_2, \nu_1$  - границы  $r$ -того частотного интервала, учитывая

$$\begin{aligned} 0 \leq \nu_r < \nu_{r+1} \leq \pi \\ \nu_{r+1} - \nu_r = \pi/R \end{aligned} \quad (1.7)$$

Тогда с использованием ф.1.7 рисунки 1.2Б – 1.5Б будут представлены в следующем виде:



**Рисунок 1.6 – Распределение энергии для сигналов, соответствующих звукам «А» (А) и «В» (Б)**



**Рисунок 1.7 – Распределение энергии для сигналов, соответствующим звуку «Н» (А) и паузе (Б)**

## ГЛАВА 2 ОБЗОР МЕТОДОВ КОДИРОВАНИЯ РЕЧЕВОГО СИГНАЛА

Широкое применение в информационно – телекоммуникационных системах нашли алгоритмы Voice Activity Detector (VAD) [4, 6, 21, 41]. Реализация алгоритмов VAD основана на различиях речевого сигнала и шума. При этом основное внимание уделяется следующим особенностям:

- речь является нестационарным процессом [12, 37];
- фоновый шум стационарен на более длинном отрезке времени по сравнению с речью [7, 29];
- уровень речевого сигнала обычно выше уровня фонового шума.

### 2.1 VAD на основе сравнения с энергетическим порогом

Одной из простейших реализаций VAD является принятие решения о наличии или отсутствии полезного сигнала на основе сравнения уровня энергии фрагмента сигнала с энергетическим порогом [4,18,26].

Алгоритм данного метода выглядит следующим образом:

1. Создать и (или) загрузить звуковой файл формата \*.wav, содержащий некоторое количество периодов молчания (пауз).

2. Задать начальные параметры:

1) длину анализируемого отрезка  $N$  (значение подобрать эмпирически, исходя из предварительного анализа длительности звуков и пауз в речевых сигналах);

2) начальное значение энергетического порога  $H$ , которое должно быть максимальным (т.е. настолько большим, чтобы первый отрезок идентифицировался как пауза, поскольку мы предполагаем, что любой звуковой сигнал начинается с периода молчания).

3. Создать цикл анализа отрезков сигнала.

4. В теле цикла для отрезка анализа:

4.1. Вычислить величину энергии сигнала:

$$P = \sum_{i=1}^N x(i)^2, \quad (2.1)$$

где  $\vec{x} = (x_1, \dots, x_i, \dots, x_N)^T$  - анализируемый отрезок сигнала,  $i = \overline{1, N}$ , где  $N$  - длина окна анализа;  $P$  - значение энергии сигнала  $\vec{x}$ .

4.2. Сравнить величину энергии с пороговым значением, т.е. проверять условие  $P < H$ ;

4.3. Если условие выполняется, то отрезок сигнала идентифицируется как пауза и следует:

а) вычислить усредненное значение энергии  $P_{cp}$  (с учетом значений предыдущих отрезков анализа);

б) вычислить значение порога, уточненное по среднему значению энергии:

$$H = K \cdot P_{cp}, \quad (2.2)$$

где  $K$  – коэффициент, определенный эмпирически;

в) кодировать паузу (для кодирования пауз можно использовать различные способы, например, запомнить номер отрезка анализа, или запоминать номер начального отсчета и длину паузы);

4.4. Если условие не выполняется, то перейти к анализу следующего отрезка.

Такая реализация целесообразна лишь в том случае, когда уровень фонового шума низкий по сравнению с уровнем сигнала, порождаемого звуками речи. К тому же к недостаткам метода можно отнести то, что он не

адаптивный, т.к. коэффициент  $K$  подбирается эмпирически. Поэтому при программной реализации было решено вычислять значение энергетического порога как:

$$H = \sum_{i=1}^{N \cdot M} X(i)^2 / M, \quad (2.3)$$

где  $\vec{X} = (X_1, \dots, X_i, \dots, X_{N \cdot M})^T$  - анализируемая запись,  $i = \overline{1, N}$ , где  $N$  - длина окна анализа,  $M$  - количество отрезков анализа.

Программная реализация предложенного метода в среде Matlab представлена в ПРИЛОЖЕНИИ Б.

## 2.2 VAD на основе коэффициентов линейного предсказания

В алгоритмах VAD широко используются коэффициенты автокорреляции для определения энергетического уровня сигнала и его стационарности [8,23]. Решение о наличии речевого сигнала принимается в том случае, если энергия сигнала превышает пороговое значение и сигнал является нестационарным [17,18].

$$Df_n = A(0)r(0) + 2 \sum_{i=1}^p A(i) \frac{r(i)}{r(0)}, \quad (2.4)$$

$$r(i) = \sum_{k=0}^{N-i} x(k)x(k+i), \quad i = \overline{0, p}, \quad (2.5)$$

$$A(i) = \sum_{k=0}^{p-i} a(k)a(k+i), \quad i = \overline{0, p}, \quad (2.6)$$

где  $Df_n$  - среднее значение автокорреляции сигнала, вычисленное для  $n$ -го отрезка;  $r_n(i)$  - коэффициенты автокорреляции  $n$ -го отрезка входного сигнала;  $A_n(i)$  - коэффициенты автокорреляции средних значений коэффициентов линейного предсказания  $n$ -го отрезка;  $p$  - порядок модели;

$\vec{x} = (x_1, \dots, x_j, \dots, x_N)^T$  - анализируемый сигнал,  $j = \overline{1, N}$ ;  $N$  - длина окна анализа;  $a$  - средние значения коэффициентов линейного предсказания, рассчитываемые на основе средних коэффициентов автокорреляции с использованием алгоритма Левинсона - Дарбина [38] (см. ПРИЛОЖЕНИЕ А).

Если абсолютное значение разности между значениями  $Df$  текущего и последующего отрезков больше, чем установленное пороговое значение, то текущий отрезок считается нестационарным, в противном случае – стационарным.

В дополнение к среднему значению автокорреляции сигнала  $Df$ , можно вычислять значение остаточной энергии:

$$E = A(0)r(0) + 2 \sum_{i=1}^p A(i)r(i), \quad (2.7)$$

где  $E$  - остаточная энергия;  $r(i)$  - коэффициенты автокорреляции входного сигнала;  $A(i)$  - коэффициенты автокорреляции средних значений коэффициентов линейного предсказания;  $p$  - порядок модели.

Для определения пороговых значений величин  $Df$  и  $E$ , сравнивается разница этих значений для текущего и предыдущего отрезков, которые составляют совокупность обучающей выборки (т.е. заведомо отнесены к шуму). Для этого обучающая выборка разбивалась во временной области на отрезки фиксированной длины  $N$  (128 и 256 отсчётов при  $f_s = 8000 \text{ Гц}$ ) со сдвигов в 5 отсчётов. Для каждого фрагмента вычислялись значения остаточной энергии  $E$  (2.7) и среднее значение коэффициентов автокорреляции  $Df$  (2.4). В качестве пороговых значений выступают максимальные значения остаточной энергии  $E$ , вычисленные для отрезка из

обучающей выборки, и максимальное значение абсолютных величин разностей между  $Df$  соседних отрезков, полученное на этапе обучения.

Программные реализации предложенного метода в среде Matlab представлены в ПРИЛОЖЕНИИ В и ПРИЛОЖЕНИИ Г.

### 2.3 Метод обнаружения пауз на основе субполосного анализа

Распределение долей энергии по частотным интервалам может быть использовано для обнаружения пауз в речевом сигнале. Решающая функция для проверки гипотезы о том, что анализируемый отрезок сигнала соответствует паузе между звуками данных имеет вид [17,18]:

$$S = \max(P_r / P_r^{\Pi}) \text{ для любых } r = \overline{1, R}, \quad (2.8)$$

где  $P_r$  - значение части энергии  $r$ -ого частотного интервала для отрезка сигнала, соответствующего активной речи;  $P_r^{\Pi}$  - значение части энергии  $r$ -ого частотного интервала для отрезка сигнала, соответствующего паузе;  $S$  - решающая функция.

При этом значение  $P_r^{\Pi}$  получается на этапе обучения за счёт усреднения по большому количеству отрезков сигнала, заведомо относящимся к паузам:

$$P_r^{\Pi} = \frac{\sum_{k=1}^{N_y} (P_r)_k^{\Pi}}{N_y}, \quad (2.9)$$

где  $N_y$  - количество отрезков сигнала в паузе, которые используются для усреднения.

Если выполняется неравенство (2.10), то принимается, что анализируемый отрезок относится к паузе.

$$S > h_\alpha, \quad (2.10)$$

где  $h_\alpha$  - пороговое значение, обеспечивающее заданный уровень вероятности ложной тревоги, т.е.  $P\{S > h_\alpha\} \leq \alpha \ll 1$ .

Пороговое значение  $h_\alpha$  можно получить из основанного на неравенстве Чебышева выражении (2.11).

$$h_\alpha \leq \bar{S}_\Pi + D_\Pi / d \cdot \sqrt{\alpha}, \quad (2.11)$$

где  $\bar{S}_\Pi$  - пороговое значение, обеспечивающее заданный уровень вероятности ложной тревоги, т.е.  $P\{S > h_\alpha\} \leq \alpha \ll 1$ .

### Словесное описание алгоритма

1. Ввести количество  $R$  частотных интервалов и значение длительности обрабатываемого отрезка сигнала  $N$ ;

2. Для  $r = 1, \dots, R$ ;  $i, k = 1, \dots, N$  осуществить вычисления элементов матриц (см. ф. 1.6);

3. Ввести заведомо относящиеся к паузе данные длительностью  $N_y N$ ;

4. Для  $k = 1, \dots, N_y$  выполнить:

4.1. Ввести вектор значений обрабатываемого отрезка сигнала

$$\bar{x}_k = (x_{k1}, \dots, x_{kN})^T;$$

4.2. Вычислить вектор  $\vec{y}_k = (\vec{y}_{1r}^T, \dots, \vec{y}_{1R}^T)^T = A A \bar{x}_k$

4.3. Для  $r = 1, \dots, R$  вычислить  $(P_r)_k^\Pi = \sum_{j=1}^J (y_{jr}^2)_1$ ;

5. Для  $r = 1, \dots, R$  вычислить  $P_r^\Pi = \sum_{k=1}^{N_y} (P_r)_k^\Pi / N_y$

6. Для  $k = 1, \dots, N_y$  вычислить  $S_k^{\Pi} = \max((P_r)_k^{\Pi} / P_r^{\Pi}), \forall r = 1, \dots, R$ ;
7. Вычислить  $\bar{S}_{\Pi} = \sum_{k=1}^{N_y} (S_k^{\Pi}) / N_y$  и  $D_{\Pi}^2 = \sum_{k=1}^{N_y} (S_k^{\Pi})^2 / N_y - \bar{S}_{\Pi}^2$ ;
8. Ввести значение вероятности ложной тревоги  $\alpha \ll 1$ ;
9. Выполнить итерации по определению порога:
  - 9.1. Ввести  $m=0; a_0 = 2$ ;
  - 9.2. Вычислить  $h_{\alpha}^m \leq \bar{S}_{\Pi} + D_{\Pi} / a_m \sqrt{\alpha}$ ;
  - 9.3. Вычислить  $\beta_m = \sum_{k=1}^{N_y} \text{sig}(S_k^{\Pi} - h_{\alpha}^m), \text{sig}(x) = 1, x > 0; \text{sig}(x) = 0, x \leq 0$ ;
  - 9.4. Если  $|\alpha - \beta_m| \leq \alpha^2$  то положить  $h_{\alpha} = h_{\alpha}^m$  и прекратить итерации;
  - 9.5. В противном случае при  $\alpha > \beta_m$  вычислить  $a_{m+1} := (1 - \alpha a_m) * a_m$ , если же выполняется неравенство  $\alpha < \beta_m$ , то вычислить  $a_{m+1} := (1 + \alpha a_m) * a_m$ , положить  $m = m + 1$  и перейти к выполнению 9.2;
10. Задать длительность  $M \geq I * N, I = 1, 2, \dots$ , предназначенных для сжатия за счёт обнаружения и кодирования пауз данных;
11. Ввести предназначенные для обработки данные  $\vec{z} = (z_1, \dots, z_M)^T$ ;
12. Задать шаг смещения вдоль отрезка обрабатываемых данных  $K$ , такой, что отношение  $U = N / K$  является целым числом;
13. Выразить  $MM = (I - 1) * U$ ;
14. Для  $k = 1, \dots, MM$  выполнить:
  - 14.1. Положить  $\vec{x}_k = (x_{1k}, \dots, x_{Nk})^T, x_{mk} = z_{(k-1)K+m}$ ;
  - 14.2. Вычислить вектор  $y_{\vec{y}}_k = (\bar{y}_{1r}^T, \dots, \bar{y}_{1R}^T)^T = AA\vec{x}_k$ ;
  - 14.3. Для  $r = 1, \dots, R$  вычислить  $(P_r)_k = \sum_{j=1}^{J_r} (y_{jr}^2)_1$ ;
  - 14.4. Вычислить  $S_k = \max((P_r)_k / P_r^{\Pi}), \forall r = 1, \dots, R$ ;
  - 14.5. Если  $S_k > h_{\alpha}$ , то принимается решение о наличии в анализируемом отрезке данных, порождаемых звуками речи, полагается

$k = k + 1$  и при выполнении условия  $k < MM$  выполняются предыдущие операции, начиная с подпункта 14.1, а при выполнении неравенства  $k \geq MM$  вычисления заканчиваются. В противном случае (неравенство  $S_k > h_\alpha$  не выполняется) принимается решение о принадлежности анализируемого отрезка данных паузе и выполняются следующие действия:

14.6. Если на предыдущем этапе (предыдущем  $k$ ) не принималось решения о наличии паузы, то началу новой паузы присваивается код,  $C_n = (k_n - 1)K$ , а длительности присваивается значение  $E_n = K$ , где  $k_n$  - номер итерации при которой обнаружено начало  $n$ -той паузы в обрабатываемой совокупности данных. Если же и на предыдущем этапе отрезок анализируемых данных был отнесен к паузе номером  $n$ , то код начала паузы не изменяется, а длительность увеличивается на величину шага смещения, то есть вычисляется  $E_n = E_n + K$ ;

14.7. Полагается  $k = k + 1$  и при выполнении условия  $k < MM$  выполняются предыдущие операции, начиная с подпункта 17.1, а при выполнении неравенства  $k \geq MM$  вычисления заканчиваются.

15. Конец.

Программная реализация предложенного метода в среде Matlab представлена в ПРИЛОЖЕНИИ Д.

## ГЛАВА 3 ИССЛЕДОВАНИЕ АЛГОРИТМОВ ВЫЯВЛЕНИЯ АКТИВНОСТИ РЕЧИ

### 3.1 Сравнительная оценка работоспособности методов обнаружения пауз

Для проведения эксперимента был записан речевой материал, соответствующий фразе «Он пошёл каменистой тропинкою вверх от залива». Речевой материал состоит не только из активной речи, но и из периода паузы, в самом начале записи (для обучающей выборки). Фраза была записана при частоте дискретизации 8000 Гц и 16 уровней квантования.

