



УДК 621.391.037.372

ОБ ИЗБИРАТЕЛЬНОМ ВОЗДЕЙСТВИИ НА ЧАСТОТНЫЕ КОМПОНЕНТЫ РЕЧЕВЫХ СИГНАЛОВ В ЗАДАЧЕ СЖАТИЯ

А. В. БОЛДЫШЕВ*Белгородский
государственный
университет**e-mail:
boldyshev@bsu.edu.ru*

В работе изложен новый подход к избирательному воздействию на частотные компоненты речевых сигналов, основанный на использовании субполосных преобразований по составным частотным интервалам. Приведены результаты вычислительных экспериментов по оценке эффективности разработанного метода.

Ключевые слова: информационно-телекоммуникационные технологии, сжатие речевых данных, субполосное преобразование, частотный интервал, заданная доля энергии.

Введение

Одной из основных тенденций развития информационно-телекоммуникационных технологий является обеспечение естественных для человека форм информационного обмена (речь, визуальные отображения действительности). 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. 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. Решение этой проблемы позволяет минимизировать затраты объемов компьютерной памяти для хранения данных и времени их передачи в информационно-телекоммуникационных системах (ИТС).

Можно указать достаточно много направлений и областей, для которых решение проблемы минимизации объемов битовых представлений речевых данных имеет существенное значение:

- корпоративные информационно-телекоммуникационные системы, в которых используются средства аудио- и видеоконференцсвязи;
- системы постоянного мониторинга речевого и визуального обмена (аэропорты, видеонаблюдение, вокзалы и т.п.);
- хранение и передача речевых данных средствами Интернет;
- информационно-телекоммуникационные системы удаленного взаимодействия, в том числе системы дистанционного образования.

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.

Актуальность этой проблемы также можно проиллюстрировать наличием ряда исследовательских компаний и институтов, которые занимаются обширными и многоаспектными исследованиями для решения различных задач в области обработки речевых данных, например, консорциум «Российские речевые технологии»,



ООО «Сакрамент» (Беларусь), Институт проблем передачи информации РАН; Институт Фраунгофера и фирма Thomson (формат MP3 pro).

Для сжатия речевых данных разработаны различные процедуры обработки, основой которых служат необратимые преобразования исходных данных, например, за счет более грубого квантования по уровню. Существующие методы сжатия звуковых данных с использованием грубого квантования по уровню основываются на психоакустической модели (например, формат MP3 pro) [1], что приводит к необходимости применения так называемых субполосных преобразований отрезков (векторов) отсчетов речевых сигналов, позволяющих получить другие векторы, подвекторы которых отражают частотные свойства исходного вектора в выбранных диапазонах оси частот. Компоненты этих подвекторов подвергаются квантованию по уровню. Для субполосного преобразования обычно используется процедура прореживания выходных последовательностей КИХ-фильтров, настроенных на соответствующие участки оси частот. Однако, такая процедура субполосного преобразования не является оптимальной в смысле достижения минимальных погрешностей аппроксимации в выбранных частотных диапазонах трансформант Фурье исходных векторов. Это приводит к увеличению погрешностей восстановления данных по квантованным значениям и, как следствие, к ухудшению качества воспроизводимой речи.

Таким образом, центральной, с точки зрения хранения и передачи, проблемой реализации в ИТС речевого обмена является создание эффективных методов сжатия, полученных на этапе регистрации речевых данных с возможностью дальнейшего воспроизведения исходных речевых сообщений с приемлемым для пользователя качеством.

