

ИССЛЕДОВАНИЕ АЛГОРИТМА ОБНАРУЖЕНИЯ ПАУЗ В РЕЧЕВЫХ СООБЩЕНИЯХ

А.С. Белов

Белгородский государственный университет, 308015, г. Белгород, ул. Победы, 85, Belov@bsu.edu.ru

Предложен способ обнаружения пауз в речевых сообщениях, основанный на учете различий в распределении частотных энергетических составляющих спектра сигнала в речи от составляющих спектра сигнала в паузе, применяемый для очистки входного речевого сигнала от шумов.

Ключевые слова: частотные представления, способ обнаружения пауз в речевых сообщениях, КИХ-фильтры, алгоритм вычисления точного значения доли энергии отрезка сигнала, которая сосредоточена в любом конечном частотном интервале, оценивание вероятностей правильного обнаружения пауз и ложного их обнаружения.

ВВЕДЕНИЕ

Потерей слуха в той или иной форме страдают многие люди. Эта потеря становится критической, когда начинает препятствовать нормальному речевому общению. Возрастное ухудшение слуха, перенесенные болезни уха, звуковые травмы, ототоксическое действие некоторых лекарств и другие причины приводят к тому, что около 2% населения Земли для того, чтобы быть полноценными членами общества, нуждаются в применении слуховых аппаратов (СА), которые до определённой степени позволяют компенсировать ухудшение физиологических характеристик слуховой системы человека.

Для достижения этой цели в СА осуществлена специальная обработка сигналов, регистрируемых их антеннами. Основным элементом СА является усилитель звука. К нему подключены микрофон, воспринимающий внешние звуки (в последних моделях СА используются микрофоны с автоматической настройкой диаграмм направленности), и телефон (динамик), воспроизводящий усиленный звук. Одной из проблем является селекция участков обработки сигналов в зависимости от того, принадлежат они звуку речи или паузе.

В современных цифровых СА используют методы обнаружения пауз, основанные на статистическом анализе входного сигнала и распознавании речи от шума, которые контролируют распределение интенсивности входного сигнала на коротких временных интервалах в 15 каналах. С помощью КИХ-фильтров с большой погрешностью определяют долю энергии в заданном частотном интервале, что не позволяет точно определить отношение сигнал/шум и не позволяет полностью использовать энергетические характеристики сигнала при принятии решения отношения анализируемого отрезка сигнала к паузе или звуковым данным.

Таким образом, становится ясно, что используемые алгоритмы неоптимальны, с точки зрения адекватного описания энергетических характеристик сигналов речевых данных.

Целью данной работы является разработка и исследование алгоритма, предназначенного для обработки звуковых сигналов при обнаружении и удалении пауз на основе математического аппарата [1], и построение алгоритма вычисления точного значения доли энергии отрезка сигнала, которая сосредоточена в любом конечном частотном интервале. Это дает возможность адекватно учесть свойство сосредоточенности энергии звуков речи в малом количестве достаточно узких частотных интервалов и принимать решение об отсутствии паузы на основе сравнения попадающих в них долей энергий отрезков со звуковыми данными (и аддитивным шумом) и отдельно шумов. Ясно, что тем самым достигается максимальный энергетический выигрыш, а следовательно, минимизируется вероятность принять отрезки со звуками речи за паузу.



Новизна заключается в том, что используется метод вычисления точных значений долей энергии, попадающих в заданные интервалы частотной полосы, и учитывается различие в распределении частотных энергетических составляющих спектра сигнала в речи от составляющих спектра сигнала в паузе (шума), точно вычисляют их значение в каждом из анализируемых достаточно узких частотных интервалов [1,2].

Экспериментальные исследования проводятся с целью проверки работоспособности алгоритма сжатия речевых данных за счет обнаружения и кодирования пауз на основе сравнения распределений энергии шума и смеси сигнал+шум в заданных частотных интервалах.

В частности, проведено оценивание вероятностей правильного обнаружения пауз (ПОП), ложного обнаружения пауз (ЛОП) и коэффициента сжатия.

В основе вычислительных экспериментов по обработке отрезков речевых сигналов лежит разработанный алгоритм сжатия за счёт кодирования пауз.

I. *Формируем основную гипотезу H_0* : анализируемый отрезок $\vec{x} = (x_1, \dots, x_n)^T$ длительностью N отсчетов принадлежит паузе в речевых сообщениях и используется следующее решающее правило.