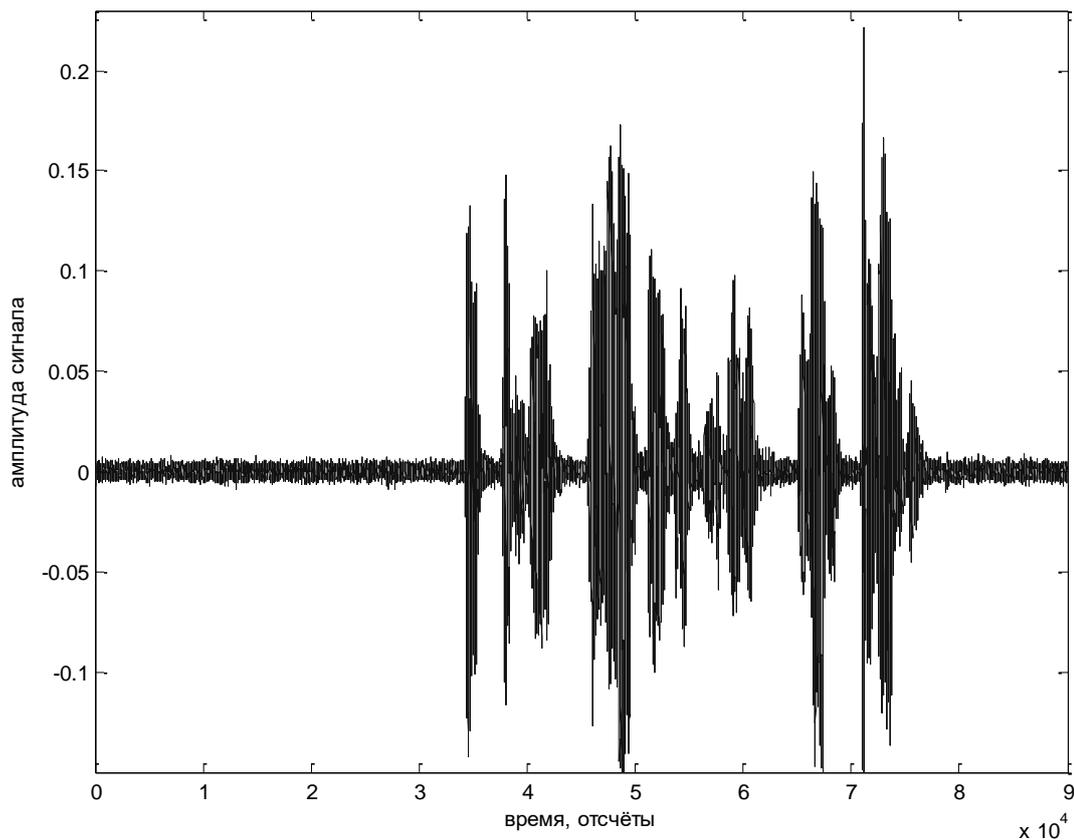
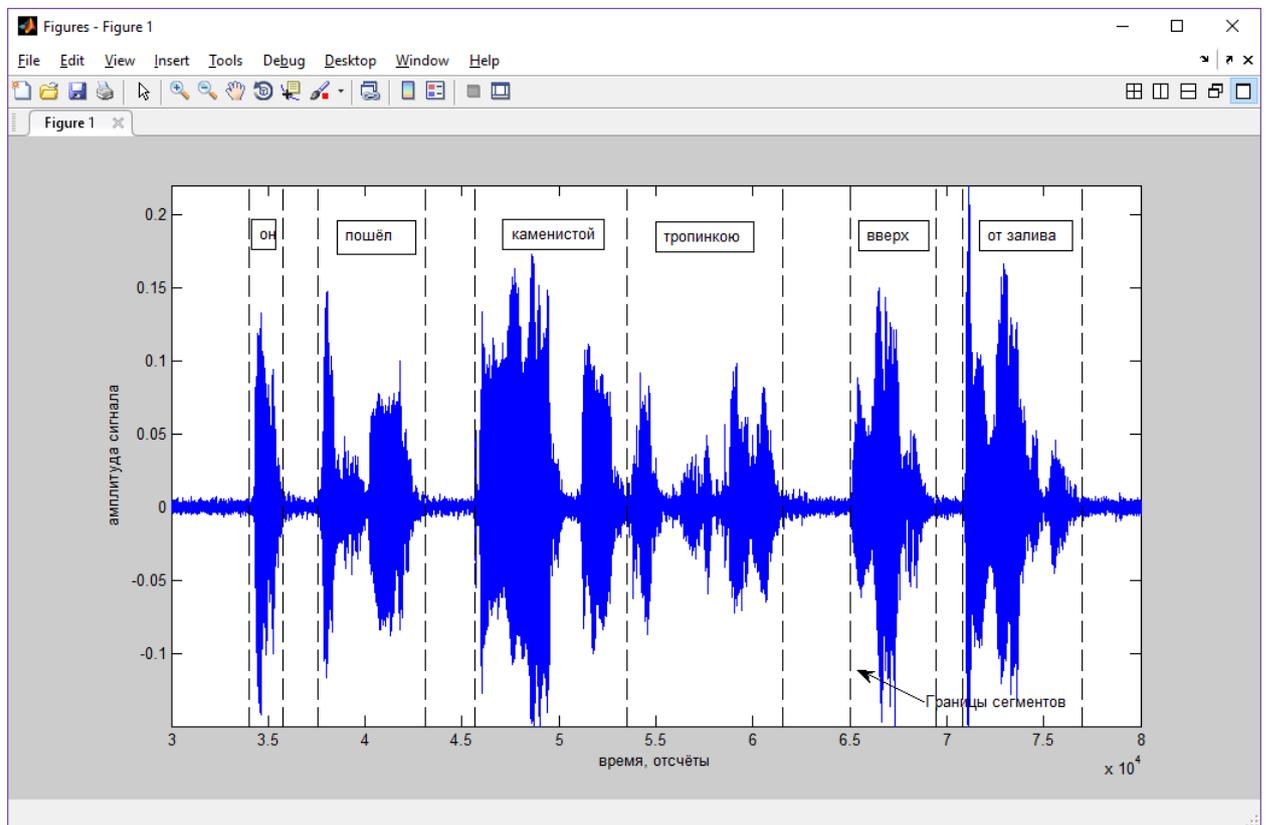


Рисунок 3.1 – Обрабатываемый сигнал во временной области

Для оценки адекватности работы алгоритмов, кодирование пауз было проведено вручную, т.е. границы сегментов были определены по осциллограмме и зафиксированы их значения.

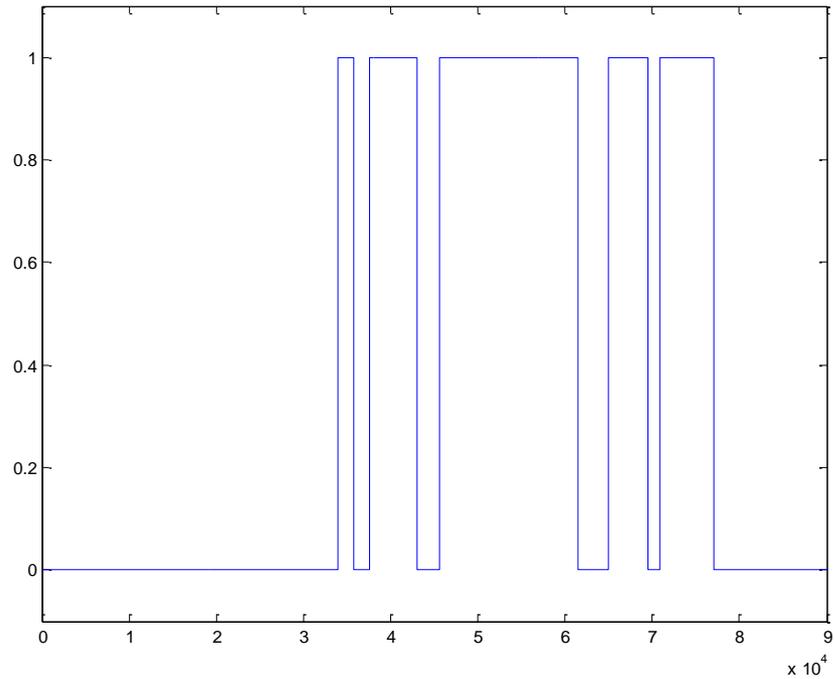
**Таблица 3.1 - Значения отсчётов, соответствующие началу и окончанию сегментов**

	он	пошёл	каменистой	тропинкою	вверх	от залива
№ отсчёта начала сегмента	34000	37569	45661	53478	65002	70831
№ отсчёта конца сегмента	35784	43107	53477	61509	69466	77000



**Рисунок 3.2 – Обрабатываемый сигнал во временной области, разбитый на сегменты**

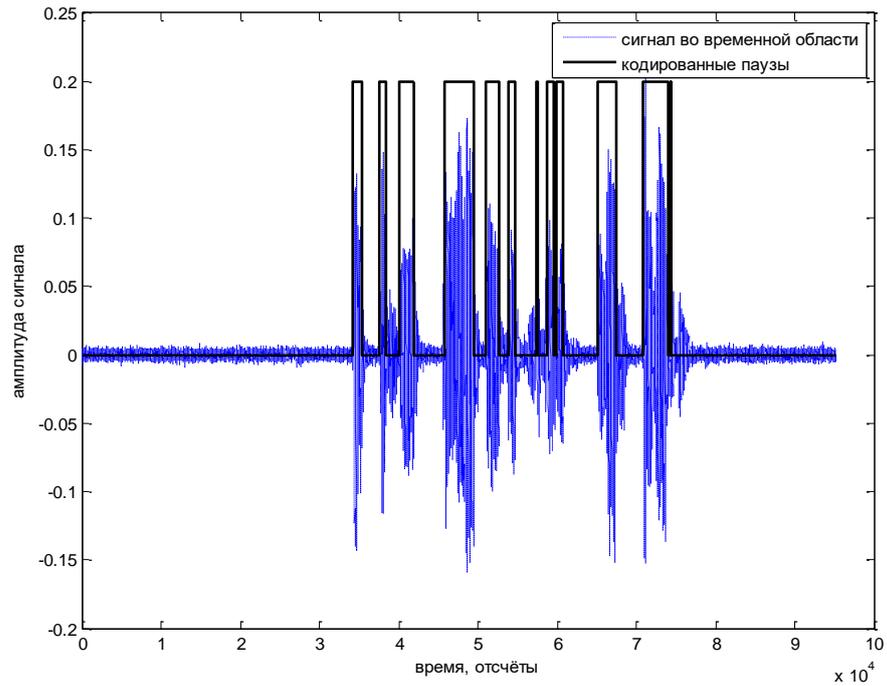
Если представить входную последовательность в виде бит, где 0 означает, что данный отсчёт представляет собой шум, а 1 – активный сигнал, то сигнал примет следующий вид:



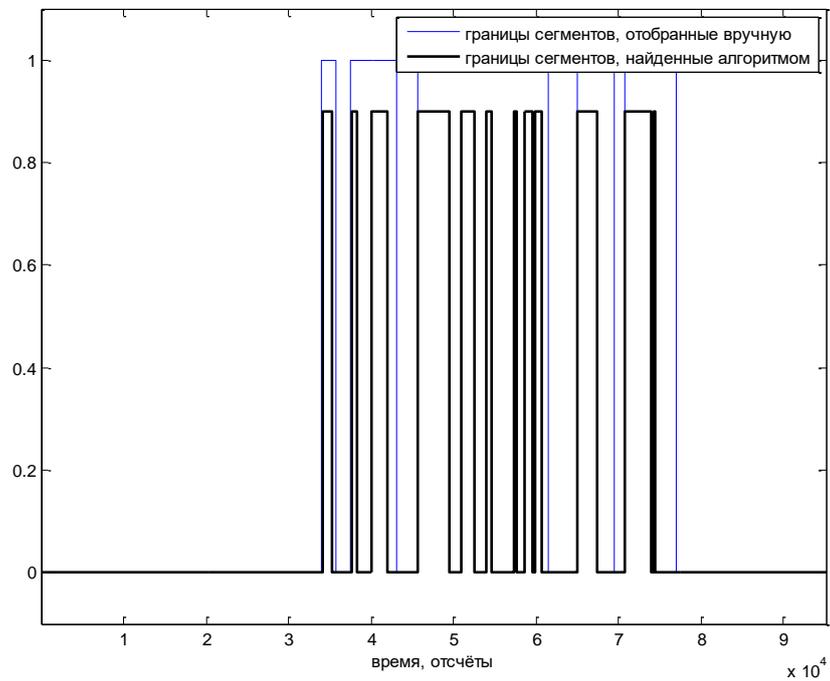
**Рисунок 3.3 – Обрабатываемый сигнал, представленный в виде двоичной последовательности**

Благодаря такому представлению легко визуально видеть разницу в границах сегментов, найденный вручную и с помощью алгоритма кодирования пауз.

Для того, чтобы продемонстрировать работу алгоритмов в одинаковых условиях, была выбрана единая для всех длина отрезка анализа  $N$ , равная 256 отсчётам. Для обучения используются первые 30000 отсчётов сигнала.



**Рисунок 3.4 А – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное во временной области**

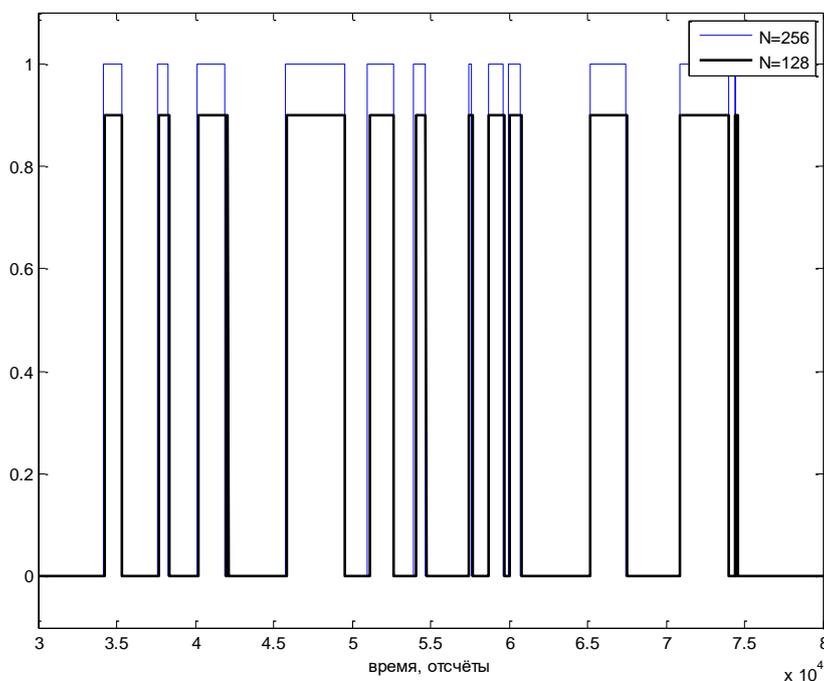


**Рисунок 3.4 Б – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное в виде двоичной последовательности**

Из рисунков 3.4 А и 3.4 Б можно отметить, что довольно-таки много отрезков, когда активная речь была воспринята как пауза. Поэтому подобный метод лучше использовать только при малых внутренних шумах. В случаях, когда в речи используются малоэнергетичные (или шумоподобные) звуки (например "ч"), детектор может принять их за паузу.

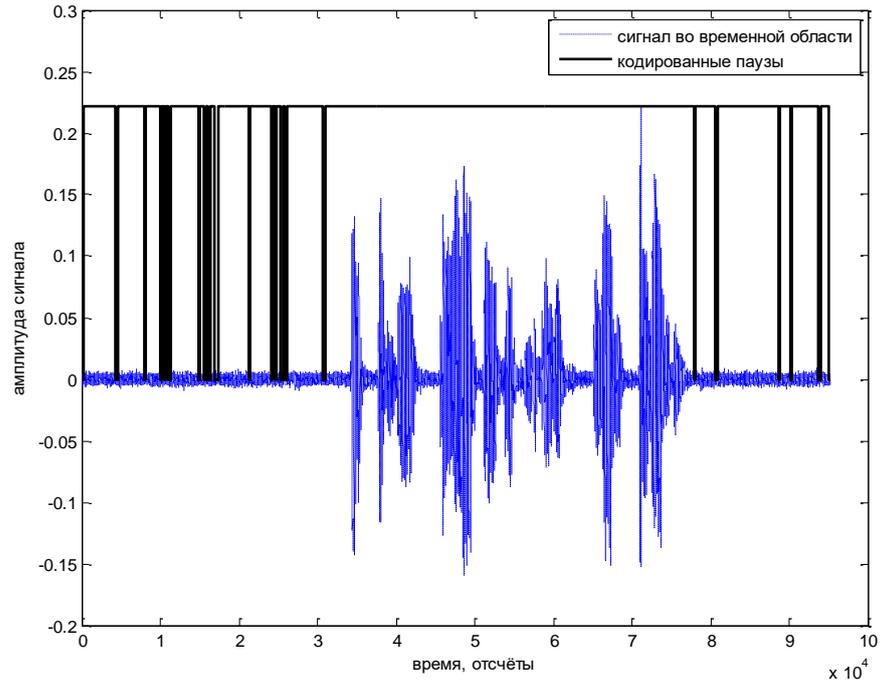
В ПРИЛОЖЕНИИ Е представлены результаты работы алгоритма при изменении длины отрезка анализа  $N$ , а также изменении формулы вычисления значения энергетического порога  $H$  с ф. (2.2) на ф. (2.3).

