



# ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 621.391.037.372

## СЖАТИЕ РЕЧЕВЫХ ДАННЫХ НА ОСНОВЕ ПРОРЕЖИВАНИЯ РЕЗУЛЬТАТОВ ОПТИМАЛЬНОЙ ЛИНЕЙНОЙ ЧАСТОТНОЙ ФИЛЬТРАЦИИ

**Е.Г. ЖИЛЯКОВ**  
**А.В. БОЛДЫШЕВ**  
**Н.А. ЧЕКАНОВ**

*Белгородский  
государственный национальный  
исследовательский университет*

*e-mail: boldyshev@bsu.edu.ru*

В статье изложен подход к сжатию речевых данных на основе прореживания результатов оптимальной линейной фильтрации речевого сигнала. Приведены результаты вычислительных экспериментов по оценке эффективности предлагаемого подхода.

Ключевые слова: информационно-телекоммуникационные системы, сжатие речевых данных, оптимальная линейная фильтрация, заданная доля энергии.

### **Введение.**

Информационный обмен является важнейшим средством развития общественных процессов, включая производственные силы. Одной из наиболее удобных и естественных форм информационного обмена для человека являются речевые конструкции (речевые сообщения). Реализация информационного обмена речевыми сообщениями, включая их архивное хранение и передачу, осуществляется с помощью компьютерных технологий. При этом речевые сигналы хранятся и передаются в виде некоторых кодовых комбинаций, совокупность которых естественно называть речевыми данными. Совокупность бит, используемых для кодирования речевых данных, называется объемом битовых представлений.

Одной из основных проблем современных информационно-телекоммуникационных систем является ограниченность их ресурсов, необходимых для реализации информационного обмена, а именно пропускной способности для передачи данных и объемы памяти жестких носителей для хранения.

Поэтому не вызывает сомнения необходимость выбора такого способа кодирования, который обеспечивает минимум объемов битовых представлений хранимых и передаваемых данных при сохранении приемлемого, с точки зрения пользователя, качества воспроизведения исходных речевых сообщений.

### **Теоретические основы.**

В данной работе приводится описание алгоритма сжатия исходных речевых данных на основе прореживания результатов оптимальной линейной частотной фильтрации [1].

Пусть  $\vec{f} = (f_1, \dots, f_L)^T$  исходный речевой сигнал длительностью  $L$  отсчетов, где  $T$  – операция транспонирования.



Обработка речевого сигнала осуществляется по отрезкам, для этого исходный сигнал  $\vec{f}$  разбивается на отрезки равной длины:

$$\vec{x}_i = (f_{(i-1)*N+1}, \dots, f_{iN})^T, i=1, \dots, M \tag{1}$$

где  $N$  – длительность окна анализа;

$M=L/N$  – количество отрезков.

Частотный диапазон  $[0, \pi]$  разбивается на  $R$  частотных интервалов следующим образом:

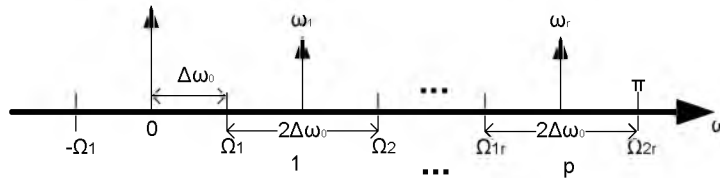


Рис.1. Пример разбиения частотной оси

где  $\pi=(2r+1)\Delta\omega_0$ ;  $\Omega_{1r}=(2r-1)\Delta\omega_0$  и  $\Omega_{2r}=\Omega_{1r}+2\Delta\omega_0$ ,  $r=1, \dots, R-1$  – нижняя и верхняя границы заданного частотного интервала;  $\Delta\omega_0=(\Omega_{2r}-\Omega_{1r})/2=2\pi/(N-1)$  – задает полуширину частотного интервала;  $\omega_r=(\Omega_{2r}+\Omega_{1r})/2$  – центральная частота  $r$ -го частотного интервала.

