УДК 621.391.037.372

МЕТОД СЖАТИЯ РЕЧЕВЫХ ДАННЫХ НА ОСНОВЕ ОПТИМАЛЬНОГО СУБПОЛОСНОГО ПРЕОБРАЗОВАНИЯ ПО СОСТАВНЫМ ЧАСТОТНЫМ ИНТЕРВАЛАМ

А.В. БОЛДЫШЕВ

Белгородский государственный университет

e-mail: boldyshev@bsu.edu.ru

В статье изложен подход к сжатию речевых данных на основе квантования по уровню оптимальных субполосных преобразований отрезков речевых сигналов по составным частотным интервалам. Приведены результаты вычислительных экспериментов по оценке эффективности разработанного метода.

Ключевые слова: информационно-телекоммуникационные технологии, сжатие речевых данных, оптимальное субполосное преобразование, составные частотные интервалы, 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. На сегодняшний день актуальной считается проблема 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. Решение этой проблемы позволяет минимизировать затраты объемов компьютерной памяти для хранения данных и времени их передачи в информационно-телекоммуникационных системах (ИТС).

В качестве примера можно указать следующие направления и области использования ИТС, для которых эта проблема имеет существенное значение:

– корпоративные информационно – телекоммуникационные системы, в которых используются средства аудио и видео конференцсвязи;

– системы постоянного мониторинга речевого и визуального обмена (аэропорты, видеонаблюдение, вокзалы и т.п.);

– хранение и передача речевых данных средствами Интернет (IP-телефония, голосовая почта, системы экспресс сообщений);

– информационно – телекоммуникационные системы удаленного взаимодействия, в том числе системы дистанционного образования;

– мультисервисные сети связи и сети радиодоступа.

Таким образом, проблема уменьшения объемов битовых представлений речевых данных (сжатия) является актуальной, а её решение позволит существенно повысить эффективность использования средств ИТС при реализации современного информационного обмена на основе речевых сообщений.

Теоретические основы

Одной из особенностей звуков русской речи является сосредоточенность энергии в достаточно узких частотных диапазонах, суммарная ширина которых гораздо меньше частоты дискретизации [1,2]. Эта особенность может быть использована в раз-

личных направлениях области обработки речевых сообщений: сжатие речевых данных, обнаружение и кодирование пауз, распознавание речи, повышения качества звучания речевых сообщений. В [3,4] приведены результаты исследований по оценке частотной концентрации звуков русской речи, т.е. оценке минимального количества частотных интервалов, в которых сосредоточена заданная доля энергии. Результаты проведенных исследований показали, что для большинства звуков русской речи величина частотной концентрации составляет порядка 0.35 и только для шумоподобных звуков – порядка 0.55-0.60. Полученные сведения о количестве и расположении частотных интервалов, в которых сосредоточена заданная доля энергии, можно осуществить сжатие речевых данных за счет хранения только составляющих речевого сигнала, соответствующих этим частотным интервалам.

Одним из способов получения составляющих речевого сигнала, соответствующих выбранным частотным интервалам является субполосное преобразование. В настоящее время наибольшее распространение получил метод субполосного преобразования на основе банка КИХ-фильтров, однако, этот метод обладает рядом недостатков, которые приводят к увеличению погрешностей восстановления данных [5].

В ряде публикаций [5,6] описывается метод субполосного преобразования, оптимальный с точки зрения минимума среднеквадратической погрешности аппроксимации трансформант Фурье исходного отрезка речевого сигнала в заданном частотном интервале, также в них показаны преимущества этого метода перед современными аналогами. В основе метода лежит математический аппарат с использованием субполосной матрицы вида:

$$A_r = \{a_{ik}^r\} = \{\sin(v_r(i-k)) - \sin(v_{r-1}(i-k))\} / \pi(i-k), i,k = 1,..,N \ , \qquad (1)$$

где v_r и v_{r-1} верхняя и нижняя границы частотного интервала.

Эта матрица является симметричной и неотрицательно определённой, поэтому она обладает полной системой ортонормальных собственных векторов, соответствующих неотрицательным собственным числам [7].