Можно отметить, что при уменьшении длины отрезка анализа, улучшения в определении границ активной речи и пауз не наблюдается:

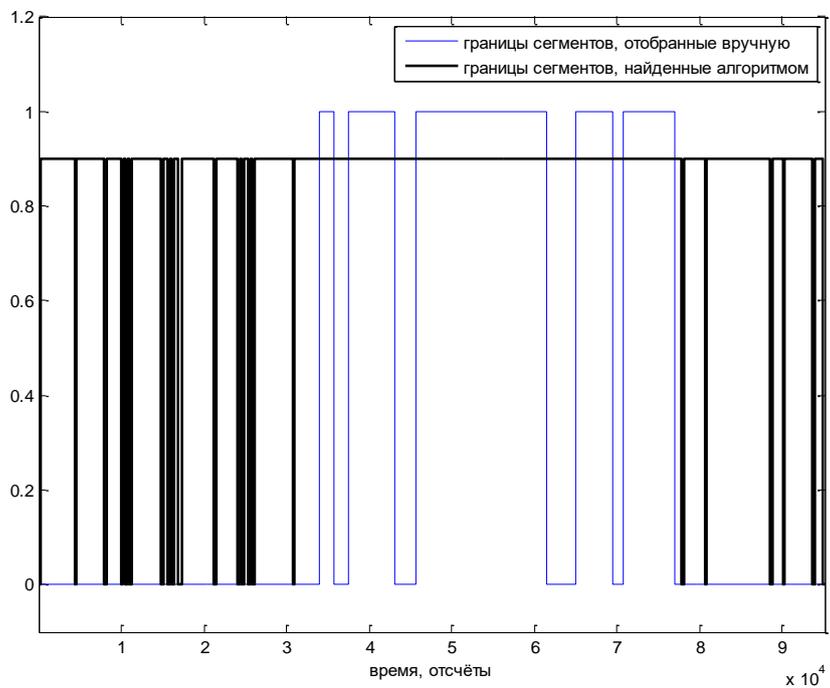


**Рисунок 3.5 – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное в виде двоичной последовательности при разной длине отрезка анализа  $N$**

Проведём аналогичные исследования для метода VAD на основе коэффициентов линейного предсказания:



**Рисунок 3.6 – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное во временной области ( $N=256, p=8$ )**

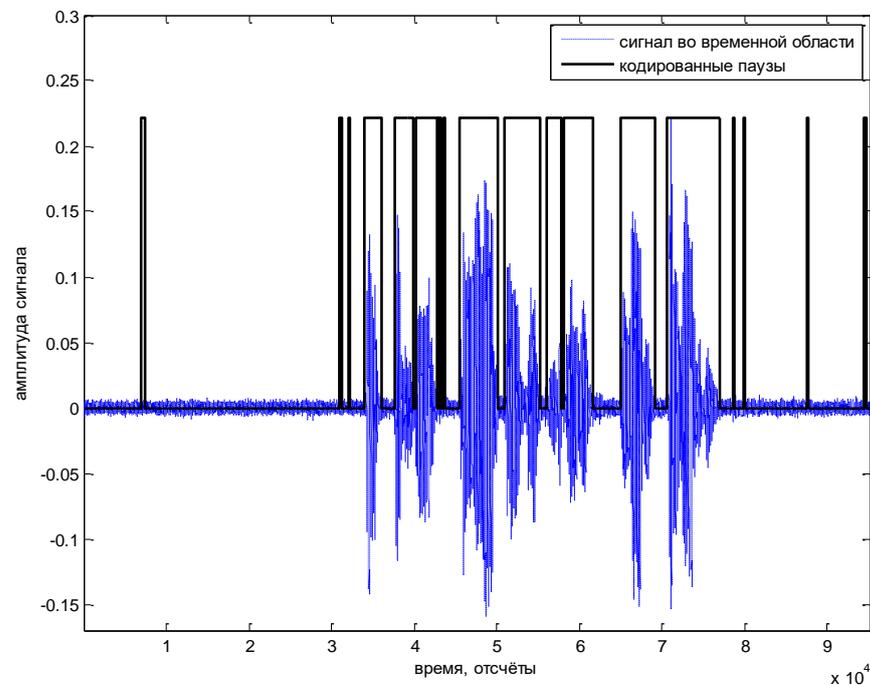


**Рисунок 3.7 – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное в виде двоичной последовательности ( $N=256, p=8$ )**

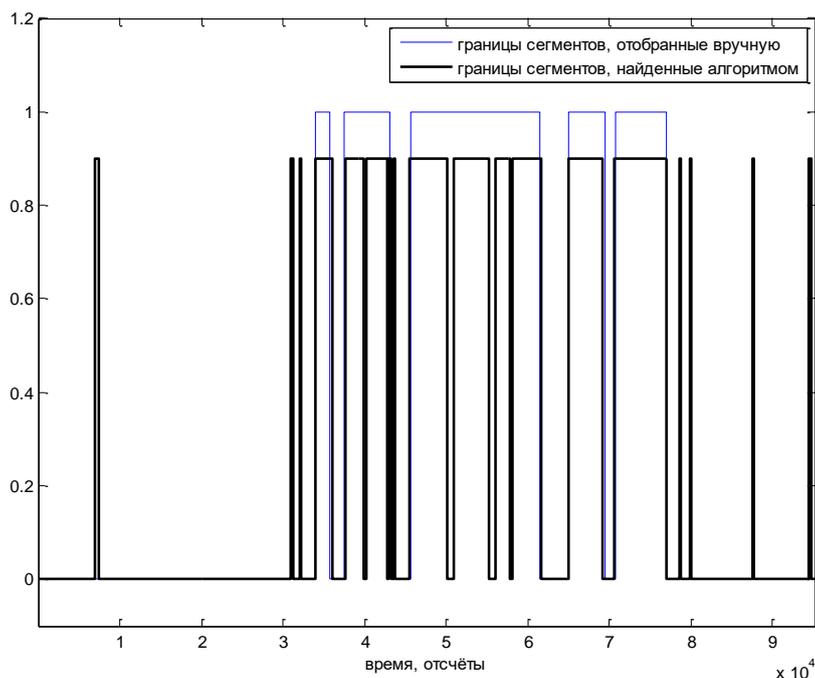
Из рисунков 3.6-3.7 можно отметить, что не было обнаружено ни одной паузы между словами в произнесённой фразе, а также, что даже те отрезки, на которых происходило обучение, восприняты как активная речь.

В ПРИЛОЖЕНИИ Ж представлены результаты работы алгоритма при изменении длины отрезка анализа  $N$  и порядка модели  $p$ .

В рамках эксперимента, было исследовано, как искусственное увеличение пороговых значений влияет на кодирование пауз (см. рисунок 3.7).



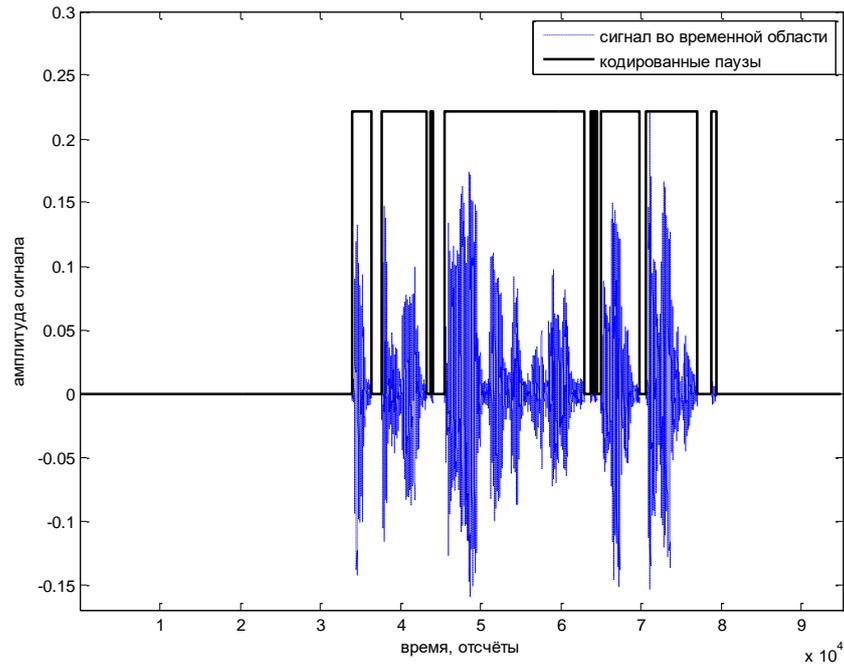
**Рисунок 3.8 – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное во временной области ( $N=256$ ,  $p=8$ ) при увеличении пороговых значений в 10 раз**



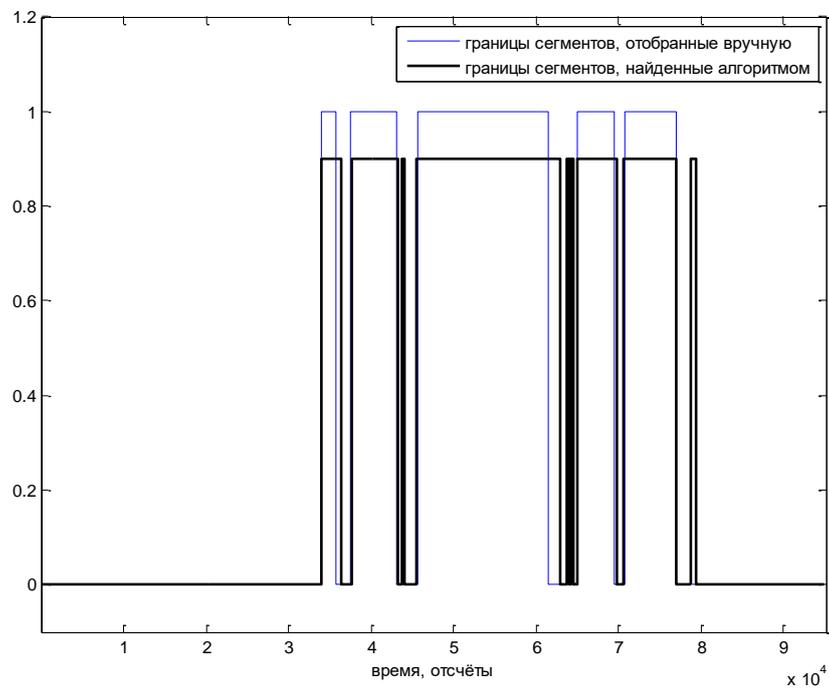
**Рисунок 3.9 – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное в виде двоичной последовательности ( $N=256, p=8$ ) при увеличении пороговых значений в 10 раз**

Можно отметить, что с увеличением пороговых значений  $Df$  и  $E$ , способность алгоритма VAD на основе коэффициентов линейного предсказания обнаружить паузы значительно улучшилась. Поэтому можно утверждать, что подобный алгоритм требует улучшений при обучении.

Проведём аналогичные исследования для метода обнаружения пауз на основе субполосного анализа:



**Рисунок 3.10 – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=256, R=16$ )**



**Рисунок 3.11 – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=256, R=16$ ), представленное в виде двоичной последовательности**

В ПРИЛОЖЕНИИ 3 представлены результаты работы алгоритма при изменении длины отрезка анализа  $N$  и количества интервалов, на которые разбивается ось частот  $R$ .

Таким образом на основании рисунков 3.3 А – 3.8 Б можно сделать вывод, что методы VAD являются работоспособными только при эмпирической подстройке параметров во время обучения, т.е. не являются адаптивными, в отличие от метода обнаружения пауз на основе субполосного анализа.

### 3.2 Оценка выявления исключения пауз на ёмкость речевого материала

Оценить, насколько исключение пауз влияет на общий объём речевого материала, введём величину  $K_{сж}$  - коэффициент сжатия, который рассчитывается в соответствии с ф.3.1.

$$K_{сж} = \frac{N_{исх}}{N_{акт.р}}, \quad (3.1)$$

где  $N_{исх}$  - количество отсчётов исходного сигнала до обработки;  $N_{акт.р}$  - количество отсчётов сигнала, содержащего только активную речь (т.е. без пауз).

Для того, чтобы адекватно оценить результаты эксперимента, вычислим значение  $K_{сж}$  для идеального случая, т.е. когда паузы были определены вручную:

$$K_{сж} = \frac{95192 \text{ отсчётов}}{33808 \text{ отсчётов}} = 2,8157 \text{ раз}$$

Таким образом, если полученные в результате экспериментов значения  $K_{сж}$  будут больше значения 2,8, то активная речь была воспринята как пауза и информация в сообщении утеряна, а если меньше 2,8 - значит сжатия за счёт удаления пауз происходит не эффективно.

Эксперименты проводились при разных параметрах:

- для метода VAD на основе сравнения с энергетическим порогом - менялись длина обрабатываемого отрезка  $N$ .

- для метода VAD на основе коэффициентов линейного предсказания - менялись длина обрабатываемого отрезка  $N$ , и порядок модели  $p$ .

- для метода обнаружения пауз на основе субполосного анализа - менялись длина обрабатываемого отрезка  $N$ , и количество интервалов, на которые разбивается ось частот  $R$ .

Все полученные результаты сведены в таблицы 3.2 – 3.4.

**Таблица 3.2 – Значение коэффициента сжатия  $K_{сж}$  для метода VAD на основе сравнения с энергетическим порогом**

	$N = 128$ отсчётов	$N = 256$ отсчётов
$K_{сж}$ , раз	5,5851	5,4467

**Таблица 3.3 – Значение коэффициента сжатия  $K_{сж}$  для метода VAD на основе коэффициентов линейного предсказания**

$K_{сж}$ , раз	$p = 8$	$p = 10$	$p = 15$	$p = 20$
$N = 128$ отсчётов	1,0564	1,0445	1,0549	1,0685
$N = 256$ отсчётов	1,0655	1,0599	1,1067	1,0685

**Таблица 3.4 – Значение коэффициента сжатия  $K_{сж}$  для метода обнаружения пауз на основе субполосного анализа**

	$N = 128$ отсчётов			$N = 256$ отсчётов		
	$R = 16$	$R = 32$	$R = 64$	$R = 16$	$R = 32$	$R = 64$
$K_{сж}$ , раз	2,5372	2,5635	2,6178	2,5119	2,3983	2,4951

Таким образом, на основании таблиц 3.2-3.4 можно сделать вывод о том, что метод обнаружения пауз на основе субполосного анализа является оптимальным с точки зрения сжатия речевого материала при исключении пауз. Это обосновывается тем, что значения  $K_{сж}$  при подобном методе наиболее близки к значению коэффициента сжатия, рассчитанного вручную.

### **3.3 Оценка выявления точности определения границы пауза/речь на вероятность ошибки декодирования**

Оценка эффективности работы алгоритмов осуществлялась на основе определения вероятностей ошибок I и II рода. При этом за основную принималась гипотеза о наличии паузы.

Вероятность ошибки I рода  $P_{I \text{ рода}}$  («ложная тревога») рассчитывалась по формуле:

$$P_{I \text{ рода}} = \frac{N_{\text{пр-пауз}}}{N_{\text{сигн}}}, \quad (3.2)$$

где  $N_{\text{пр-пауз}}$  - количество отсчётов, принятых за паузу в заведомо известном полезном сигнале;  $N_{\text{сигн}}$  - количество отсчётов в заведомо известном полезном сигнале.

Вероятность ошибки II рода  $P_{II\text{рода}}$  («пропуск цели») рассчитывалась по формуле:

$$P_{II\text{рода}} = \frac{N_{\text{пр-сигн}}}{N_{\text{пауз}}}, \quad (3.3)$$

где  $N_{\text{пр-сигн}}$  - количество отсчётов, принятых за полезный сигнал в заведомо известном фрагменте пауз;  $N_{\text{пауз}}$  - количество отсчётов в заведомо известном фрагменте пауз.

Эксперименты проводились при разных параметрах:

- для метода VAD на основе сравнения с энергетическим порогом - менялись длина обрабатываемого отрезка  $N$ .

- для метода VAD на основе коэффициентов линейного предсказания - менялись длина обрабатываемого отрезка  $N$ , и порядок модели  $p$ .

- для метода обнаружения пауз на основе субполосного анализа - менялись длина обрабатываемого отрезка  $N$ , и количество интервалов, на которые разбивается ось частот  $R$ .

Все полученные результаты сведены в таблицы 3.5 – 3.7.

**Таблица 3.5 – Значение вероятностей ошибок  $P_{I\text{рода}}$  и  $P_{II\text{рода}}$  для метода VAD на основе сравнения с энергетическим порогом**

$N = 128$ отсчётов		$N = 256$ отсчётов	
$P_{I\text{рода}}$	$P_{II\text{рода}}$	$P_{I\text{рода}}$	$P_{II\text{рода}}$
0,4959	0	0,4849	0,0010

**Таблица 3.6 – Значение вероятностей ошибок  $P_{I\text{рода}}$  и  $P_{II\text{рода}}$  для метода VAD на основе коэффициентов линейного предсказания**

	$N = 128$ отсчётов		$N = 256$ отсчётов	
	$P_{I\text{рода}}$	$P_{II\text{рода}}$	$P_{I\text{рода}}$	$P_{II\text{рода}}$
$p = 8$	0	0,9172	0	0,9047
$p = 10$	0	0,9339	0	0,9131
$p = 15$	0	0,9193	0	0,8505
$p = 20$	0	0,9006	0	0,9006

При этом наиболее адекватные результаты для метода VAD на основе коэффициентов линейного предсказания фиксируются при увеличении пороговых значений в 10 раз ( $N=256, p=8$ ):  $P_{I\text{рода}} = 0,0824$ ,  $P_{II\text{рода}} = 0,0935$ .

**Таблица 3.7 – Значение вероятностей ошибок  $P_{I\text{рода}}$  и  $P_{II\text{рода}}$  для метода обнаружения пауз на основе субполосного анализа**

	$N = 128$ отсчётов			$N = 256$ отсчётов		
	$R = 16$	$R = 32$	$R = 64$	$R = 16$	$R = 32$	$R = 64$
$P_{I\text{рода}}$	0,0115	0,0152	0,0190	0,0039	0,0039	0,0039
$P_{II\text{рода}}$	0,0668	0,0626	0,0521	0,0688	0,0980	0,0729

Основную опасность при обработке сигналов представляют ошибки II рода, поэтому подобные алгоритмы стараются минимизировать вероятность ошибки II рода при фиксированном значении вероятности ошибки I рода.