Частотный интервал, расположенный вблизи нуля (с границами  $[0, \Omega_{1r})$  и центральной частотой  $\omega=0$ ), будем называть нулевым.

При выборе количества частотных интервалов должно выполняться условие  $N/R \geq 4$ .

Для каждого частотного интервала рассчитывается субполосная матрица, причем для нулевого интервала она имеет вид:

$$A_0 = \{a_{ik}^0\} = \frac{\sin[\Delta\omega_0(i-k)]}{\pi(i-k)}, i, k=1, \dots, N, r=0; \tag{2}$$

для остальных частотных интервалов:

$$A_r = \{a_{ik}^r\} = \frac{\sin[\Omega_{2r}(i-k)] - \sin[\Omega_{1r}(i-k)]}{\pi(i-k)}, r=1, \dots, R-1, i, k=1, \dots, N \tag{3}$$

Субполосные матрицы являются симметричными и неотрицательно определенными, поэтому их можно представить в виде разложения по их собственным векторам и числам [1]:

$$A_r = Q_r L_r Q_r^T, r=0, 1, \dots, R-1 \tag{4}$$

где  $Q_r = \{\vec{q}_1^r, \vec{q}_2^r, \dots, \vec{q}_N^r\}$  – матрица собственных векторов;  $L_r = \text{diag}(\lambda_1^r, \dots, \lambda_N^r)$  – диагональная матрица собственных чисел.

$$\lambda_1^r > \lambda_2^r > \dots > \lambda_N^r > 0$$

Собственные числа количественно равны сосредоточенным в выбранных частотных интервалах долям энергий соответствующих собственных векторов [2]. Т.е. если собственное число равно единице, значит вся энергия соответствующего собственного вектора, сосредоточена в выбранном частотном интервале, иначе энергия этого вектора просачивается за пределы этого интервала.

В [2] показано, что величина собственных чисел, индексы которых превосходят значение

$$J = 2 \left[ \frac{N * 2 * \Delta\omega_0}{2\pi} \right] + 4 \tag{5}$$

пренебрежимо мала по сравнению с единицей (квадратные скобки означают целую часть от результата).

В ряде работ [3,4] описана процедура определения минимального количества частотных интервалов, в которых сосредоточена  $m$ -ая (подавляющая, порядка 0.92) доля



энергии отрезка речевого сигнала. Для последовательности изложения необходимо напомнить основные моменты этой процедуры: вычисление распределения энергии по частотным интервалам, упорядочивание их по убыванию, определение количества и порядковых номеров частотных интервалов, сумма которых составляет  $m$ -ую долю энергии от общей:

$$P_r = \vec{x}^T A_r \vec{x} \quad (6)$$

$$\sum_{k=1}^{U_m} P_{(k)} \geq m \|\vec{x}\|^2 = m \sum_{i=1}^N x_i^2 \quad (7)$$

где  $U_m$  – минимальное количество частотных интервалов (множество  $R_m$ ), в которых сосредоточена  $m$ -ая доля энергии;  $P_r$  – энергия в выбранном частотном интервале;  $P_{(k)}$  – порядковая статистика  $P_{(k+1)} \geq P_{(k)}, P_{(k)} \in \{P_1, \dots, P_R\}$ ;  $\|\vec{x}\|^2$  – энергия анализируемого отрезка сигнала.

Таким образом, имея сведения о номерах частотных интервалов, можно с помощью оптимальной линейной частотной фильтрации выделить частотные компоненты, соответствующие только этим интервалам [1]:

$$\vec{y}_r = A_r \vec{x}^T = \sum_{k=1}^J \lambda_k^r \alpha_k^r \vec{q}_k^r, r \in R_m \quad (8)$$

где  $\alpha_k^r = (\vec{x}, \vec{q}_k^r)$  – скалярное произведение анализируемого отрезка сигнала и собственного вектора.