Этот математический аппарат можно использовать для получения компонент исходного речевого сигнала, соответствующих выбранным частотным интервалам. Для этого необходимо сформировать специальную составную матрицу, которая вычисляется как сумма субполосных матриц, соответствующих выбранных частотным интервалам, составляющих заданную долю энергии $\boldsymbol{m}$:

$$A_\Sigma = \sum_{i=1}^{l_{NR}^m} A_i \ , \qquad (2)$$

где l_{NR}^m – минимальное количество частотных интервалов, в которых сосредоточена заданная доля энергии отрезка речевого сигнала;

$\boldsymbol{t}$ – обозначает один из анализируемых речевых отрезков, порождаемых звуком русской речи; $\boldsymbol{R}$ – количество частотных интервалов, на которые разбивается частотный диапазон; $\boldsymbol{N}$ – длительность анализируемого отрезка; $\boldsymbol{m}$ – доля общей энергии, задаваемая для определения минимального количества частотных интервалов, в которых она сосредоточена[3]; A_i – субполосные матрицы, соответствующие тем частотным интервалам, которые составляют заданную долю энергии m.

Составная матрица обладает полным системой ортонормальных собственных векторов (3), соответствующих неотрицательным собственным числам (4):

$$Q_\Sigma = \{\vec{q}_{\Sigma 1}, \vec{q}_{\Sigma 2}, .., \vec{q}_{\Sigma N}\}, \qquad (3)$$

$$L_\Sigma = diag(\lambda_{\Sigma 1}, .., \lambda_{\Sigma N}). \qquad (4)$$

Необходимо отметить, что собственные числа количественно равны сосредоточенным в выбранных частотных интервалах долям энергий соответствующих собственных векторов и удовлетворяют условию:

$$0 \le \lambda_{\Sigma k} \le 1, k = 1,..., N. \qquad (5)$$

Для того, чтобы получить субполосный вектор, который будет отражать частотные свойства исходного отрезка речевых данных можно воспользоваться следующим выражением:

$$\vec{y}_\Sigma = \sqrt{L_\Sigma} Q_\Sigma^T \vec{x}\,, \tag{6}$$

где $\vec{x} = (x_1,...,x_N)^T$ – анализируемый отрезок речевых данных; Q_Σ^T – матрица собственных векторов; $\sqrt{L_\Sigma}$ – корень из диагонального элемента, соответствующего определенному собственному вектору.

Энергию отрезка речевого сигнала сосредоточенную в выбранных частотных интервалах можно определить как [5]:

$$P_\Sigma = \sum_{i=1}^{N} \lambda_{\Sigma i} y_{\Sigma i}^2\,. \tag{7}$$

С точки зрения сжатия речевых данных, можно поставить задачу нахождения минимального количества собственных значений составной матрицы, при оставлении которых будет достигаться максимальная степень сжатия. Сжатие исходных речевых данных будет осуществляться за счет хранения вектора значений размерностью равной минимальному количеству ненулевых собственных значений. При этом важным условием является минимизация погрешности восстановления исходного отрезка речевых данных, т.е. обеспечение высокого качества воспроизведения исходного речевого сообщения.

Представим выражение (7) в виде двух слагаемых:

$$P_\Sigma = \sum_{i=1}^{J_\Sigma} y_{\Sigma i}^2 \lambda_i + \sum_{i=J_{\Sigma+1}}^{N} y_{\Sigma i}^2 \lambda_i\,, \tag{8}$$

где $\sum_{i=1}^{J_\Sigma} y_{\Sigma i}^2 \lambda_i$ – первое слагаемое, в котором λ_i – собственные значения суммарной матрицы, величина, которых достаточно большая, $\sum_{i=J_{\Sigma+1}}^{N} y_{\Sigma i}^2 \lambda_i$ – второе слагаемое, в котором λ_i – собственные значения суммарной матрицы, величина, которых достаточно мала (близка к 0).

Доля этой энергии, которую составляет второе слагаемое, настолько мала, что предполагается ей можно пренебречь без получения существенных искажений. Таким образом, для оценки минимального количества собственных значений J_Σ, необходимых для восстановления исходного отрезка речевого сигнала без существенных потерь, можно использовать следующее выражение:

$$\sum_{i=1}^{J_\Sigma} \lambda_i \Big/ \sum_{i=1}^{N} \lambda_i \geq c\,, \tag{9}$$

где c – некий порог, который показывает, какую долю составляют собственные значения, величина которых близка к 0.

В качестве примера в таблице 1 приведено минимальное количество собственных значений J_Σ для звука русской речи «Б». При проведения экспериментальных исследований были выбраны следующие параметры: порог $c = 0.89 \div 0.995$, длительность отрезка речевых данных N=160, количество частотных интервалов R=16, заданная доля энергии отрезка речевого сигнала m=0.86÷0.98 в скобках указан параметр l_{NR}^{m}).

Как видно из приведенных результатов вычислительных экспериментов, минимальное количество собственных значений составной матрицы для данного звука в среднем составляет порядка 40, что позволяет говорить о возможность четырехкратного сокращения объема памяти, требуемого для хранения сведений о данном звука.