Несмотря на то, что минимальное значение вероятности ошибки II рода соответствует методу VAD на основе сравнения с энергетическим порогом, его нельзя назвать оптимальным, т.к. значение вероятности ошибки I рода достаточно велико, что свидетельствует о том, что в некоторых случаях активная речь была воспринята как пауза. Поэтому с точки зрения

минимального значения ошибок I и II рода оптимальным является метод обнаружения пауз на основе субполосного анализа.

## ЗАКЛЮЧЕНИЕ

В данной исследовательской работе проведён сопоставительный анализ эффективности нескольких наиболее распространённых методов кодирования речевого сигнала. Данные методы должны быть способны обнаружить границу «активная речь/сигнал», и являются первым шагом в алгоритмах внедрения информации в речь посредством методов стеганографии.

В работе были реализованы алгоритмы VAD на основе сравнения с энергетическим порогом, VAD на основе коэффициентов линейного предсказания, а также метод обнаружения пауз на основе субполосного анализа. Эффективность методов оценивалась с помощью следующих характеристик:

- на основе  $K_{сж}$  - коэффициент сжатия, который характеризует насколько исключение пауз влияет на общий объём речевого материала

- на основе определения вероятностей ошибок декодирования I и II рода, а именно сколько отсчётов сигнала отсчётов активной речи было воспринято как пауза и наоборот.

В результате сопоставительного анализа были сформулированы следующие выводы:

- методы VAD являются работоспособными только при эмпирической подстройке параметров во время обучения, т.е. не являются адаптивными, в отличии от метода обнаружения пауз на основе субполосного анализа.

- метод обнаружения пауз на основе субполосного анализа является оптимальным с точки зрения сжатия речевого материала при исключении пауз. Это обосновывается тем, что значения  $K_{сж}$  при подобном методе наиболее близки к значению коэффициента сжатия, рассчитанного вручную. В методе VAD на основе сравнения с энергетическим порогом

активная речь была воспринята как пауза и информация в сообщении утеряна, а в методе VAD на основе коэффициентов линейного предсказания  $K_{сжс}$  слишком мал, что означает, что сжатие за счёт удаления пауз происходит не эффективно.

- Несмотря на то, что минимальное значение вероятности ошибки II рода соответствует методу VAD на основе сравнения с энергетическим порогом, его нельзя назвать оптимальным, т.к. значение вероятности ошибки I рода достаточно велико, что свидетельствует о том, что в некоторых случаях активная речь была воспринята как пауза. Поэтому с точки зрения минимального значения ошибок I и II рода оптимальным является метод обнаружения пауз на основе субполосного анализа.

Таким образом, выявление участков активности речи наиболее эффективно осуществляется на основе метода обнаружения пауз на основе субполосного анализа, который учитывает основное свойство речевых сигналов, а именно концентрацию энергии сигнала в узкой полосе частот. Поэтому его использование обосновано в задачах поиска активной речи при использовании методов стеганографии.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Аграновский А.В. Стеганография, цифровые водяные знаки и стеганоанализ [Текст] / А.В. Аграновский, А.В. Балакин, В.Г. Грибунин, С.А. Сапожников.- М.: Вузовская книга, 2009.- 220 с.: ил.
2. Аграновский, А.В. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов [Текст] / А.В. Аграновский, Д.А. Леднов – М.: Радио и связь, 2004. – 164 с.
3. Айфичер Э. Цифровая обработка сигналов: практический подход, 2-е издание [Текст]: Пер. с англ. / Айфичер Э., Джервис Б. – М.: Издательский дом «Вильямс», 2004. – 992 с.:ил.
4. Белов С.П. Об уменьшении объема трафика при пакетной передаче речевых сообщений за счет кодирования пауз [Текст] / С.П. Белов, Е.И. Прохоренко // Научные ведомости Белгородского государственного университета. Серия: Информатика и прикладная математика. - Белгород, 2006. - № 1 (21). – Вып. 2. - С. 141–148.
5. Бондарев, В.Н. Цифровая обработка сигналов. Методы и средства [Текст]: Учеб. Пособие для вузов. 2-е изд / Бондарев В.Н., Трестре Г., Чернега В.С. – Х: Конус, 2001. – 398 с.:ил.
6. Волченков В. А. Методы и алгоритмы детектирования активности речи [Текст] / В.А. Волченков, В.В. Витязев // Цифровая Обработка Сигналов. – 2013. - № 1. – С. 54-60.
7. Гантмахер, В.Е. Шумоподобные сигналы. Анализ, синтез, обработка [Текст] / В.Е. Гантмахер, Н.Е. Быстров, Д.В. Чеботарев. – СПб.: Наука и Техника, 2005. – 400 с.: ил.
8. Герасимов А.В. Применение метода модифицированного линейного предсказания к задачам выделения акустических признаков речевых сигналов [Текст] / А.В. Герасимов, О.А. Морозов, В.Р. Фидельман // Радиотехника и Электроника. – 2005. – том 50. №10. – с. 1287 – 1292.

9. Голд, Б. Цифровая обработка сигналов [Текст] / Б. Голд, Ч. Рейдер. – М.: Сов. Радио, 1973. – 368 с.
10. Гольденберг Л.М. Цифровая обработка сигналов. [Текст] / Л.М. Гольденберг. – М.: Радио и связь, 1985. – 312 с.
11. Грибунин Г.В. Цифровая стеганография [Текст] / В. Г. Грибунин, И. Н. Оков, И. В. Туринцев.- М.: СОЛОН-Пресс, 2002.- 272с.: ил.
12. Гудонавичюс, Р.В. Распознавание речевых сигналов по их структурным свойствам [Текст] / Р.В. Гудонавичюс, П.П. Кемешис, А.Б. Читавичюс. – Л., «Энергия», 1977. – 64 с.: ил.
13. Динамические спектры речевых сигналов [Текст] М.Ф. Деркач, Р.Я. Гумецкий, Б.М. Гура, М.Е. Чабан. – Львов: Вища школа. Изд-во при Львов. Ун-те, 1983. - 168 с.
14. Жиялков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. [Текст] / Е.Г. Жиялков. - Белгород: Изд-во БелГУ, 2007. – 160 с.
15. Жиялков Е.Г. Методы анализа и построения функций по эмпирическим данным на основе частотных представлений: монография [Текст] / Е.Г. Жиялков. - Белгород: Изд-во БелГУ, 2007. – 160 с.
16. Жиялков Е.Г. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений [Текст] / Е.Г. Жиялков, С.П.Белов, Е.И. Прохоренко. – Белгород: БелГУ, 2007. – 136 с.
17. Жиялков Е. Г. Об анализе и синтезе речевых сигналов в IP-телефонии [Текст] / Е.Г. Жиялков, А.А. Фирсова // Вестник Национального технического университета Харьковский политехнический институт. Серия: Информатика и моделирование. – 2009. - № 43. – С. 84-91.
18. Жиялков Е.Г. Об эффективности различных подходов, к сегментации речевых сигналов на основе обнаружения пауз [Текст] / Е.Г. Жиялков, С.П. Белов, А.С. Белов и др. ; БелГУ // Научные ведомости БелГУ.

Сер. История. Политология. Экономика. Информатика. - 2010. - №7(78), вып.14/1.-С. 187-193.

19. Жилияков, Е.Г. Сегментация речевых сигналов на основе анализа распределения энергии по частотным интервалам [Текст] / Е.Г. Жилияков, Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова, М.В. Фатова // Научные ведомости Белгородского государственного университета. Серия: История. Политология. Экономика. Информатика, Том 18 – 2011. - №7-1 (102). – С. 187-196.

20. Залманзон Л.А. Преобразование Фурье, Уолша, Хаара и их применение в управлении, связи и других областях. [Текст] / Залманзон Л.А. – М.: Наука. Гл. ред. Физ.-мат.лит., 1989 год. 496 с.

21. Зверев, В.А. Выделение сигналов из помех численными методами [Текст] / В.А. Зверев, А.В. Стромков. – Нижний Новгород: ИПФ РАН, 2001. – 188 с.

22. Конахович Г.Ф. Компьютерная стеганография. Теория и практика [Текст] / Г.Ф. Конахович, А.Ю. Пузыренко.- К.: МК-Пресс, 2006.- 288с.: ил.

23. Коротаев Г.А. Некоторые аспекты линейного предсказания при анализе речевого сигнала [Текст] / Г.А. Коротаев // Зарубежная радиоэлектроника. – 1991. - №7. – с.13-31

24. Лайонс, Р. Цифровая обработка сигналов [Текст] / Лайонс Р; - 2-е изд. ; Пер. с англ. – М.: ООО "Бином-Пресс", 2006 – 656 с.: ил.

25. Марпл-мл, С.Л. Цифровой спектральный анализ и его приложения [Текст] / Марпл-мл. С.Л.; Пер. с англ. –М.: Мир, 1990.

26. Методы цифровой обработки сигналов для решения прикладных задач. Монография [Текст] / Под ред. В.И. Марчука. – М.: Радиотехника, 2012. – 128 с.:ил.

27. Михайлов В.Г. Измерение параметров речи [Текст] / В.Г. Михайлов, Л.В. Златоустова; Под ред. М.А. Сапожкова. – М.: Радио и связь, 1987. – 168 с.: ил.

28. Назаров М.В. Методы цифровой обработки и передачи речевых сигналов [Текст] / М.В. Назаров, Ю.Н. Прохоров. – М.: Радио и связь, 1985. – 176 с.
29. Обнаружение радиосигналов [Текст] / П.С. Акимов, Ф.Ф. Евстратов, С.И. Захаров и др.; Под. ред. А.А. Колосова. – М.: Радио и связь, 1989. – 288 с.: ил.
30. Петраков, А.В. Защитные информационные технологии аудиовидеоэлектросвязи: Учеб. Пособие [Текст] / А.В. Петраков. – М.: Машиностроение, 2010. – 616 с.: ил.
31. Пономарь М. О. Состояние разработки алгоритмов скрытого встраивания информации в просодические параметры речи [Текст] / Н.О. Пономарь // Вестник Московского государственного лингвистического университета. – 2011. - № 619. – С. 175-182.
32. Рабинер Л.Р. Цифровая обработка речевых сигналов [Текст] / Л.Р. Рабинер, Р.В. Шафер; под ред. М.В. Назарова, Ю.Н. Прохорова. – М.: Радио и связь, 1981. – 495 с.
33. Рандалл, Р. Б. Частотный анализ [Текст] / Р. Б. Рандалл. – Глоструп, Дания: К. Ларсен и сын А/О, 1989. – 389 с.
34. Раушер К. Основы спектрального анализа [Текст]: Пер. с англ. С. М. Смольского / Под ред. Ю. А. Гребенко - М.: Горячая линия-Телеком, 2006. – 225 с.
35. Романюк Ю.А. Дискретное преобразование Фурье в цифровом спектральном анализе. Учебное пособие [Текст] / Ю.А. Романюк – М.: МФТИ, 2007. – 120 с.
36. Салагай, М.О. Просодические средства защиты смысловой информации [Текст]: автореф. дис. на соиск. учен. степ. канд. филол. наук (10.02.21) / Салагай Марина Олеговна; Московский технический университет связи и информатики. – Москва, 2011. – 29 с.
37. Сапожников, М.А. Вокодерная связь [Текст]/ М.А. Сапожников, В.Г. Михайлов.-М: Радио и связь , 1983. – 248 с.: ил.

38. Сергиенко А.Б. Цифровая обработка сигналов [Текст] / А.Б. Сергиенко. – СПб.: Питер, 2003. – 604 с.
39. Смит С. Цифровая обработка сигналов: практическое руководство для инженеров и научных работников [Текст] / Стивен Смит ; пер. с англ. А. Ю. Линовича, С. В. Витязева. - Москва : Додэка-XXI, 2008. – 718 с.
40. Солони́на, А.И. Основы цифровой обработки сигналов [Текст]/ А.И. Солони́на, Д.А. Улахович, С.М. Арбузов, Е.Б.Соловьёва. – СПб.: БХВ-Петербург, 2005. – 768с.: ил.
41. Сорокин В.Н. Сегментация речи на кардинальные элементы [Текст] / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы, 2006, Т. 6, №3, с. 177-207.
42. Сюзев В.В. Основы теории цифровой обработки сигналов [Текст] / В.А. Сюзев – М.: РТСофт, 2014. – 752 с.
43. Трахтман А.М. Введение в обобщённую спектральную теорию сигналов [Текст] / А.М. Трахтман– М.: Изд-во «Советское радио», 1972. – 352 с.
44. Фланаган, Джеймс Л. Анализ, синтез и восприятие речи Speech analysis synthesis and perception [Текст] / Джеймс Л. Фланаган.;Пер. с англ. А.А. Пирогова; Под ред. А.А. Пирогова. – М.: Издательство «Связь», 1968. – 360с.: ил.
45. Харкевич А.А. Спектры и анализ [Текст] Изд. 5-е / А.А. Харкевич. – М.: Книжный дом «ЛИБРОКОМ», 2009. – 240 с.
46. Шелухин О.И. Цифровая обработка и передача речи [Текст] / О.И. Шелухин, Н.Ф. Лукьянцев; под ред. О.И. Шелухина. – М.: Радио и связь, 2000. – 456 с.
47. Kondoз А.М. Digital Speech. Coding for Low Bit Rate Communication Systems. – John Wiley & Sons, Ltd. 2004.– 442 p.

## ПРИЛОЖЕНИЕ А

### Алгоритм Левинсона - Дарбина [38]:

Вычисляются параметры предсказания первого порядка  $b_1^{(1)}$  (А.1) и средний квадрат ошибки предсказания  $E_1$  (А.2).

$$b_1^{(1)} = \frac{R_n(1)}{R_n(0)} \quad (\text{А.1})$$

$$E_1 = R_n(0) \cdot \left(1 - (b_1^{(1)})^2\right) \quad (\text{А.2})$$

Вычисляется добавляемый элемент вектора коэффициентов предсказания при  $k = \overline{2, p}$ :

$$b_k^{(k)} = \frac{R_n(k) - \rho_{k-1}^T \mathbf{V}_{k-1}^{(k-1)}}{E_{k-1}}, \quad (\text{А.3})$$

где  $\rho_k$  - перевёрнутый вектор - столбец значений корреляционной функции;

$\mathbf{V}_k^k$  - вектор - столбец коэффициентов предсказания  $k$ -го порядка, полученный на  $k$ -ой итерации;

$b_i^k$  -  $i$ -тый элемент вектора коэффициентов предсказания, полученный на  $k$ -ой итерации.

Производится обновление остальных коэффициентов предсказания:

$$\mathbf{V}_{k-1}^{(k)} = \mathbf{V}_{k-1}^{(k-1)} - \beta_{k-1}^{(k-1)} b_k^{(k)}, \quad (\text{А.4})$$

где  $\beta_k^k$  - перевернутый вектор - столбец коэффициентов предсказания  $k$ -го порядка, полученный на  $k$ -ой итерации.