Используя тригонометрические преобразования ( $\sin(\alpha) - \sin(\beta)$ ), можно преобразовать выражение (3) к следующему виду:

$$\begin{aligned} A_r &= 2 \frac{\sin\left[\frac{\Omega_{2r} - \Omega_{1r}}{2}(i-k)\right]}{\pi(i-k)} \cos\left[\frac{\Omega_{2r} + \Omega_{1r}}{2}(i-k)\right] = \\ &= 2 \frac{\sin[\Delta\omega_0(i-k)]}{\pi(i-k)} \cos[\omega_r(i-k)] = 2\alpha_{ik}^0 \cos[\omega_r(i-k)] \end{aligned} \quad (9)$$

В [5] показана справедливость следующих аппроксимаций собственных векторов субполосных матриц  $A_r$  для любого частотного интервала:

$$q_{k,2i-1}^r = q_{k,i}^0 \cos[\omega_r(i-1)] \text{ и } q_{k,2i}^r = q_{k,i}^0 \sin[\omega_r(i-1)], r \in R_m \quad (10)$$

Выражение (8) с учетом (10) можно преобразовать к виду:

$$\vec{y}_r = G_c \sum_{k=1}^{J/2} \lambda_{2k-1}^r \alpha_{2k-1}^r \vec{q}_k^0 + G_s \sum_{k=1}^{J/2} \lambda_{2k}^r \alpha_{2k}^r \vec{q}_k^0 = G_c \vec{z}_r^c + G_s \vec{z}_r^s, r \in R_m \quad (11)$$

где  $G_c = \text{diag}(\cos(\omega_r(0)), \cos(\omega_r(1)), \dots, \cos(\omega_r(N-1)))$ ;

$G_s = \text{diag}(\sin(\omega_r(0)), \sin(\omega_r(1)), \dots, \sin(\omega_r(N-1)))$ ;

Выражение (11) представляет собой сумму произведений синусов и косинусов на соответствующие огибающие  $\vec{z}_r^c$  и  $\vec{z}_r^s$ . Выделить эти огибающие можно следующим образом:

$$\begin{aligned} \vec{z}_r^c &= Q_0 Q_0^T G_c \vec{y}_r, r \in R_m \\ \vec{z}_r^s &= Q_0 Q_0^T G_s \vec{y}_r, r \in R_m \end{aligned} \quad (12)$$

Уменьшение количества исходных речевых данных осуществляется за счет прореживания компонент выделенных огибающих  $\vec{z}_r^c$  и  $\vec{z}_r^s$  с шагом:

$$D = 2p + 1 = 2(R-1) + 1 = 2R - 1 \quad (13)$$

Таким образом, количество значений, которое необходимо хранить равно:

$$M = U_m * 2([N/D] + 1) \quad (14)$$

где  $[]$  – целая часть от результата.



Для того, чтобы восстановить сигнал по его прореженным значениям необходимо сперва сформировать вектор, который будет состоять из прореженных значений и нулей на местах интерполируемых значений.

$$\vec{w}_{r,i}^c = \begin{cases} z_{r,i}^c, i = 1, D, 2D, \dots, ID \\ 0, i \neq 1, D, 2D, \dots, ID \end{cases}, l = [N/D] \tag{15}$$

$$\vec{w}_{r,i}^s = \begin{cases} z_{r,i}^s, i = 1, D, 2D, \dots, ID \\ 0, i \neq 1, D, 2D, \dots, ID \end{cases}, l = [N/D]$$

где  $[\ ]$  – целая часть от результата.

Фактически при прореживании (12) прореживается именно компоненты  $w_r$ , которая отражает сосредоточенность энергии в выбранном частотном интервале, поэтому необходимо получить частотную компоненту, энергия которой не будет просачиваться за пределы частотного интервала. Данное преобразование осуществляется аналогично (12).

$$\begin{aligned} \tilde{z}_r^c &= Q_0 Q_0^T \vec{w}_r^c, r \in R_m \\ \tilde{z}_r^s &= Q_0 Q_0^T \vec{w}_r^s, r \in R_m \end{aligned} \tag{16}$$