Если

$$S = \max(P_r / P_r^\Pi) > h_\alpha, \quad r = 1, \dots, R; \quad (1)$$

то гипотеза отвергается.

Здесь

$$P_r = \frac{1}{2\pi} \int_{\omega \in V_r} |X(\nu)|^2 d\nu = \vec{x}^T A_r \vec{x}; \quad (2)$$

доля энергии анализируемого отрезка в r -том частотном интервале

$$V_r = [-\nu_r, -\nu_{r-1}] \cup [\nu_{r-1}, \nu_r]; \quad (3)$$

A_r – субполосная матрица:

$$A_r = \{a_{ik}^r\}, \quad i, k = 1, \dots, N; \\ a_{ik}^r = \frac{\{\sin(\nu_r(i-k)) - \sin(\nu_{r-1}(i-k))\}}{\pi(i-k)}, \quad i, k = 1, \dots, N; \\ \nu_r - \nu_{r-1} = \frac{\pi}{R}; \quad (4)$$

R – общее количество частотных интервалов;

P_r^Π – оценка математического ожидания доли энергии сигнала в паузе в соответствующем частотном интервале;

h_α – порог, удовлетворяющий условию

$$\int_0^{h_\alpha} W(S) dS \geq 1 - \alpha, \quad (5)$$

$1 - \alpha$ – выбранный уровень правильного обнаружения пауз (α – вероятность ложных тревог); W – функция плотности вероятности решающей функции S .

II. *Так же, как и в реальных информационных технологиях, для оценивания P_r^Π и h_α используется этап обучения*. На этапе обучения осуществляется:

1) вычисление среднего значения энергии сигнала, заведомо относящегося к паузе по формуле

$$P_r^\Pi = \sum_{k=1}^{N_y} (P_r)_k^\Pi / N_y, \quad (6)$$

где N_y – количество отрезков сигнала в паузе, которые используются для усреднения (обучения), что соответствует оцениванию математических ожиданий вычисляемых долей энергий в соответствующих частотных интервалах;



$(P_r)_k^\Pi$ – энергия анализируемого отрезка, относящегося к паузе в r -том частотном интервале, вычисляемая по формуле (12).

2) На основе неравенства Чебышева проводится итерация по определению порога, обеспечивающего заданный уровень вероятности ложной тревоги

$$h_\alpha \leq \bar{S}_\Pi + a * D_\Pi / \sqrt{\alpha}, \quad (7)$$

где h_α – порог, α – заданный уровень вероятности ложной тревоги, a – коэффициент, значение которого больше 2-х, а его конкретная оценка проводится в процессе обучения по следующему принципу:

$$\beta_m = \frac{\sum_{k=1}^{N_y} \text{sig}(S_k^\Pi - h_\alpha^m)}{N_y}, \quad \text{sig}(x) = 1, x > 0; \text{sig}(x) = 0, x \leq 0.$$

Если $|\alpha - \beta_m| \leq \alpha^2$, то $h_\alpha = h_\alpha^m$, и необходимо прекратить итерации.

В противном случае при $\alpha > \beta_m$ положить $a_{m+1} := (1 - \alpha \times a_m) \times a_m$, если же выполняется неравенство $\alpha < \beta_m$, то положить $a_{m+1} := (1 + \alpha \times a_m) \times a_m$, положить $m = m + 1$ и продолжить итерации.

До этого при использовании обучающей выборки относящихся к паузе данных вычисляются оценки математических ожиданий вида (6). Затем вычисляются оценки математического ожидания и дисперсии решающей функции:

$$\bar{S}_\Pi = \sum_{k=1}^{N_y} (S_k^\Pi) / N_y, \quad (8)$$

$$D_\Pi^2 = \sum_{k=1}^{N_y} (S_k^\Pi)^2 / N_y - \bar{S}_\Pi^2. \quad (9)$$

Здесь символ S_k^Π означает значение решающей функции на k -том анализируемом отрезке заведомо относящихся к паузе данных.