Пересчитывается средний квадрат ошибки предсказания:

$$E_k = E_{k-1} \left( 1 - \left( b_k^{(k)} \right)^2 \right) \quad (\text{A.5})$$

При этом коэффициенты линейного предсказания определяются как:

$$a_n = B_n^{(p)}, n = \overline{1, p} \quad (\text{A.6})$$

**Листинг для VAD на основе сравнения с энергетическим порогом**

```

clc
clear
[X,Fs,bits]=wavread('путь_файла\имя_файла'); % загрузка файла
En=sum(X.^2);
N=(задать длину анализируемого отрезка);
H=(En/(length(X)/N));
sdv=5; % сдвиг окна
kol_vo_int=floor((length(X)-N)/sdv)+1;

i=0;
for n=1:kol_vo_int
    x=X(n*sdv-(sdv-1):n*sdv-(sdv-1)+(N-1));
    En_sh=sum(x.^2);
    if En_sh<H
        i=i+1;
        Sh(i)=En_sh;
        sr_Sh=mean(Sh);
        H=(En/(length(X)/N))+sr_Sh;
        Nomer(i)=n*sdv-(sdv-1);
    end
end

k=1;
kod(k,1)=Nomer(1);
for i=1:length(Nomer)-1
    if Nomer(i+1)-Nomer(i)>sdv
        k=k+1;
        kod(k,1)=Nomer(i+1);
        kod(k-1,2)=Nomer(i);
    end
end
kod(k,2)=length(X); % кодовая книга, содержащая границы пауз

```

**Листинг этапа обучения для VAD на основе коэффициентов  
линейного предсказания**

```

clc
clear
[X,Fs,bits]=wavread('путь_файла\имя_файла'); % загрузка файла
N=128;
p=15;
for nn=1:30000;
    sdv=1;
    x=X(1+sdv*(nn-1):N+sdv*(nn-1));

    for k=0:p
        s=0;
        for m=1:(N-k)
            s=s+(x(m))*(x(m+k));
        end
        RR(k+1)=s;
    end
    for k=1:p+1
        R(k)=RR(k)/RR(1);
    end
    clear k s m

    E(1)=R(1);
    ka(1)=R(2)/E(1);
    al(1,1)=ka(1);
    for i=2:p
        E(i)=(1-ka(i-1)^2)*E(i-1);
        s=0;
        for j=1:(i-1)
            s=s+al(i-1,j)*R(i-j+1);
        end
        ka(i)=(R(i+1)-s)/E(i);
        al(i,i)=ka(i);
        for j=(i-1):-1:1
            al(i,j)=al(i-1,j)-ka(i)*al(i-1,i-j);
        end
    end
    clear E i s j ka R

    for i=1:p

```

```

    alfa(i)=al(p,i);
end
clear i al

```

```

for i=0:p
    Aa=0;
    for k=1:(p-i)
        Aa=Aa+(alfa(k))*(alfa(i+k));
    end
    A(i+1)=Aa;
end
clear Aa i k

```

```

s=0;
ss=0;
for i=1:p
    s=s+A(i+1)*(RR(i+1)/RR(1));
    ss=ss+A(i+1)*RR(i+1);
end
Df(nn)=A(1)*RR(1)+2*s;
E(nn)=A(1)*RR(1)+2*ss;
for i=1:length(Df)-1
    Dff=abs(Df(i+1)-Df(i));
end
clear s ss
end

```

**Df=max(Dff) %пороговое значение разницы автокорреляции сигнала**  
**E=max(E) % пороговое значение остаточной энергии**

**Листинг для VAD на основе коэффициентов линейного предсказания**

```

clc
clear
N=256;

Df_h=(пороговое значение);
E_h=(пороговое значение);
p=8;
Df1=(сначала вычислить Df1 для первого отрезка анализа. Вычисляется как Df2 при nn=1);

```

```

[X,Fs,bits]=wavread('путь_файла\имя_файла'); % загрузка файла
Y=zeros(1,length(X));

```

```

for nn=2:floor(length(X)/N)
x=X(nn*N-(N-1):nn*N);

```

```

for k=0:p
s=0;
for m=1:(N-k)
s=s+(x(m))*(x(m+k));
end
RR(k+1)=s;
End

```

```

for k=1:p+1
R(k)=RR(k)/RR(1);
end
clear k s m

```

```

E(1)=R(1);
ka(1)=R(2)/E(1);
al(1,1)=ka(1);
for i=2:p
E(i)=(1-ka(i-1)^2)*E(i-1);
s=0;
for j=1:(i-1)
s=s+al(i-1,j)*R(i-j+1);
end

```

```

ka(i)=(R(i+1)-s)/E(i);
al(i,i)=ka(i);
for j=(i-1):-1:1
    al(i,j)=al(i-1,j)-ka(i)*al(i-1,i-j);
end
end
clear E i s j ka Ry

for i=1:p
    alfa(i)=al(p,i);
end
clear i al

for i=0:p
    Aa=0;
    for k=1:(p-i)
        Aa=Aa+(alfa(k))*(alfa(i+k));
    end
    A(i+1)=Aa;
end
clear Aa i k

s=0;
ss=0;
for i=1:p
    s=s+A(i+1)*(RR(i+1)/RR(1));
    ss=ss+A(i+1)*RR(i+1);
end
Df2=A(1)*RR(1)+2*s;
E=A(1)*RR(1)+2*ss;
clear s ss

if abs(Df2-Df1)>Df_h
    Y(nn*N-(N-1):nn*N)=1;
else if E>E_h
    Y(nn*N-(N-1):nn*N)=1;
    else Y(nn*N-(N-1):nn*N)=0;
end
end
Df1=Df2;
end

```

Y – кодовая книга, состоящая из {0} {1}, где {0} кодируется пауза, а {1} – активная речь

**Листинг для обнаружения пауз на основе субполосного анализа**

```

clc
clear
close all
%%
[X,Fs,bits]=wavread('путь_файла\имя_файла'); % загрузка файла
X=X';
R=% 'количество интервалов';
N=128; % длина отрезка

u=[0:pi/R:pi];
for r=1:R
    for i=1:N
        for k=1:N
            if i~=k
                A(i,k,r)=(sin(u(r+1)*(i-k))-sin(u(r)*(i-k)))/(pi*(i-k));
            else    A(i,k,r)=(u(r+1)-u(r))/pi;
            end
        end
    end
end

n1='отсчёт начала паузы'
n2='отсчёт конец паузы'
pause=X(n1:n2);
k=floor((length(pause))/N);
for i=1:k
    R_pause=pause(N*(i-1)+1:N*i);
    for r=1:R
        P_pause(i,r)=R_pause*A(:,r)*R_pause';
    end
end
Ps_pause=sum(P_pause)/k;
for i=1:k
    for r=1:R
        S_pause(i,r)=P_pause(i,r)/Ps_pause(r);
    end
    Sm_pause(i)=max( S_pause(i,:));
end
S=sum(Sm_pause)/k;
D=sqrt(sum(Sm_pause.^2)/k-S^2);

```

```

alpha=0.1; % вероятность тревоги
m=0; a=2;
h=S+D/a*sqrt(alpha);
beta=0;
for i=1:k
    b=Sm_pause(i)-h;
    if b>0
        beta = beta+1;
    end
end
while abs(alpha-beta)>alpha
    if alpha<beta
        a=(1-alpha*a)*a;
    elseif alpha>beta
        a=(1+alpha*a)*a;
    end
    h=S+D/a*sqrt(alpha);
    beta=0;
    for i=1:k
        b=Sm_pause(i)-h;
        if b>0
            beta = beta+1;
        end
    end
end
end;

s1=0;
k=floor((length(X))/N);
for i=1:k
    R_sign=X(N*(i-1)+1:N*i);
    for r=1:R
        P_sign(i,r)= R_sign*A(:,r)* R_sign';
        S_sign(i,r)=P_sign(i,r)/Ps_pause(r);
    end
    Sm_sign(i)=max(S_sign(i,:));
    if Sm_sign(i)>h
        XX(N*(i-1)+1:N*i)=X(N*(i-1)+1:N*i);
    else
        XX(N*(i-1)+1:N*i)=zeros(1,N);
        s1=s1+1;
        M(s1,1)=N*(i-1)+1;
        M(s1,2)=N*i;
    end
end
end

```

Результаты работы метода VAD на основе сравнения с энергетическим порогом при изменении параметров.

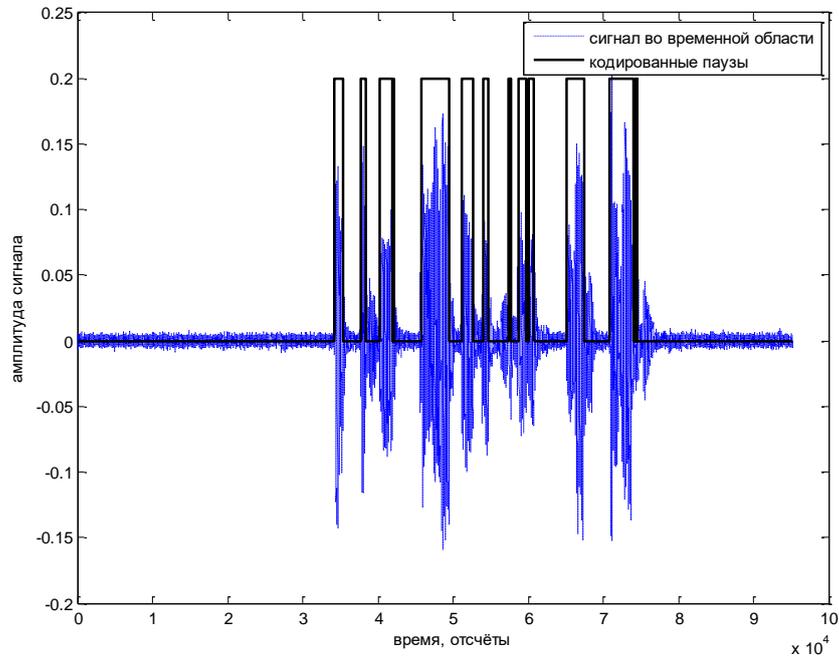


Рисунок Е.1 А – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное во временной области при  $N=128$  отсчетов

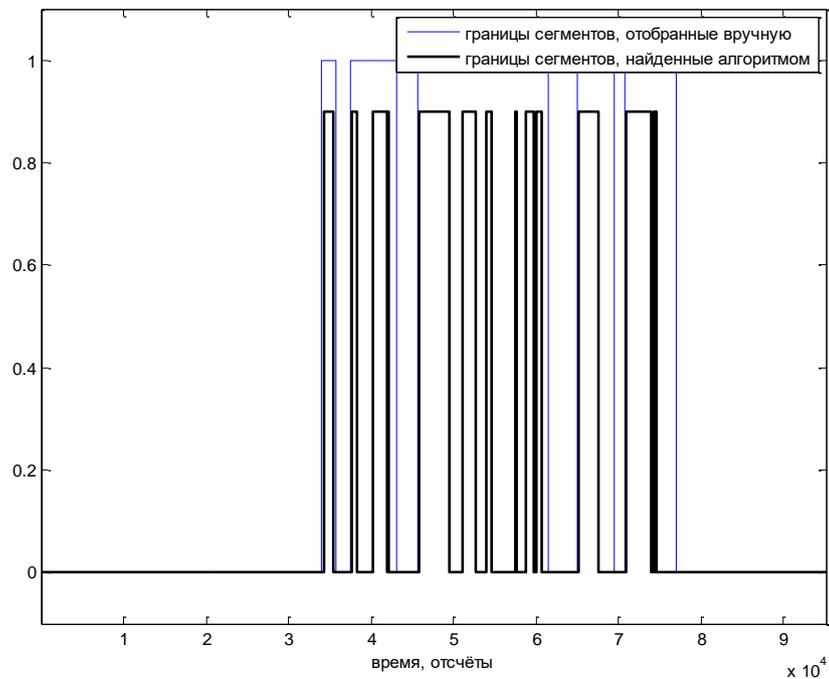
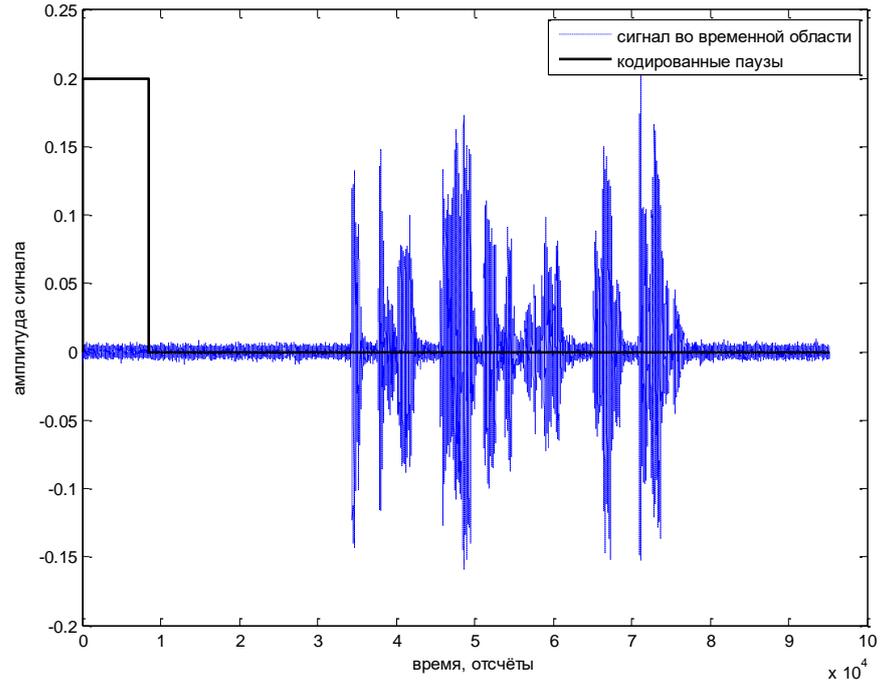
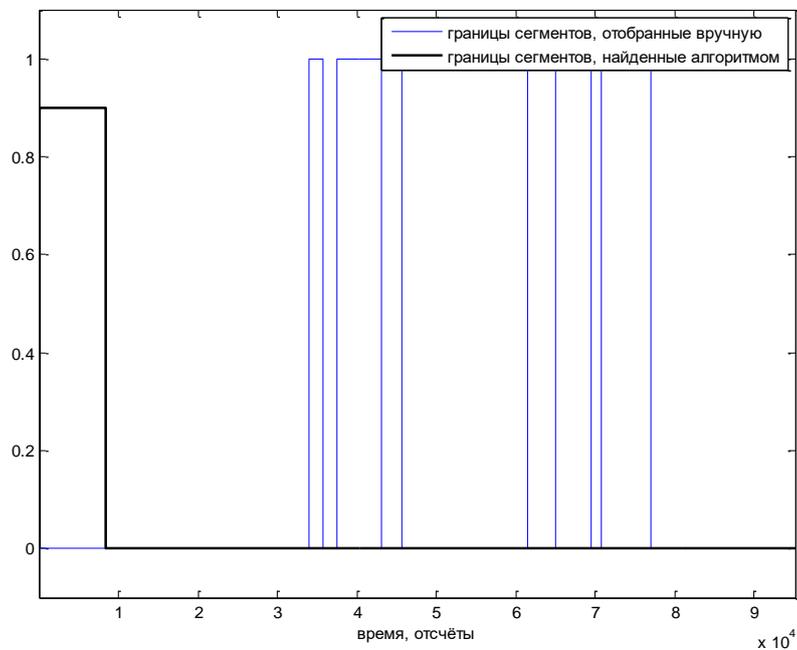


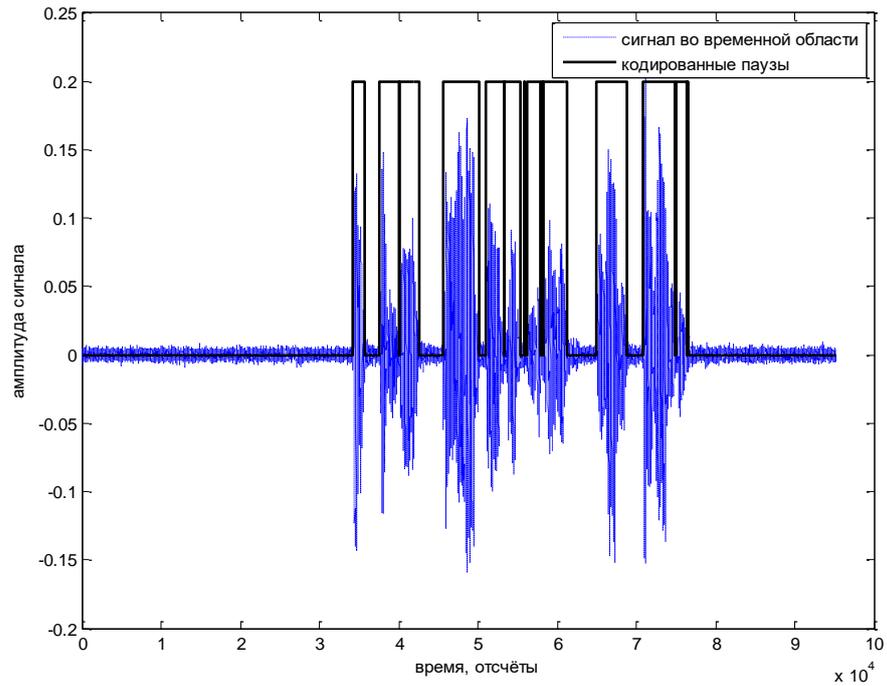
Рисунок Е.1 Б – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное в виде двоичной последовательности при  $N=128$  отсчетов



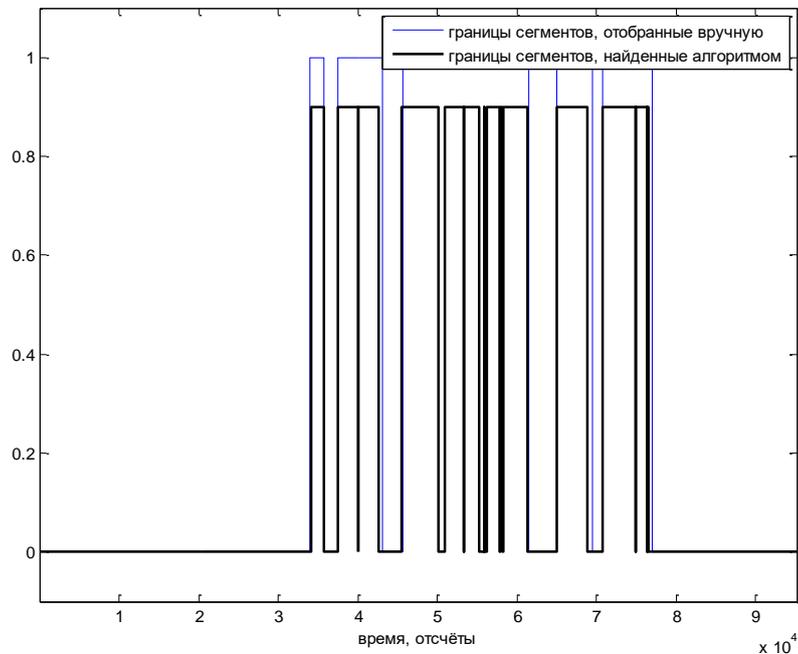
**Рисунок Е.2 А – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное во временной области при  $N=128$  отсчётов,  $H=K \cdot P_{cp}$  ( $K=0,5$ )**