Далее необходимо получить частотные компоненты, которые соответствуют частотным интервалам  $r=1, \dots, R-1$ . Согласно (10) получаем:

$$\tilde{y}_r = G_c \tilde{z}_r^c + G_s \tilde{z}_r^s, r \in R_m \tag{17}$$

Восстановление исходного отрезка речевого сигнала осуществляется путем суммирования всех частотных компонент:

$$\tilde{x} = \sum_{r=1}^{U_m} \tilde{y}_r \tag{18}$$

**Вычислительные эксперименты.**

В рамках исследования предложенного подхода к сжатию речевых данных были проведены вычислительные эксперименты. В качестве исходных данных были выбраны отрезки речевых сигналов, соответствующие различным звукам русской речи. Звуки были выделены «на слух» из различных произнесенных фраз. Длительность окна анализа  $N$  была выбрана 129 отсчетов, количество частотных интервалов  $R=32$ , что соответствует требованию  $N/R \geq 4$ .

В табл. 1 приведено количество отрезков речевого сигнала, соответствующих различным звукам, которые подвергались обработке.

Таблица 1

**Количество анализируемых звуков русской речи**

	Количество звуков		Количество звуков		Количество звуков		Количество звуков
А	71	З	21	П	23	Ч	15
Б	26	И	49	Р	30	Ш	15
В	17	Й	18	С	31	Щ	12
Г	16	К	29	Т	34	Ы	19
Д	22	Л	31	У	18	Э	18
Е	19	М	27	Ф	16	Ю	11
Ё	11	Н	33	Х	16	Я	12
Ж	15	О	26	Ц	20		

Для каждого звука был вычислен коэффициент сжатия:



$$K = \frac{N}{U_m * 2 * M}$$

где М – количество отсчетов, оставшихся после прореживания.  
Далее вычислялся средний коэффициент сжатия:

$$K_{\text{сред}} = \sum_{i=1}^{Zv} K_i / Zv$$

где Zv – общее количество каждого анализируемого звука.

В табл. 2 приведены результаты вычисления среднего коэффициента сжатия.

Таблица 2

**Коэффициенты сжатия для каждого звука речи.  
K<sub>сред</sub> – средний коэффициент сжатия. N=129, R=32**

М Звук	0,90	0,92	0,94	0,96	m Звук	0,90	0,92	0,94	0,96
А	3,63	3,34	2,98	2,57	П	4,70	4,31	3,71	3,10
Б	4,97	4,67	4,22	3,42	Р	4,52	4,08	3,63	3,05
В	4,31	3,92	3,48	2,96	С	4,20	3,05	3,33	2,80
Г	4,33	3,91	3,49	3,02	Т	3,52	3,27	2,87	2,42
Д	4,66	4,05	3,73	3,11	У	4,55	4,28	3,75	3,11
Е	4,45	4,08	3,69	3,19	Ф	4	3,65	3,20	2,65
Ё	4,44	4,06	3,62	2,99	Х	3,96	3,68	3,23	2,71
Ж	4,05	3,70	3,27	2,73	Ц	4,02	3,66	3,24	2,70
З	4,36	4,04	3,63	3,14	Ч	3,96	3,62	3,20	2,67
И	5,74	5,38	4,87	4,19	Ш	3,83	3,50	3,10	2,59
Й	5,55	5,01	4,49	3,88	Щ	3,86	3,53	3,12	2,60
К	4,14	3,81	3,39	2,84	Ы	4,90	4,48	3,77	3,13
Л	5,35	4,90	4,36	3,64	Э	4,28	3,94	3,48	2,93
М	6,66	5,94	5,22	4,21	Ю	4,41	4,11	3,61	2,98
Н	6,76	6,13	5,35	4,35	Я	4,31	3,97	3,49	2,93
О	5,02	4,58	3,99	3,40					

Средний коэффициент сжатия по всем звукам русской речи составляет порядка 4-6 раз, при m=0.92.

На рисунке 2 в качестве пример приведены графики, соответствующие исходным и восстановленным звукам «а» и «ж».