Для сокращения объема вычислительных работ используются свойства собственных векторов и матриц:

1) для каждого из частотных интервалов вычисляются матрицы A_r и соответствующие наборы собственных векторов и чисел

$$\lambda_{kr} \bar{q}_{kr} = A_r \bar{q}_{kr}, \quad k = 1, 2, \dots, J,$$

$$J = 2 \left[\frac{N}{2R} \right] + 2, \quad r = 1, \dots, R;$$

2) на основе собственных значений матриц A_r формируется матрица AA

$$AA = \begin{pmatrix} \sqrt{L_1} Q_1^T \\ \sqrt{L_2} Q_2^T \\ \dots \\ \sqrt{L_R} Q_R^T \end{pmatrix}, \quad (10)$$

где $Q_r = (\bar{q}_{1r}, \dots, \bar{q}_{Jr})$ – матрица собственных векторов, $L_r = \text{diag}(\lambda_{1r}, \dots, \lambda_{Jr})$ – диагональная матрица собственных чисел матрицы A_r ;

3) для определения приближенных значений долей энергии отрезка сигнала вычисляется вектор

$$\vec{y} = AA\vec{x} = \begin{pmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \dots \\ \bar{y}_R \end{pmatrix} \quad (11)$$



и суммы квадратов компонент соответствующих подвекторов

$$\hat{P}_r = \sum_{k=1}^J (y_{kr})^2, \quad (12)$$

$$\bar{y}_r = \sqrt{L_r} Q_r^T \bar{x}, r = 1, \dots, R,$$

что является оценкой доли энергии сигнала в r -том частотном интервале отрезка речевого сигнала.

III. Для оценки обоснованности выбора h_α используется следующий этап диагностической проверки.

1) Вычисление оценки вероятности правильного обнаружения паузы (α_{non}).

Вычисление оценки вероятности правильного обнаружения паузы проводилось на участке сигнала, заведомо относящегося к паузе по формуле

$$\alpha_{\text{non}} = \frac{D_{\text{нен}}}{k2 - k1};$$

где $D_{\text{нен}}$ – количество значений решающей функции, не превышающих порог; $k1$ – номер отсчета начала паузы; $k2$ – номер отсчета конца паузы.

2) Вычисление оценки вероятности ложного обнаружения паузы ($\alpha_{\text{лож}}$).

Вычисление оценки вероятности ложного обнаружения паузы проводилось на участке сигнала, заведомо относящегося к речи по формуле

$$\alpha_{\text{лож}} = \frac{D_n}{n2 - n1};$$

где D_n – количество значений решающей функции превышающих порог; $n1$ – номер отсчета начала речи; $n2$ – номер отсчета конца речи.

3) Определение коэффициента сжатия за счет кодирования пауз по формуле

$$K_{\text{сжат}} = \frac{x}{m};$$

где x – размер исходного сигнала в отсчётах; m – длина кодированных пауз в отсчётах.

В качестве эмпирических данных были использованы отрезки речевых файлов. Длины анализируемых отрезков выбраны равными $N=20, 60, 200, 1000$.

Область определения трансформант Фурье дискретных сигналов $[0, \pi]$ разбивается на R одинаковых частотных интервалов, таких что $\nu_{2r} - \nu_{1r} = \Delta\nu = \text{const}$, причем такие R , что M является целым числом ($N=MR$).

В ходе экспериментов для всех значений N используется различное разбиение оси частот на R -частотных интервалов, а именно:

- при $N=20$ $R=2, 5, 10$;
- при $N=60$ $R=2, 6, 10, 15, 30$;
- при $N=200$ $R=2, 20, 25, 50, 100$;
- при $N=1000$ $R=2, 10, 50, 100$.

В табл. 1 представлены предварительно вычисленное значение порога при заданном уровне вероятности ложной тревоги, оценка вероятности правильного и ложного обнаружения пауз, а также коэффициент сжатия сигнала за счет кодирования пауз на участке сигнала в 100000 отсчетов при заданных N и R .

Для иллюстрации полученных результатов ниже приведены рисунки, на которых изображены границы паузы/речи и речи/паузы (на рисунках заранее на слух определяемая граница отображена вертикальной линией).