Теоретические основы

Одной из особенностей звуков русской речи является сосредоточенность энергии в достаточно узком частотном диапазоне [2,3]. Эта особенность может быть использована в различных направлениях области обработки речевых сообщений: сжатие речевых данных, обнаружение и кодирование пауз, распознавание речи, очистка от шумов. При этом необходимо точно определять, в каком количестве частотных интервалов сосредоточена необходимая доля энергии. Далее эту особенность будем называть частотная концентрация, которая определяется минимальным количеством частотных интервалов, в которых сосредоточена заданная доля энергии [4]:

$$W_{NR} = l_{NR}^m / R, \quad (1)$$

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

m – заданное значение доли энергии сигнала;

R – количество частотных интервалов, на которые разбивается частотный диапазон;

N – длительность анализируемого отрезка.

$$l_{NR}^m = \min d_{NR}^m, \quad (2)$$

В качестве типичного примера на рис. 1 приведено распределение энергии отрезка речевого сигнала, соответствующего звуку «а» по частотным интервалам (рис. 1а), и количество частотных интервалов, соответствующих различным значениям величины m (рис. 1б).

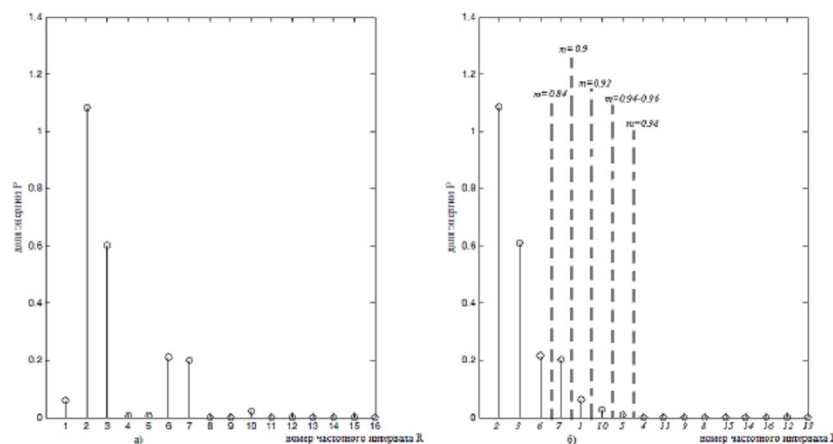


Рис. 1. Распределение энергии отрезка речевого сигнала, соответствующего звуку «а» по частотным интервалам, и количество частотных интервалов, соответствующее различным значениям m , при $N=160$ и $R=16$

Как видно из рисунка, основная энергия отрезка речевого сигнала, сосредоточена менее чем в половине частотных интервалов, т.е. её частотная концентрация менее 50%. Так для большинства звуков русской речи величина частотной концентрации составляет порядка 35% и только для шумоподобных звуков – порядка 55-60% [4].

Имея сведения о номерах частотных интервалов, в которых сосредоточена заданная доля энергии, можно осуществить сжатие речевых данных за счет избирательного воздействия (например, квантования по уровню) и 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 [2].

В ряде публикаций [5-7] описывается новый метод субполосного преобразования, оптимальный, с точки зрения минимума среднеквадратической погрешности аппроксимации, – трансформант Фурье исходного отрезка речевого сигнала в заданном частотном интервале. В этих работах также показаны преимущества этого метода перед современными аналогами.

В основе метода лежит новый математический аппарат с использованием субполосной матрицы вида:

$$A_r = \{a_{ik}^r\}; i, k = 1, \dots, N, \quad (3)$$

с элементами:

$$a_{ik}^r = \{\sin(v_r(i-k)) - \sin(v_{r-1}(i-k))\} / \pi(i-k); i, k = 1, \dots, N, \quad (4)$$

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 [8].

На основе этих матриц можно вычислять точные значения долей энергий отрезков речевых сигналов в выбранных частотных интервалах (5):



$$P_r = \bar{x} A_r \bar{x}^T, \quad (5)$$

где $\bar{x} = (x_1, \dots, x_N)^T$ – анализируемый отрезок речевых данных; T – операция транспонирования.

Данный математический аппарат можно использовать для осуществления избирательного воздействия на выбранные частотные интервалы на основе субполосного преобразования путем формирования составной матрицы, которая вычисляется как сумма субполосных матриц, соответствующих выбранным частотным интервалам, составляющим заданную долю энергии m :

$$A_\Sigma = \sum_{i=1}^{I_{NR}^m} A_i, \quad (6)$$

где A_i – субполосные матрицы, соответствующие тем частотным интервалам, которые составляют заданную долю энергии m .