Ниже на рис. 1 приведена полученная степень сжатия для всех звуков русской речи. Степень сжатия определялась следующим образом [8]:

$$K = N / J_{\Sigma}. \qquad (10)$$

Таблица 1

Звук «Б», N=160, R=16

Порог c / $m\,(l_{NR}^{m})$	0.86-0.92 (2)	0.94 (3)	0.96-0.98 (4)
0.89	18	27	36
0.9-0.91	19	28	37
0.92	19	29	37
0.93-0.94	19	29	38
0.95- 0.96	20	30	39
0.97	20	31	39
0.98	21	32	40
0.99	21	33	41
0.995	22	34	41

С учетом выбора порога c, выражения для получения вектора субполосного преобразования примет вид:

$$\vec{y}_{\Sigma} = \sqrt{L_{\Sigma}}\,\widetilde{Q}_{\Sigma}^{T}\vec{x}, \qquad (11)$$

где $\vec{x} = (x_1,\ldots,x_N)^T$ – анализируемый отрезок речевых данных, $\widetilde{Q}_{\Sigma}^{T}$ – матрица собственных векторов, количество которых соответствует J_{Σ}

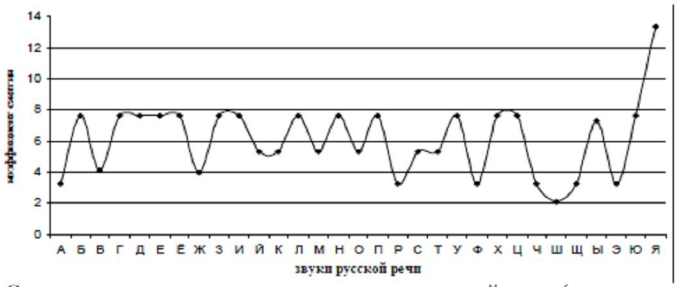


Рис. 1. Степень сжатия для различных звуков русской речи (c=0.92, m=0.92)

Для увеличения степени сжатия можно подвергнуть полученные вектора (11) квантованию по уровню. Ниже в таблице 2 приведены результаты экспериментальных исследований по оценке степени сжатия при использовании процедуры квантования по уровню для звука «Б». В качестве примера в таблице 2 приведены результаты для следующих параметров: m=0.86÷0.98, c=0.92, количество разрядов квантования n=1÷5. Коэффициент сжатия определялся следующим образом:

$$K_{сж} = V_{исх} / V_{сж}, \qquad (12)$$

где $V_{исх}$ – объем отрезка исходного сигнала, соответствующего определенному звуку, который определяется количеством бит, требуемых для хранения отсчетов исходной последовательности на жестком носителе;

$V_{сж}$ – объем сигнала, соответствующего определенному звуку, полученного в результате преобразования, определяемый количеством бит, которые должны быть выделены в памяти ЭВМ для хранения квантованных значений $\vec{y}_{\Sigma}$, так и служебной информации, в которую включаются данные о параметрах квантования и сведения о номерах частотных интервалов содержащих заданную долю энергии m.

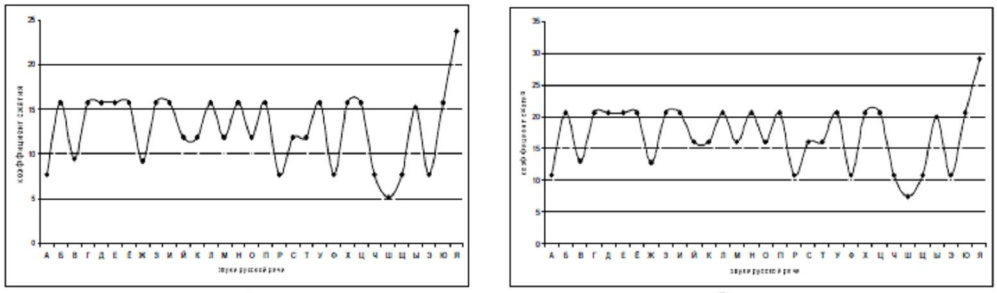
Таблица 2

Звук «Б», c=0.92

m	**0.86**	**0.88**	**0.9**	**0.92**	**0.94**	**0.96**	**0.98**
n=5	9.27	9.27	9.27	9.27	6.46	5.20	5.20
n=4	10.76	10.76	10.76	10.76	7.57	6.12	6.12
n=3	12.80	12.80	12.80	12.80	9.14	7.44	7.44
n=2	15.80	15.80	15.80	15.80	11.53	9.48	9.48
n=1	20.64	20.64	20.64	20.64	15.61	13.06	13.06

Как видно из приведенной таблицы, использование квантования по уровню результатов субполосного преобразования позволяет значительно увеличить коэффициент сжатия исходных речевых данных.