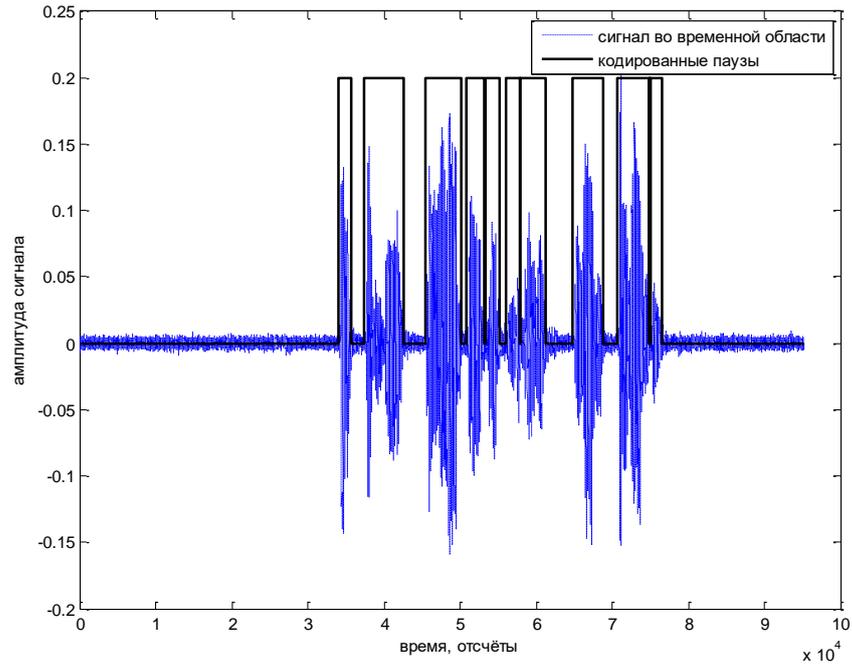
**Рисунок Е.2 Б – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное в виде двоичной последовательности при  $N=128$  отсчётов,  $H=K \cdot P_{cp}$  ( $K=0,5$ )**



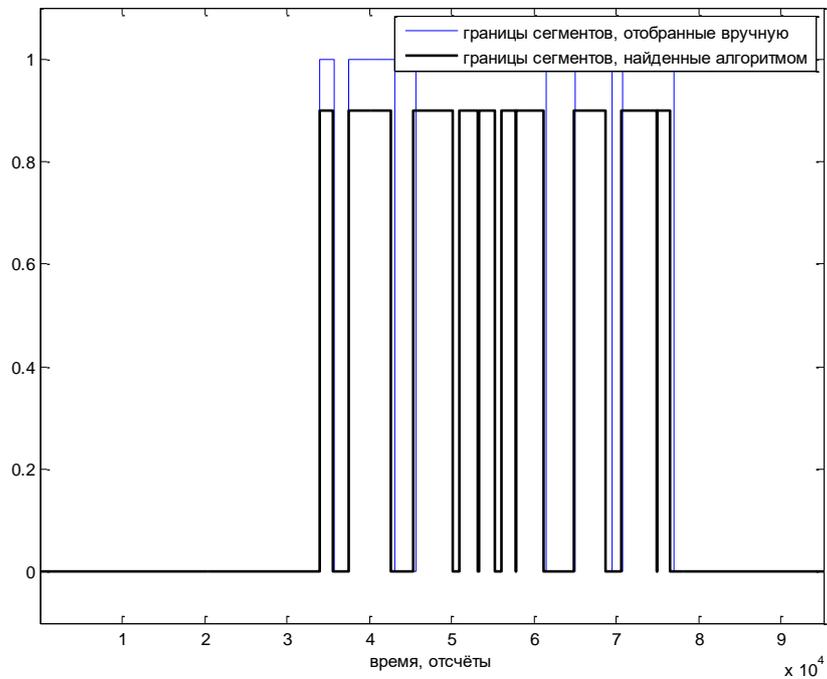
**Рисунок Е.3 А – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное во временной области при  $N=128$  отсчётов,  $H=K \cdot P_{cp}$  ( $K=5$ )**



**Рисунок Е.3 Б – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное в виде двоичной последовательности при  $N=128$  отсчётов,  $H=K \cdot P_{cp}$  ( $K=5$ )**



**Рисунок Е.4 А – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное во временной области при  $N=256$  отсчётов,  $H=K \cdot P_{cp}$  ( $K=5$ )**



**Рисунок Е.4 Б – Кодирование пауз методом VAD на основе сравнения с энергетическим порогом, представленное в виде двоичной последовательности при  $N=256$  отсчётов,  $H=K \cdot P_{cp}$  ( $K=5$ )**

Результаты работы метода VAD на основе коэффициентов линейного предсказания.

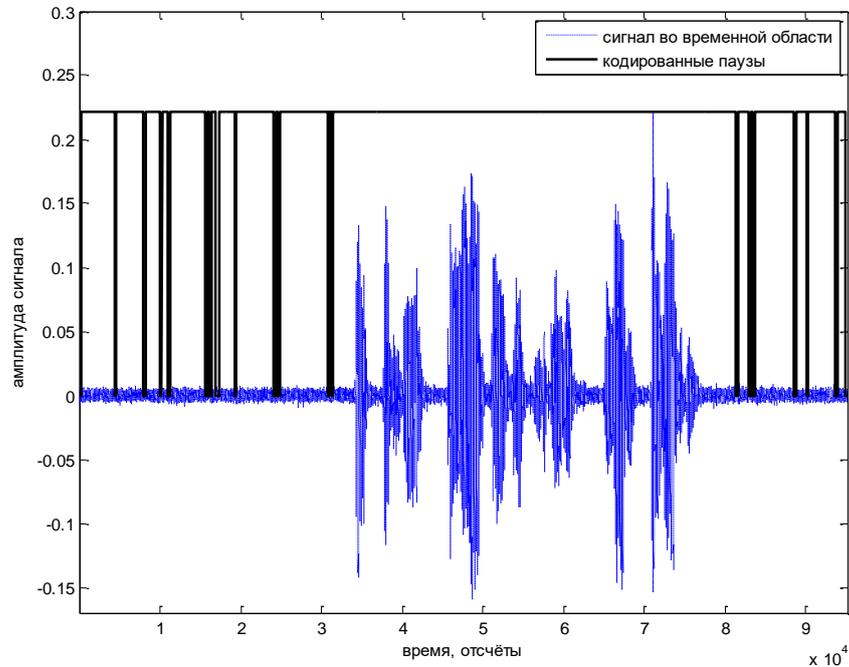


Рисунок Ж.1 А – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное во временной области ( $N=256, p=10$ )

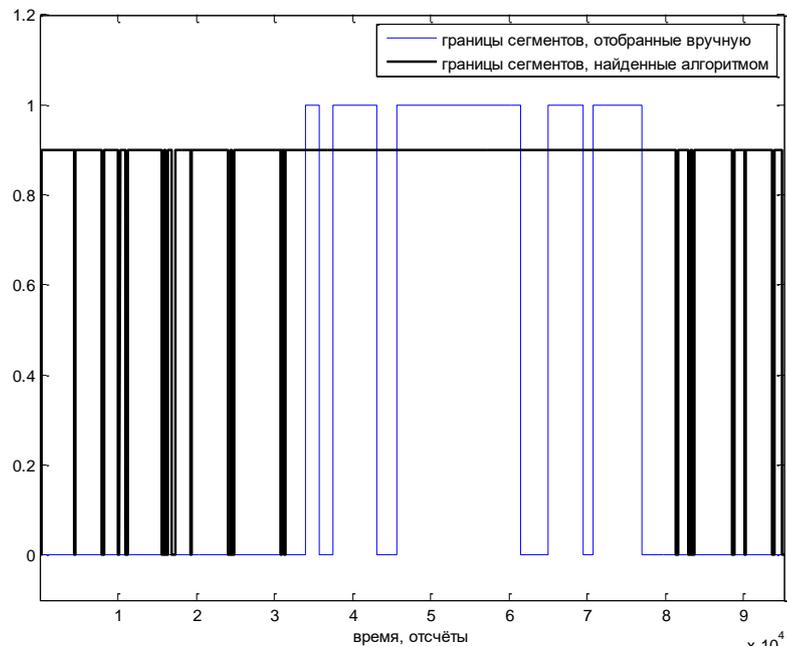
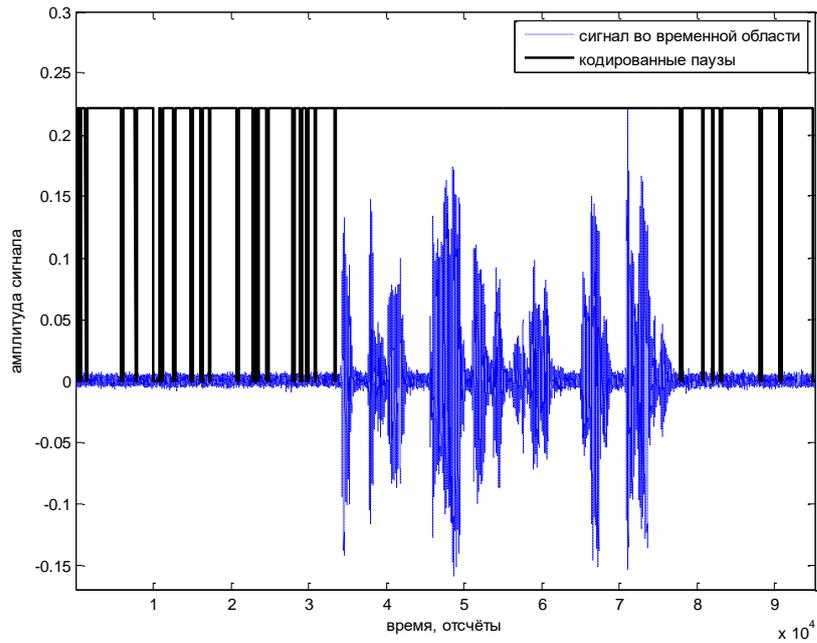
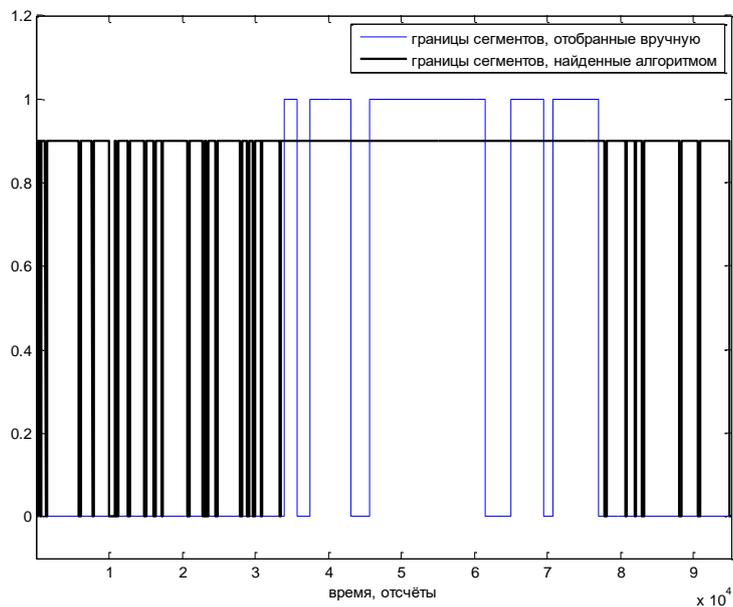


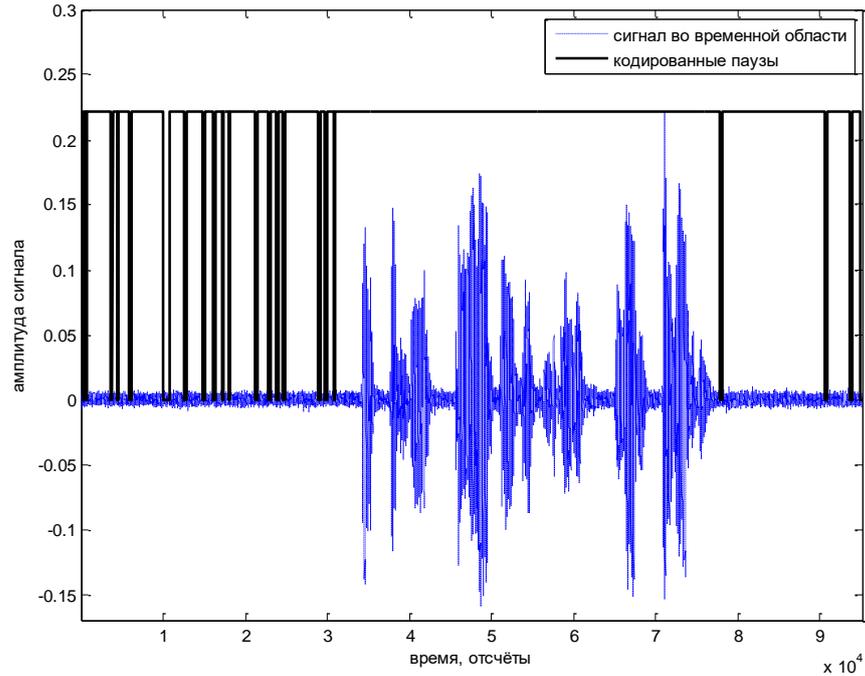
Рисунок Ж.1 Б – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное в виде двоичной последовательности ( $N=256, p=10$ )



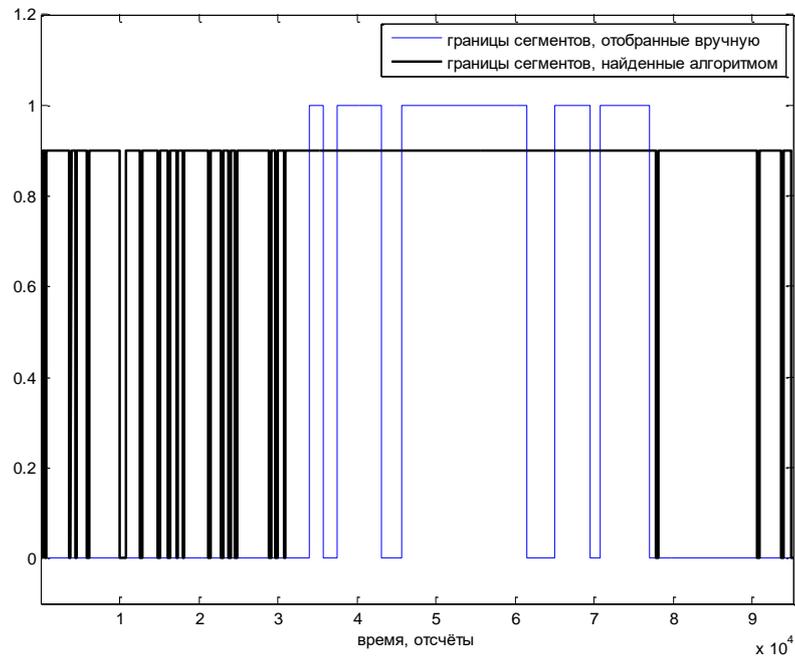
**Рисунок Ж.2 А – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное во временной области ( $N=256, p=15$ )**



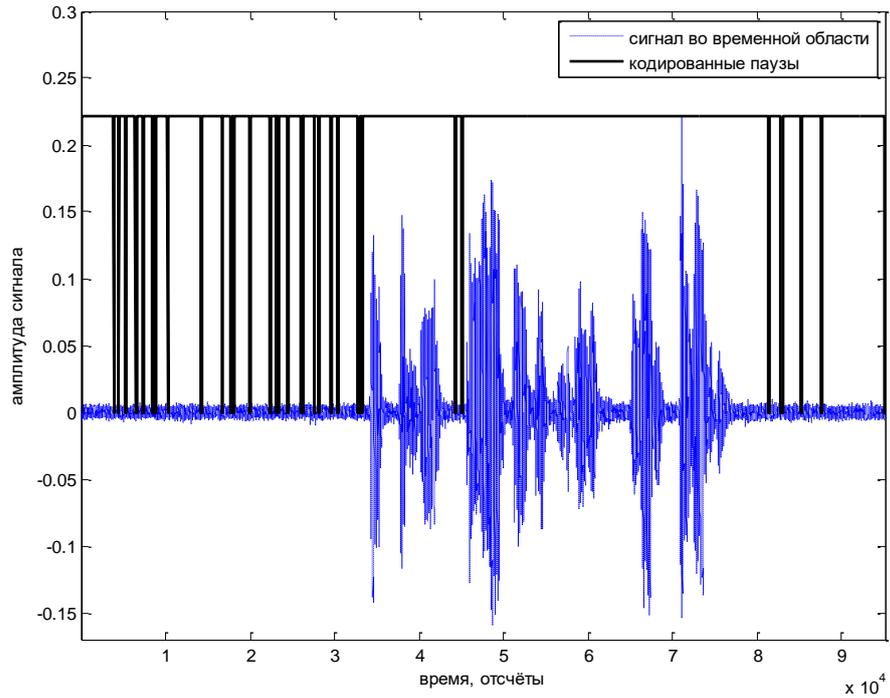
**Рисунок Ж.2 Б – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное в виде двоичной последовательности ( $N=256, p=15$ )**