Составная матрица обладает полной системой ортонормальных собственных векторов (7), соответствующих неотрицательным собственным числам (8):

$$Q_\Sigma = \{\bar{q}_{\Sigma 1}, \bar{q}_{\Sigma 2}, \dots, \bar{q}_{\Sigma N}\}, \quad (7)$$

$$L_\Sigma = \text{diag}(\lambda_{\Sigma 1}, \dots, \lambda_{\Sigma N}). \quad (8)$$

Собственные числа количественно равны сосредоточенным в выбранных частотных интервалах долям энергий, соответствующих собственным векторам и удовлетворяют условию:

$$0 \leq \lambda_{\Sigma k} \leq 1, k = 1, \dots, N. \quad (9)$$

Субполосное преобразование осуществляется следующим образом:

$$\bar{y}_\Sigma = \sqrt{L_\Sigma} Q_\Sigma^T \bar{x}, \quad (10)$$

где $\bar{x} = (x_1, \dots, x_N)^T$ – анализируемый отрезок речевых данных;

Q_Σ^T – матрица собственных векторов;

L_Σ – собственные числа, соответствующие собственным векторам.

Имеет место обратное субполосное преобразование:

$$\hat{\bar{x}} = Q_\Sigma^T \bar{y}_\Sigma. \quad (11)$$

Для оценки эффективности разработанного метода проводились вычислительные эксперименты по оценке погрешности [9] аппроксимации отрезков исходного речевого сигнала в выбранных частотных интервалах (13). В качестве исходных были использованы речевые данные, полученные в результате записи естественной речи различных дикторов, из которых выделялись отрезки, соответствующие определенным звуко сочетаниям. Для проводимых экспериментов параметры для исследования речевого материала – количество частотных интервалов $R=16$ и длительность анализируемого отрезка $N=160$.

$$\delta = \sqrt{\frac{(\bar{X} - \hat{\bar{X}}) * A_r * (\bar{X} - \hat{\bar{X}})^T}{\bar{X} * A_r * \bar{X}^T}}, \quad (12)$$

где \bar{X} – исходный отрезок речевого сигнала;

$\hat{\bar{X}}$ – восстановленный отрезок речевого сигнала;

A_r – субполосная матрица вида (3).



В табл. 1 приведены типичные результаты по оценке погрешности восстановления исходного отрезка речевого сигнала при использовании составной матрицы и КИХ-фильтра. При проведении вычислительных экспериментов величина параметра m для составной матрицы изменялась в диапазоне от 0,84 до 0,98, для КИХ-фильтра была выбрана длительность импульсной характеристики $\delta = 1024$.

Таблица 1

**Погрешность восстановления исходного отрезка речевого сигнала
в заданных частотных интервалах**

Номер частотного интервала (R)	Погрешность восстановления исходного отрезка речевого сигнала в заданных частотных интервалах	
	предлагаемый метод	КИХ-фильтр (δ_{1024})
1	$2,9896 \cdot 10^{-15}$	0,14154
2	$1,5425 \cdot 10^{-15}$	0,11291
3	$1,2771 \cdot 10^{-15}$	0,24117
4	$4,3957 \cdot 10^{-15}$	0,44707
5	$4,9288 \cdot 10^{-15}$	0,7355
6	$1,1038 \cdot 10^{-15}$	0,4721
7	$1,6473 \cdot 10^{-15}$	0,18875
8	$1,1405 \cdot 10^{-14}$	0,58453
9	$1,1175 \cdot 10^{-14}$	0,39325
10	$2,1645 \cdot 10^{-14}$	0,36162
11	$8,2484 \cdot 10^{-14}$	0,86369
12	$3,7525 \cdot 10^{-14}$	0,26371
13	$6,0482 \cdot 10^{-14}$	0,47618
14	$3,0724 \cdot 10^{-14}$	0,30675
15	$1,8253 \cdot 10^{-14}$	0,4378
16	$1,9927 \cdot 10^{-14}$	0,68117

По результатам проведенных вычислительных экспериментов можно судить о значительном превосходстве предлагаемого метода субполосного преобразования перед субполосным преобразованием на основе КИХ-фильтров.