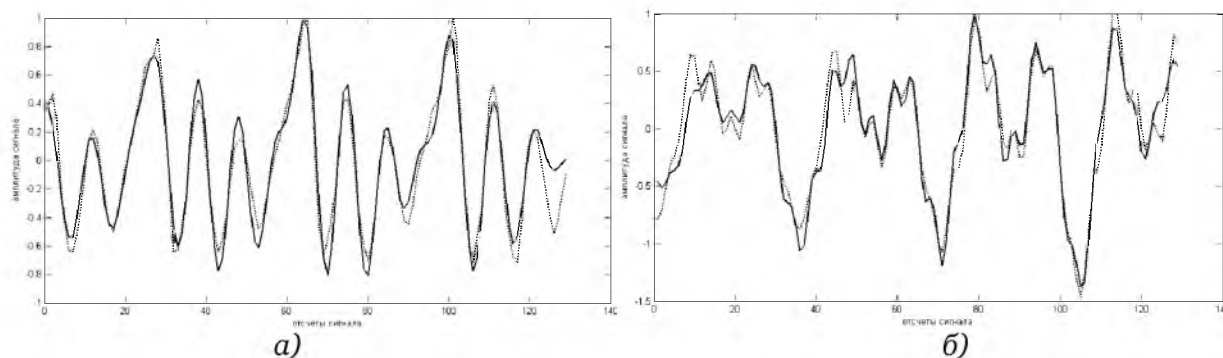


Рис. 2. Исходный сигнал (пунктирная линия) и восстановленный (сплошная линия):  
а) звук «а»; б) звук «ж» m=0.92



Как видно из приведенных выше рисунков, что форма восстановленного сигнала практически совпадает с формой исходного, исключение составляют лишь «края» сигнала. Практически полное совпадение формы восстановленного и исходного сигнала свидетельствует о возможности получения высокого качества воспроизведения восстановленного речевого сообщения.

### **Выводы.**

Проведенные вычислительные эксперименты показали, что предлагаемый подход к сжатию может обеспечить сокращение исходного размера речевого сообщения в 5 раз при сохранении достаточно высокого качества воспроизведения.

*Работа выполнена в рамках программы РНПВШ ГКН<sup>о</sup> 8.2251.2011*

### **Литература**

1. Гантмахер Ф.Р. Теория матриц / Ф.Р. Гантмахер. – М.: Физматлит, 2004. – 560с.
2. Жилияков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. / Е.Г. Жилияков. – Белгород: Изд-во, 2007. – БелГУ, 2007.– 160. Гантмахер, Ф.Р. Теория матриц / Ф.Р. Гантмахер. – М.: Физматлит, 2004. – 560с.
3. Болдышев А.В. О различиях распределения энергии звуков русской речи и шума / А.В. Болдышев, А.А. Фирсова // материалы 12-ой Международной конференции и выставке «ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ и ЕЁ ПРИМЕНЕНИЕ – DSPA'2010» 31 марта – 02 апреля 2010 года, г. Москва.
4. Прохоренко Е.И. Метод сжатия речевых данных на основе составной субполосной матрицы / Е.И. Прохоренко, А.В. Болдышев, А.В. Эсауленко // Журнал «Вопросы Радиоэлектроники», серия электроника и вычислительная техника (ЭВТ). Выпуск №1 Москва 2011. с. 60-72.
5. Жилияков Е.Г. О вычислении собственных функций субполосного ядра / Е.Г. Жилияков, С.В. Туяков // Вопросы радиоэлектроники. Сер. Электронная вычислительная техника (ЭВТ). – 2011. – Вып. 1. – С. 25-34.

## **SPEECH COMPRESSION DATA BASED ON THE THINNING RESULTS OF OPTIMAL LINEAR FREQUENCY FILTRATION**

**E.G. ZHILYAKOV  
A.V. BOLDYSHEV  
N.A. CHEKANOV**

*Belgorod National  
Research University*

*e-mail: boldyshev@bsu.edu.ru*

The article describes an approach to compression of speech data based on the thinning results of optimal linear filtration of the speech signal. The results of computational experiments to evaluate the effectiveness of the proposed approach.

Key words: Information and communication systems, speech data compression, optimal linear filtrations, given part of the energy.