Таблица 1

Оценка вероятности правильного ($\alpha_{\text{пол}}$) и ложного ($\alpha_{\text{лож}}$) обнаружения пауз, коэффициент сжатия ($K_{\text{сжат}}$)

№ эксперимента	N	R	$\alpha_{\text{пол}}$	$\alpha_{\text{лож}}$	$K_{\text{сжат}}$
1	20	2	0,98781	0,0032	1,76
2	20	5	0,98781	0,0026	1,75
3	20	10	0,98320	0,0013	1,73
4	60	2	0,98261	0,0012	1,74
5	60	6	0,98801	0,0006	1,74
6	60	10	0,98441	0,0006	1,73
7	60	15	0,99101	0,0007	1,75
8	60	30	0,98381	0,0007	1,73
9	200	2	0,99001	0,0000	1,75
10	200	20	0,99601	0,0000	1,75
11	200	25	0,99201	0,0022	1,75
12	200	50	0,99801	0,0000	1,75
13	200	100	0,99801	0,0000	1,75
14	1000	2	0,99001	0,0000	1,72
15	1000	10	0,98001	0,0000	1,71
16	1000	50	0,98001	0,0000	1,71
17	1000	100	0,98001	0,0000	1,71

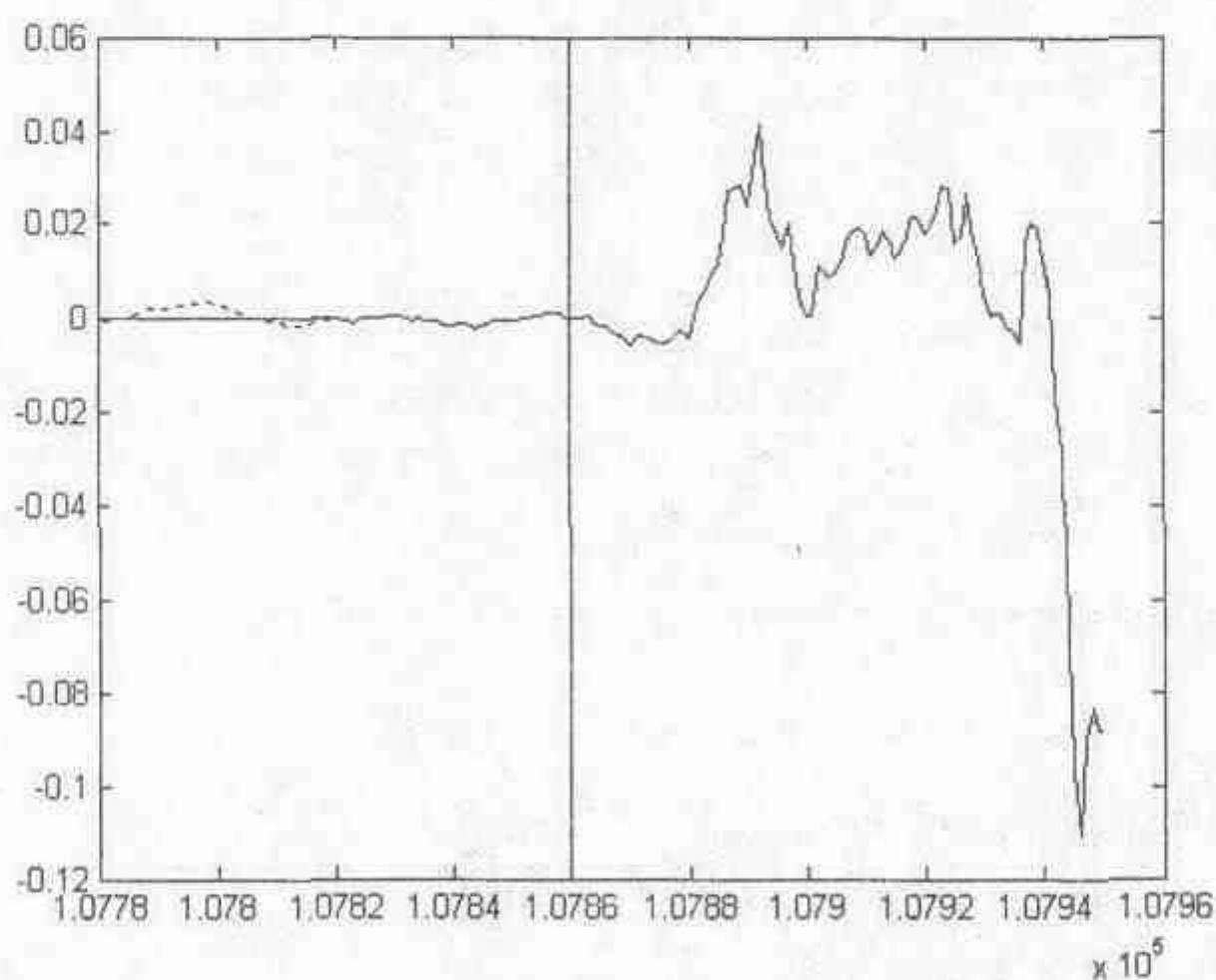


Рис. 1. Граница пауза/звук, определенная при использовании параметров $N=60$, $R=2$

При использовании значений параметров $N=60$, $R=2$ 40 отсчетов паузы определяются как речь, т.к. анализируемый отрезок достаточно велик и захватывает как речь, так и паузу перед речью. Звук при этом не искажается, но ухудшается степень сжатия речи.