Ниже на рис. 2 приведены результаты вычислительных экспериментов для всех звуков русской речи, при c=0.92, m=0.92, n=1,2.



а) б)

Рис. 2. Степень сжатия для различных звуков русской речи (c=0.92, m=0.92), а) n=2, б) n=1

Приведенные в табл. 2 и на рис. 2 результаты показывают, что предлагаемый подход к сжатию речевых данных позволяет добиться высоких показателей степени сжатия.

Таблица 3

Результаты сжатия различных речевых сигналов

Исходный речевой сигнал	**Коэффициент сжатия**				
Отрывок новостей	**Количество разрядов квантования**				
	1	2	3	4	5
	30,04	21,63	16,9	13,87	11,76
	Оценка качества воспроизведения				
	4,0	4,1	4,2	4,2	4,3
Фраза №1 диктор мужчина (Системы синтеза речи, традиционно классифицируются по способу генерации речевых сигналов)	**Количество разрядов квантования**				
	1	2	3	4	5
	18	13,15	10,37	8,55	7,28
	Оценка качества воспроизведения				
	4	4,1	4,2	4,2	4,3
Фраза №1 диктор женщина	**Количество разрядов квантования**				
	1	2	3	4	**5**
	18,18	14,15	11,09	8,95	7,52
	Оценка качества воспроизведения				
	4	4,1	4,1	4,2	4,3
Отрывок из диалога двух людей	**Количество разрядов квантования**				
	1	2	3	4	5
	20,5	14,35	11,12	9,07	7,66
	Оценка качества воспроизведения				
	4,1	4,1	4,2	4,2	4,3

Еще одним немаловажным критерием оценки методов сжатия речевых данных является оценка качества воспроизведения подвергнутых процедуре сжатия записей. Для оценки качества воспроизведения была использована шкала объективной оценки MOS [9,10]. Результаты оценки некоторых звукозаписей, подвергнутых сжатию, приведены в таблице 3.

Выводы. Проведенные вычислительные эксперименты показали высокую эффективность предлагаемого подхода к сжатию речевых данных. Предлагаемый метод позволяет сократить исходный объем речевых данных до 20-30 раз при сохранении достаточно высокого качества воспроизведения.

Работа выполнена в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы ГК № 14.740.11.0494 от 01 октября 2010.

Литература

1. Жиляков, Е.Г. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений: моногр. / Е.Г. Жиляков, С.П.Белов, Е.И. Прохоренко // Белгород, 2007. – 136 с.
2. Шелухин, О.И. Цифровая обработка и передача речи / О.И. Шелухин, Н.Ф. Лукьянцев; под ред. О.И. Шелухина // М.: Радио и связь, 2000. – 456 с.: ил.
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. Жиляков, Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. / Е.Г. Жиляков. – Белгород: Изд-во, 2007. – БелГУ, 2007.– 160.
6. Жиляков, Е.Г. Вариационные методы частотного анализа звуковых сигналов / Е.Г. Жиляков, С.П. Белов, Е.И. Прохоренко // Труды учебных заведений связи. – СПб, 2006. – № 174. – С. 163-170.
7. Гантмахер, Ф.Р. Теория матриц / Ф.Р. Гантмахер. – М.: Физматлит, 2004. – 560 с.
8. Сизиков, В.С. Математические методы обработки результатов измерений: учебник для вузов / В.С. Сизиков. – СПб.: Политехника, 2001.
9. Recomendation P.800. Methods for subjective determination transmission quality [Электронный ресурс] // http://www.itu.int: Международный союз электросвязи
10. Тропченко А.Ю., Тропченко А.А. Методы сжатия изображений, аудиосигналов и видео [текст]: Учебное пособие – СПб: СПбГУ ИТМО, 2009. – 108 с.

COMPRESSION OF SPEECH DATA BASED ON THE OPTIMAL SUBBAND TRANSFORMATION OF COMPOSITE FREQUENCY INTERVALS

A.V. BOLDYSHEV

Belgorod State University

e-mail: boldyshev@bsu.edu.ru

The article describes approach to compression of the speech data based on the quantization level of optimal subband transformations segments of speech signals in a composite frequency intervals. Results of computational experiments to evaluate the effectiveness of the method.

Key words: Information and communication technology, speech data compression, optimal subband transformation, compound frequency intervals, given part of the energy.