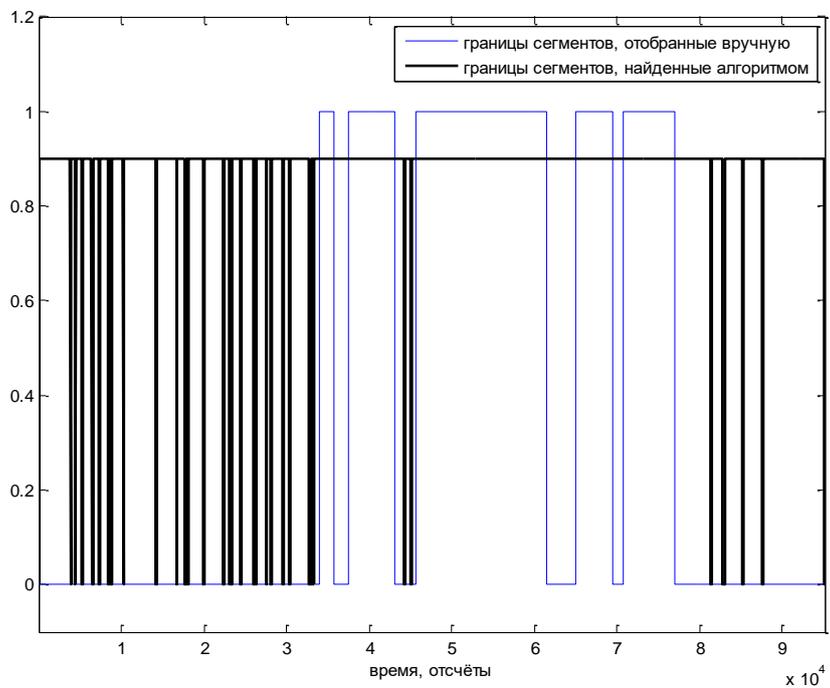
**Рисунок Ж.3 А – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное во временной области ( $N=256, p=20$ )**



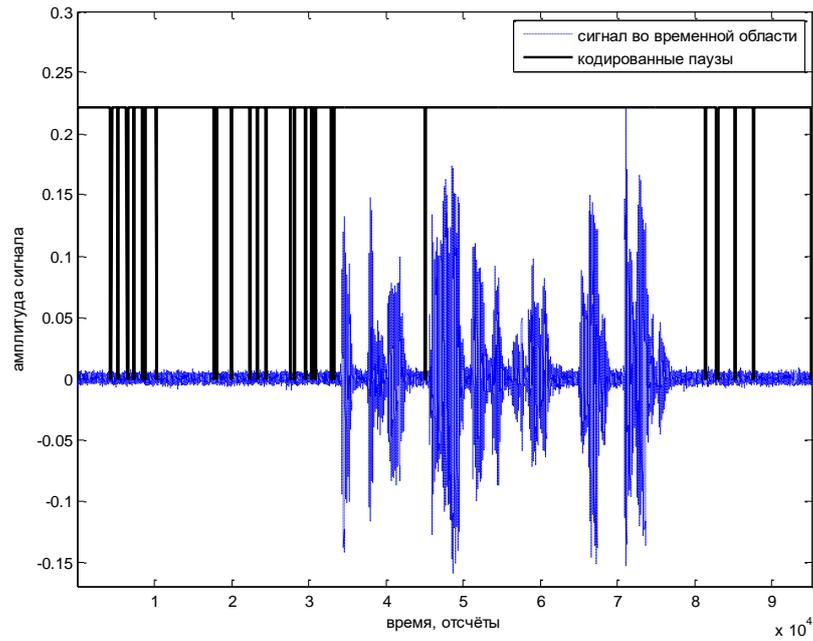
**Рисунок Ж.3 Б – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное в виде двоичной последовательности ( $N=256, p=20$ )**



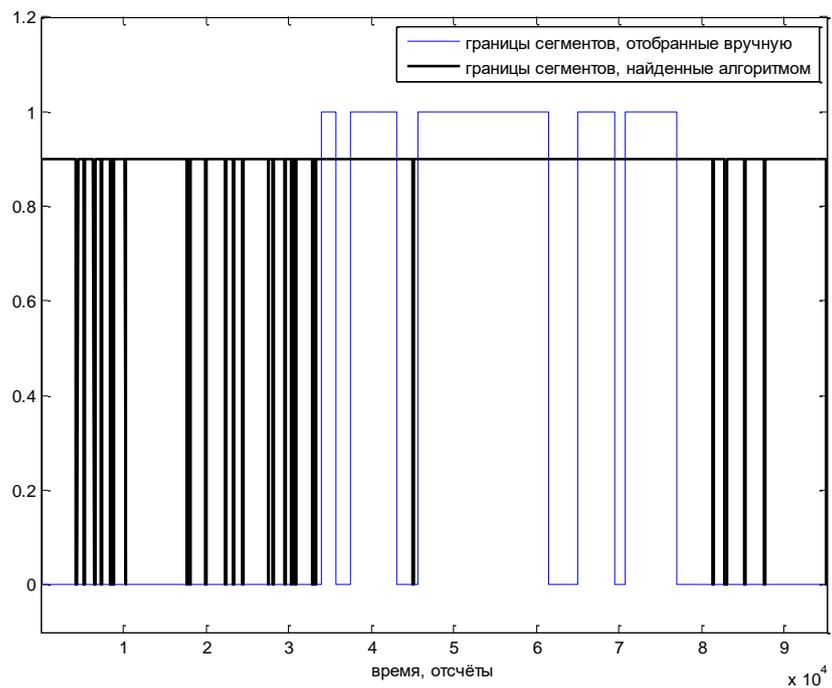
**Рисунок Ж.4 А – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное во временной области ( $N=128, p=8$ )**



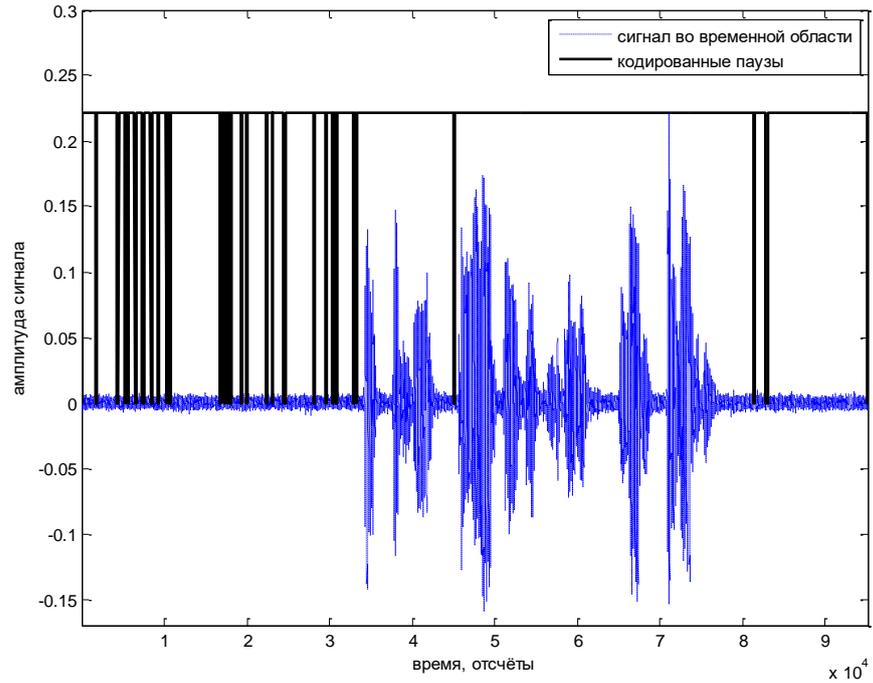
**Рисунок Ж.4 Б – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное в виде двоичной последовательности ( $N=128, p=8$ )**



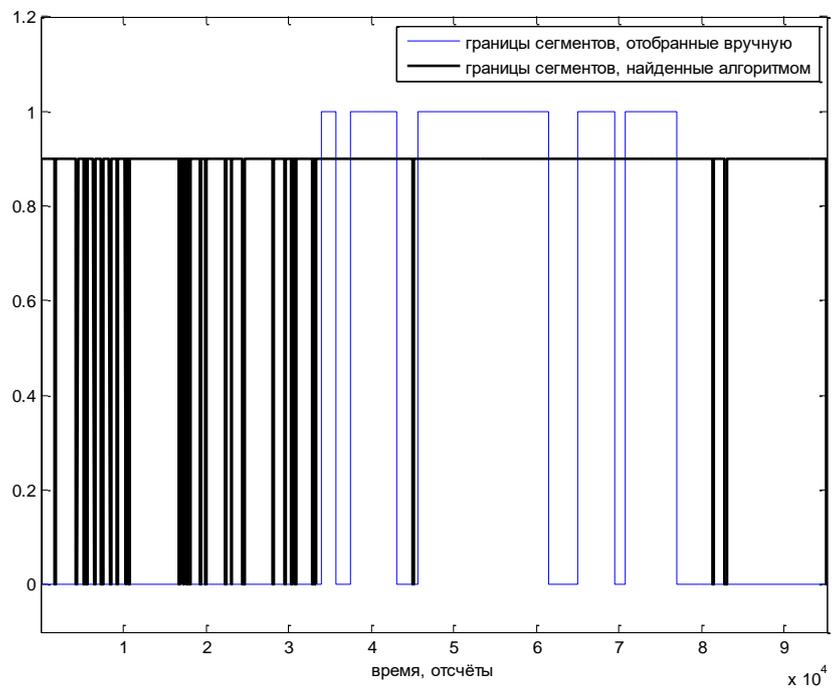
**Рисунок Ж.5 А – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное во временной области ( $N=128, p=10$ )**



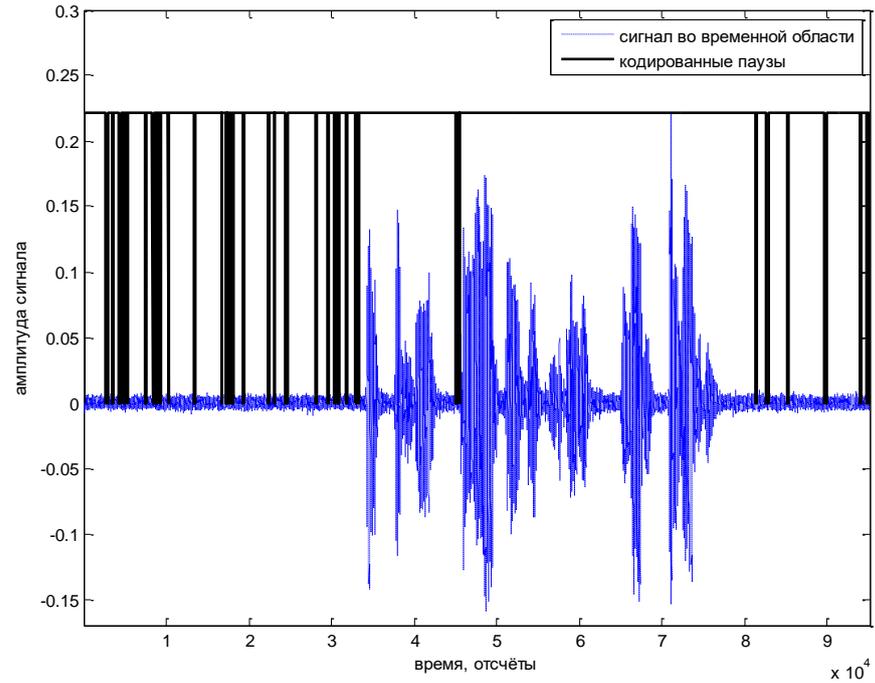
**Рисунок Ж.5 Б – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное в виде двоичной последовательности ( $N=128, p=10$ )**



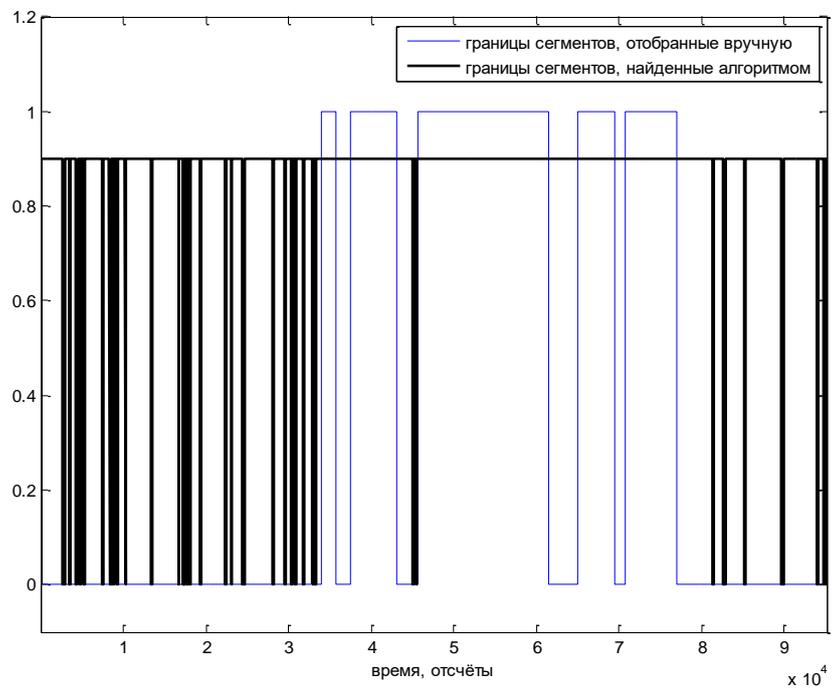
**Рисунок Ж.6 А – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное во временной области ( $N=128, p=15$ )**



**Рисунок Ж.6 Б – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное в виде двоичной последовательности ( $N=128, p=15$ )**



**Рисунок Ж.7 А – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное во временной области ( $N=128, p=20$ )**



**Рисунок Ж.7 Б – Кодирование пауз методом VAD на основе коэффициентов линейного предсказания, представленное в виде двоичной последовательности ( $N=128, p=20$ )**

Результаты работы метода обнаружения пауз на основе субполосного анализа при изменении параметров.

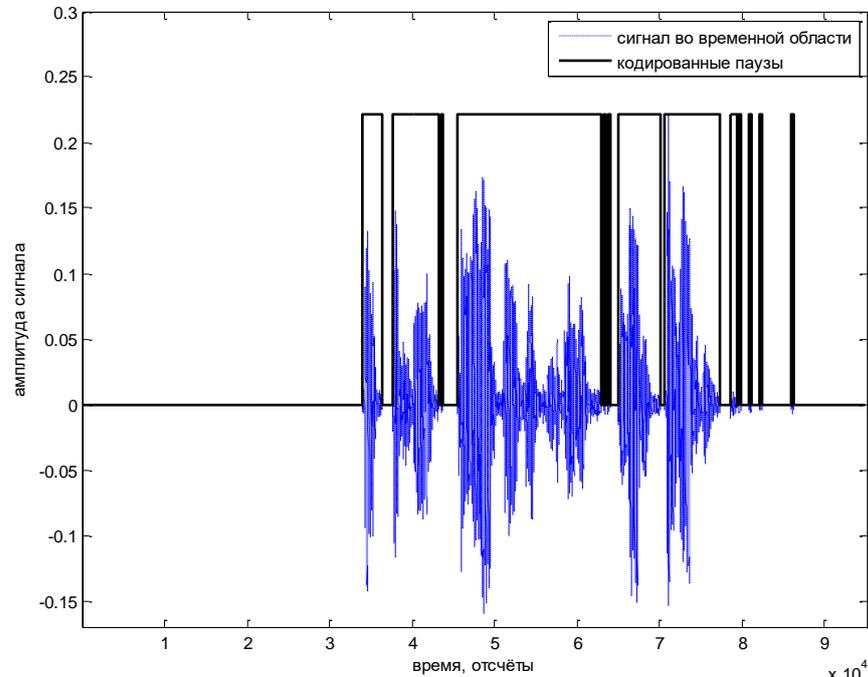


Рисунок 3.1 А – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=256$ ,  $R=32$ )