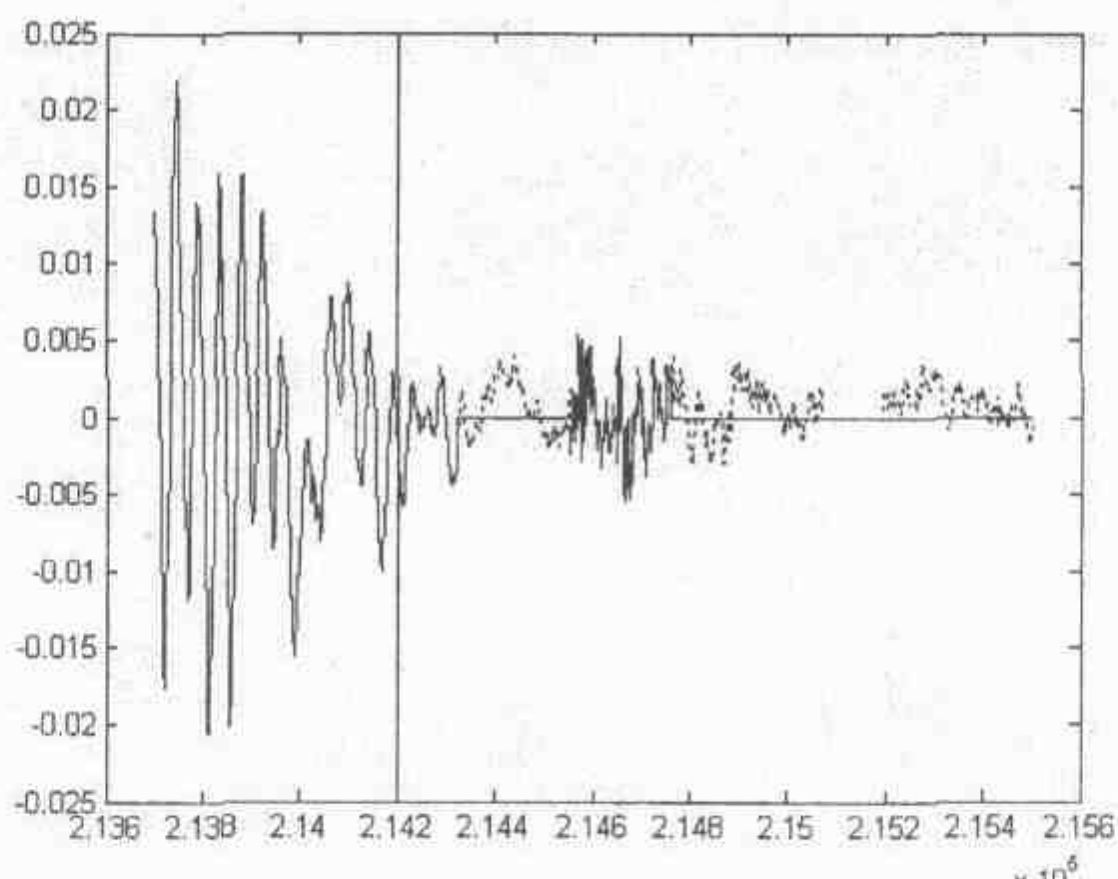


Рис. 2. Граница звук/пауза, определенная при использовании параметров $N=60$, $R=2$

При использовании значений параметров $N=60$, $R=2$ 180 отсчетов паузы определяются как речь. Звук при этом не искажается, но некоторые короткие участки паузы (например, отсчеты с 214600 по 214800) определяются как речь, что создает «треск» при воспроизведении и ухудшает степень сжатия речи.

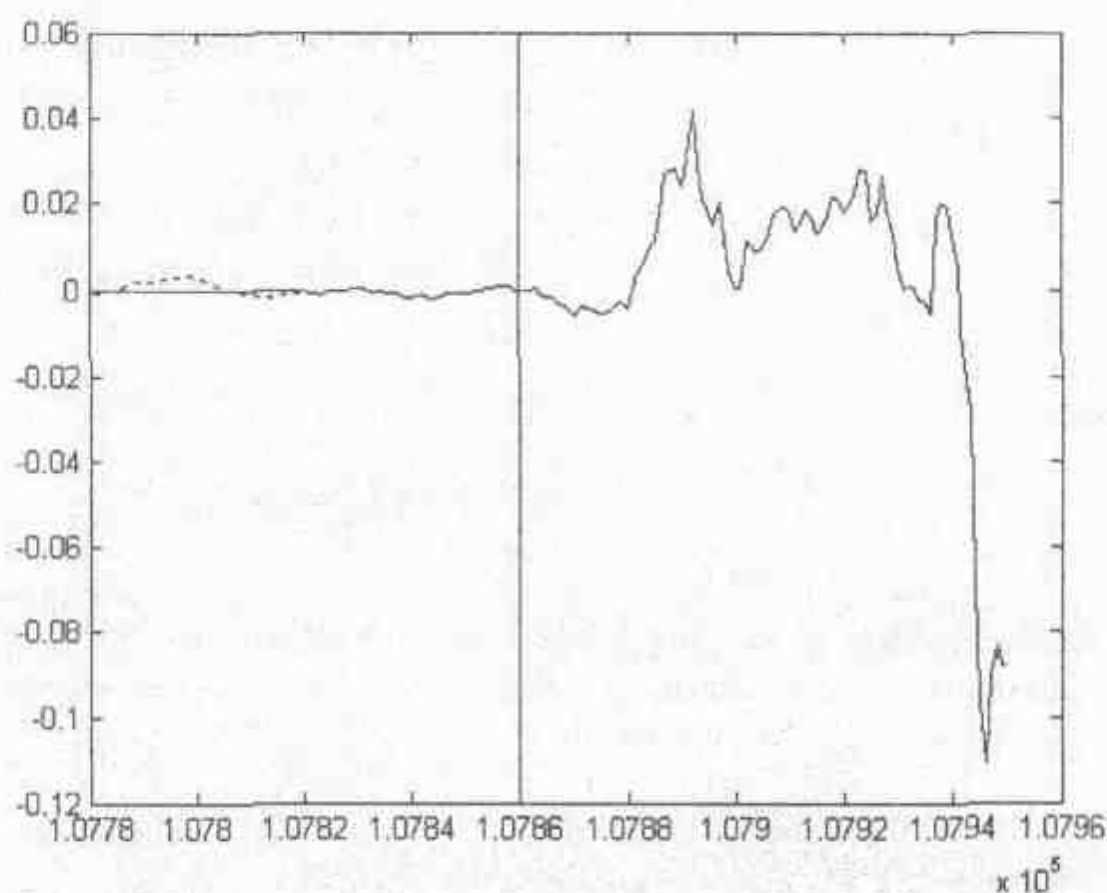


Рис. 3. Граница пауза/звук, определенная при использовании параметров $N=60$, $R=10$

При использовании значений параметров $N=60$, $R=10$ 40 отсчетов паузы определяются как речь, т.к. анализируемый отрезок достаточно велик и захватывает как речь, так и паузу перед речью. Звук при этом не искажается, но ухудшается степень сжатия речи.