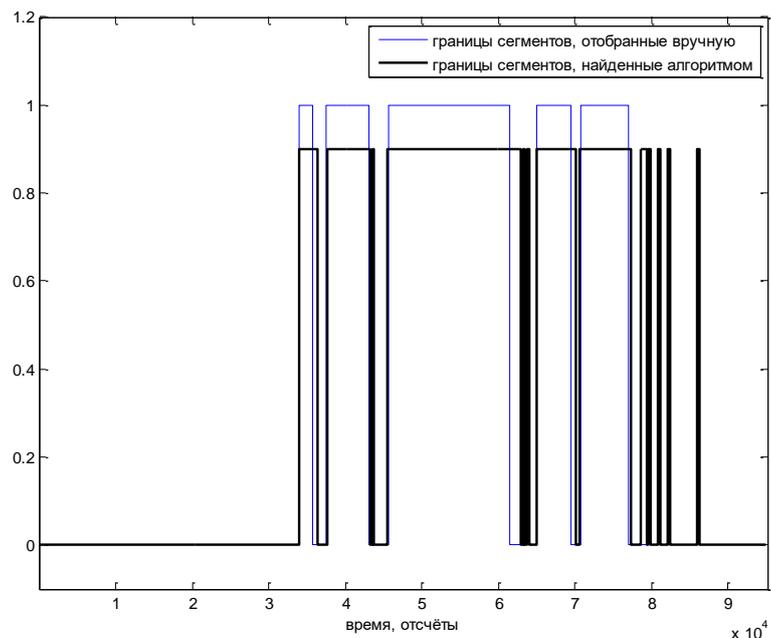
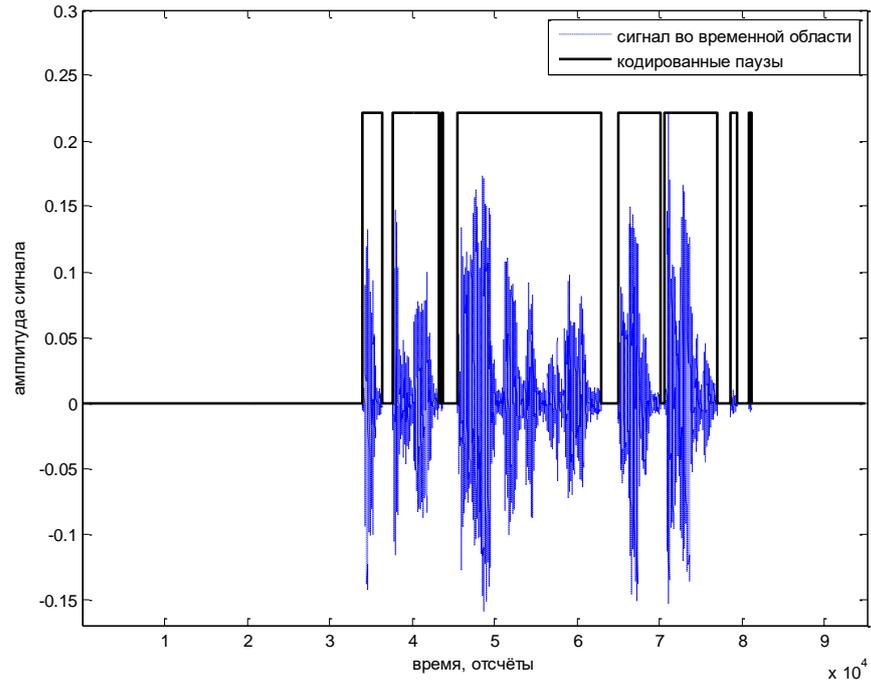
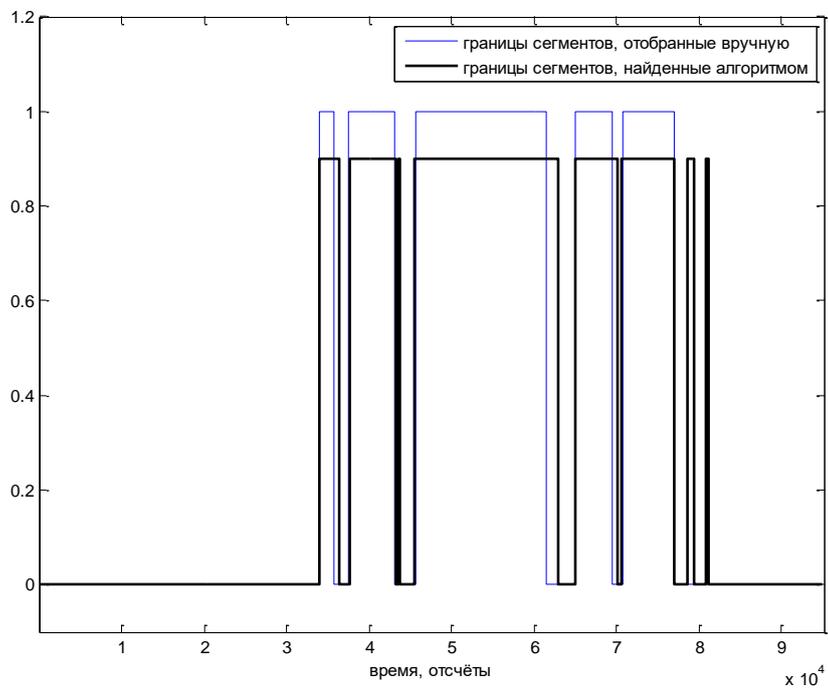


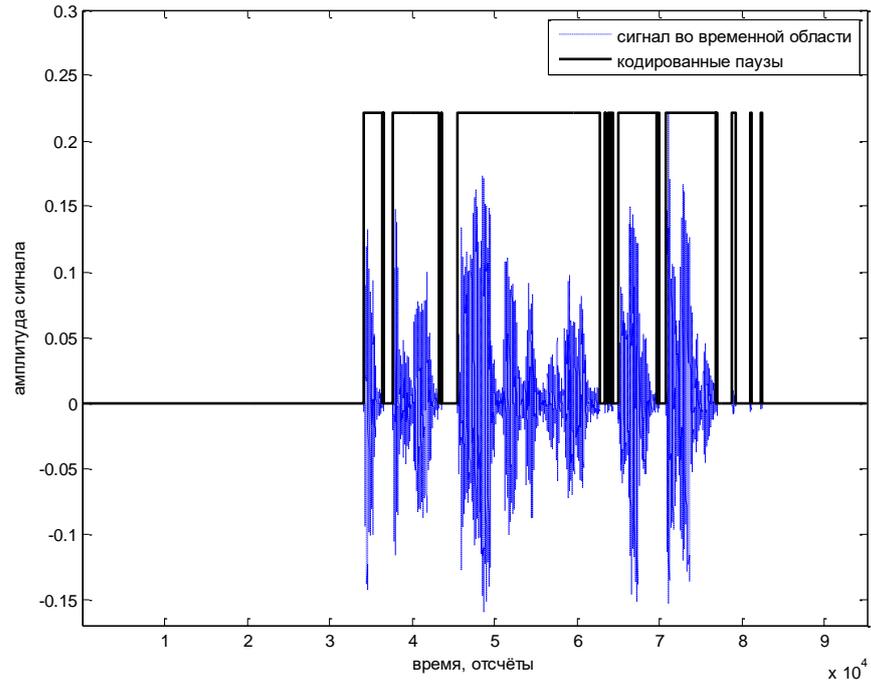
Рисунок 3.1 Б – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=256$ ,  $R=32$ ), представленное в виде двоичной последовательности



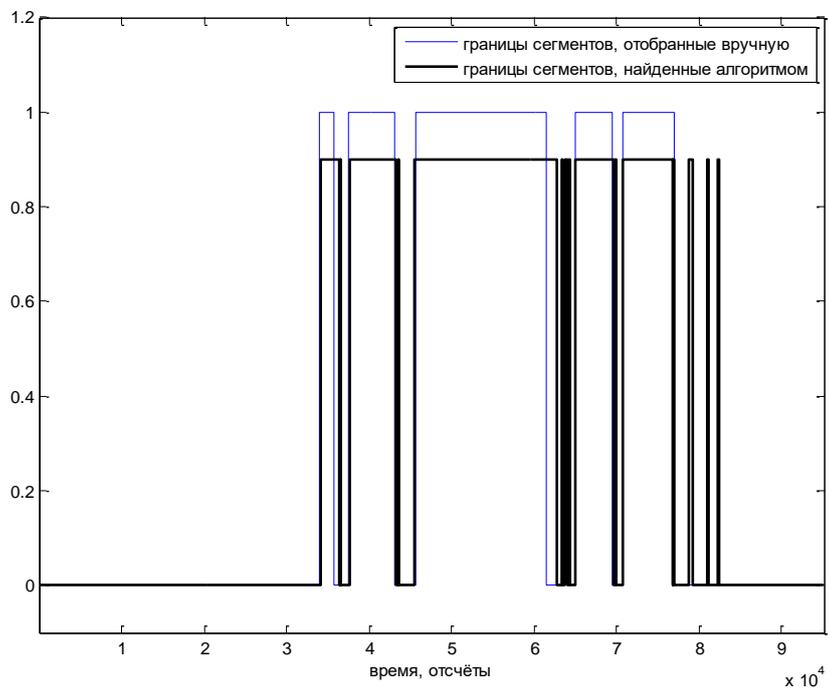
**Рисунок 3.2 А – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=256$ ,  $R=64$ )**



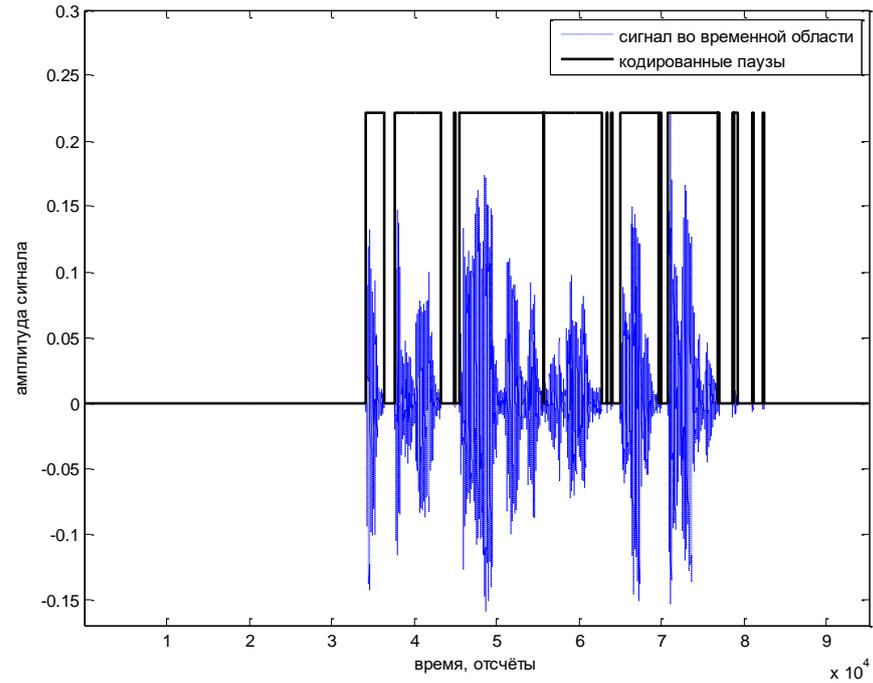
**Рисунок 3.2 Б – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=256$ ,  $R=64$ ), представленное в виде двоичной последовательности**



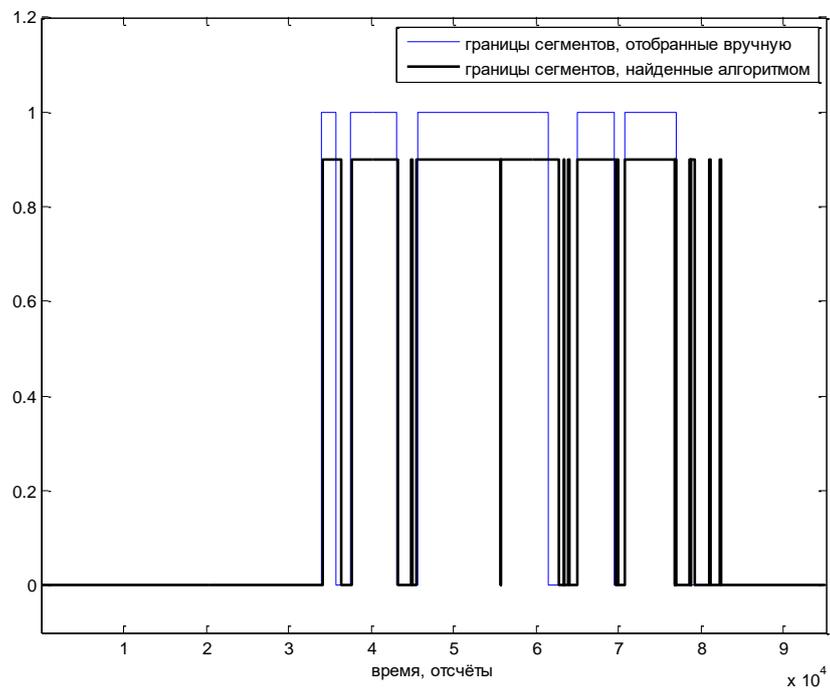
**Рисунок 3.3 А – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=128, R=16$ )**



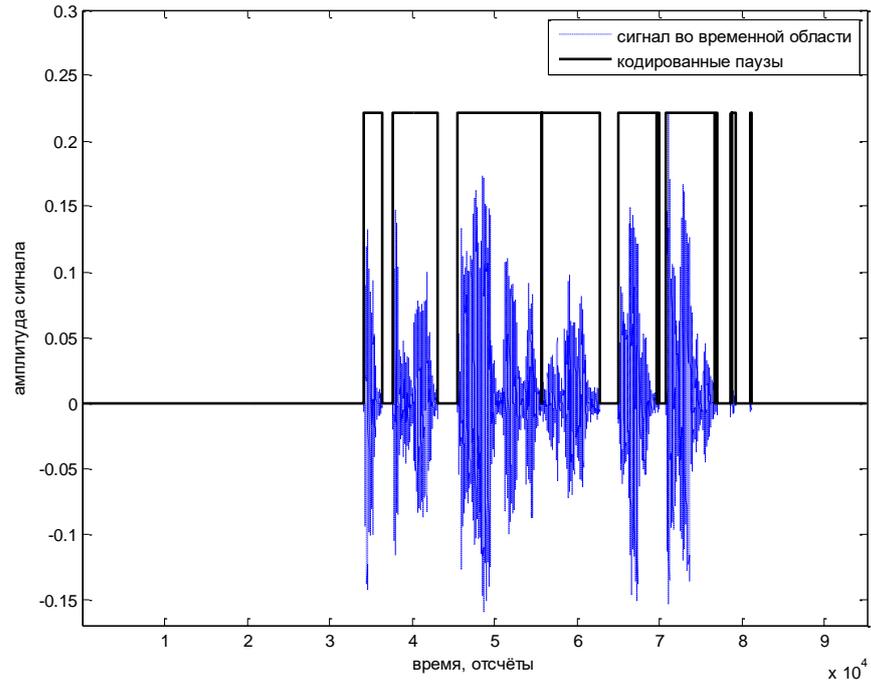
**Рисунок 3.3 Б – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=128, R=16$ ), представленное в виде двоичной последовательности**



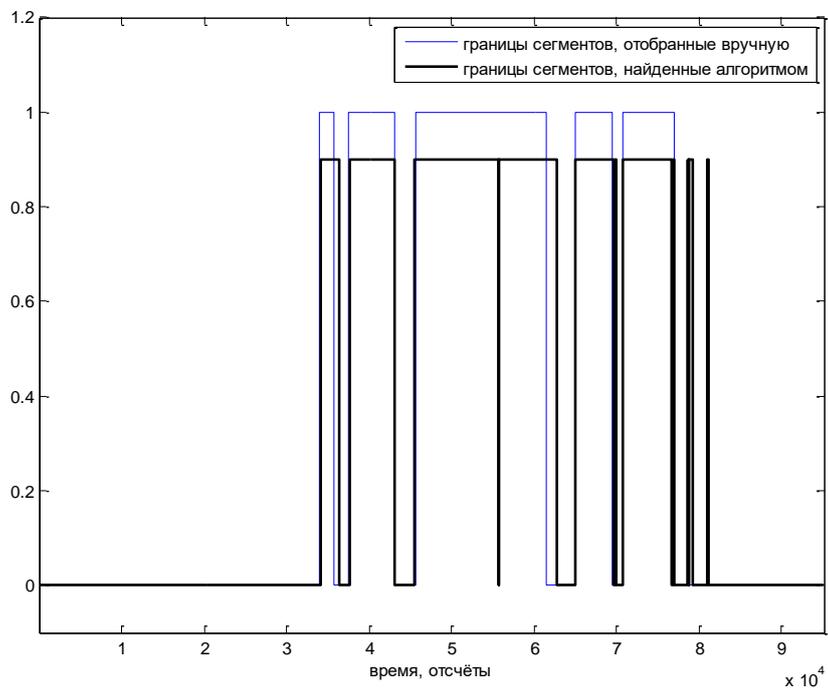
**Рисунок 3.4 А – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=128, R=32$ )**



**Рисунок 3.4 Б – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=128, R=32$ ), представленное в виде двоичной последовательности**



**Рисунок 3.5 А – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=128, R=64$ )**



**Рисунок 3.5 Б – Кодирование пауз методом обнаружения пауз на основе субполосного анализа ( $N=128, R=64$ ), представленное в виде двоичной последовательности**

Магистерская диссертация выполнена мной совершенно самостоятельно. Все использованные в работе материалы и концепции из опубликованной научной литературы и других источников имеют ссылки на них.

«\_\_\_» \_\_\_\_\_ г.

---

*(подпись)*

---

*(Ф.И.О.)*