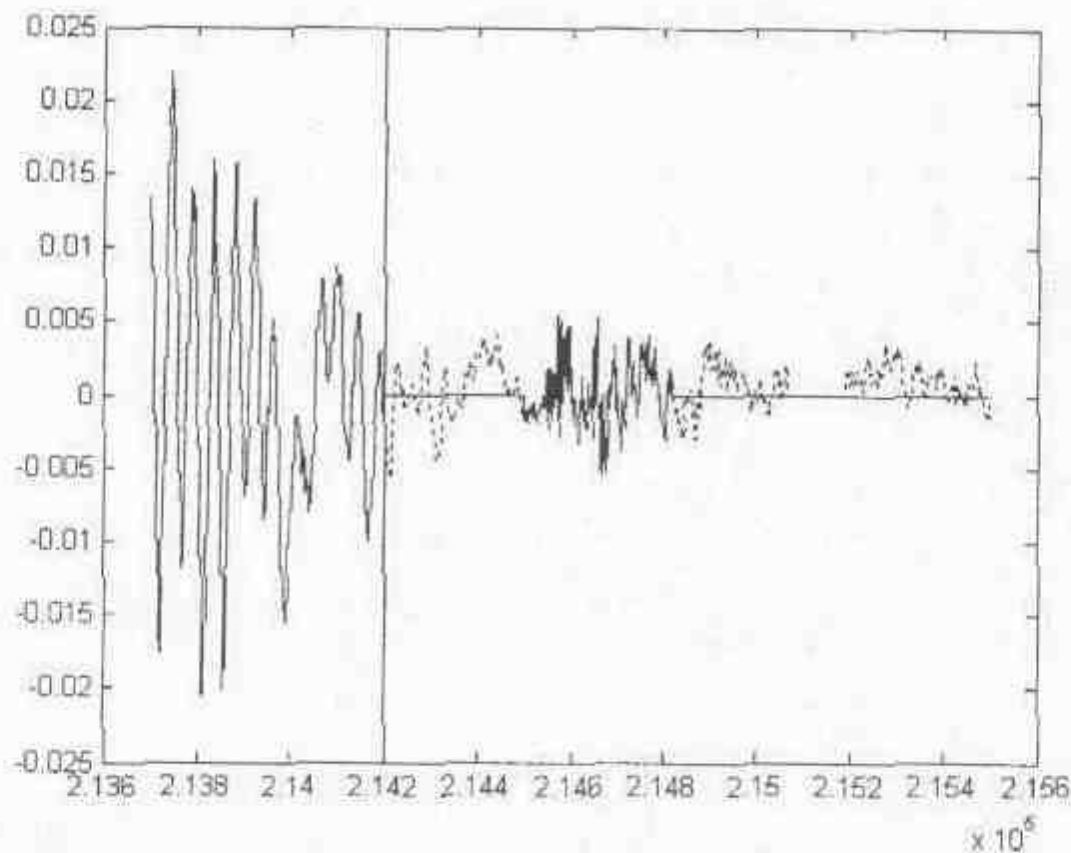


Рис. 4. Граница звук/пауза, определенная при использовании параметров $N=60$, $R=10$

При использовании значений параметров $N=60$, $R=10$ граница паузы определяется точно, но некоторые короткие участки паузы (например, отсчеты с 214600 по 214800) определяются как речь, что создает «треск» при воспроизведении и ухудшает степень сжатия речи.

ЗАКЛЮЧЕНИЕ

Предлагаемый алгоритм сжатия речевых данных за счет обнаружения и кодирования пауз на основе сравнения распределений энергии шума и смеси «сигнал+шум» в заданных частотных интервалах обладает высокой работоспособностью.

При всех использованных сочетаниях N и R вероятность правильного обнаружения пауз не менее чем 0,98, а ложного обнаружения пауз – не превосходит 0,005. Полученные при этом коэффициенты сжатия более 1,7 раза.

По результатам вычислительных экспериментов рекомендуется использовать длины анализируемых отрезков $N=60$ при количестве частотных интервалов $R=10$, т.к. при этом адекватно учитываются узость частотных интервалов, где сосредоточена энергия речевых сигналов и объем вычислительных работ.

Литература

1. Жилияков Е.Г. Уменьшение объема битового представления речевых данных на основе нового метода удаления пауз / Е.Г. Жилияков, С.П.Белов, Е.И. Прохоренко // Вопросы радиоэлектроники. Серия «Электронная вычислительная техника (ЭВТ)». – М., 2007. – Вып. 2. – С. 82-92.
2. Жилияков Е.Г. Вариационные методы частотного анализа звуковых сигналов / Е.Г. Жилияков, С.П.Белов, Е.И. Прохоренко // Труды учебных заведений связи. – СПб., 2006. – Вып. 174. – С. 163-172.

RESEARCH OF ALGORITHM OF FINDING OUT PAUSES IS IN VOCAL REPORTS

A.S. Belov

Belgorod state university, 308015, Belgorod, street of Victory, 85, Belov@bsu.edu.ru

The method of finding out pauses is offered in vocal reports, based on the account of distinctions in distributing of frequency power constituents of spectrum of signal of speech from the spectrum of signal in a pause, entrance vocal signal applied for cleaning from noises.

Key words: frequency presentations, method of finding out pauses in vocal reports, algorithm of calculation of the exact meaning of stake of energy of cutting-off of signal, which is concentrated in any eventual frequency interval, evaluation of probabilities of correct discovery of pauses and false discovery of pauses.