На рис. 2 представлены типичные результаты вычислительных экспериментов по апробации разработанного метода.

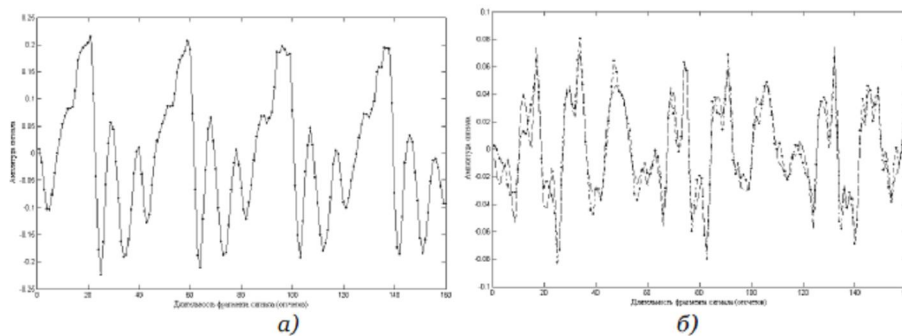


Рис. 2. Отрезок речевого сигнала, соответствующий звуку «а»:

- а) сплошная линия – исходный сигнал; пунктир (маркер точка) – восстановленный сигнал с использованием составной матрицы (СМ) при $m = 1$;
б) пунктирная линия – исходный сигнал; пунктир (маркер точка) – восстановленный сигнал с использованием банка КИХ-фильтров



Как видно из приведенных рисунков, форма сигнала, восстановленного с использованием составной субполосной матрицы, с высокой точностью повторяет форму исходного сигнала в отличие от сигнала, восстановленного с использованием банка КИХ-фильтров. Таким образом, можно утверждать о том, что при использовании метода субполосного преобразования на основе составной матрицы можно достичь высокого качества воспроизведения исходных речевых сигналов.

Для различных значений параметра m был проведен анализ зависимости количества неотрицательных собственных чисел (8), соответствующих собственным векторам (7), величина которых значительно больше нуля. Результаты представлены на рис. 3.

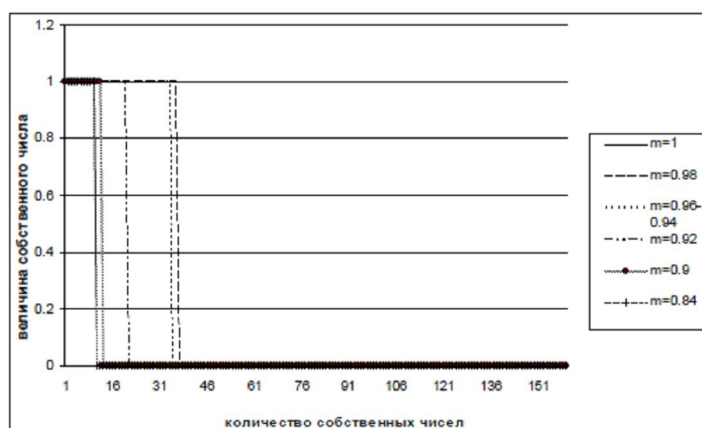


Рис. 3. Величины собственных чисел при различных значениях параметра m (звук «а»)

В табл. 2 приведены данные о количестве собственных чисел, величина которых значительно больше нуля (>0.01), при различных значениях параметра m .

Таблица 2

Количество собственных чисел, больших нуля при различных значениях параметра m (звук «а»)

Величина доли энергии (m)	Количество собственных чисел, значительно больших нуля (>0.01)
1	160
0,98	81
0.96-0,94	71
0,92	57
0,9-0,86	49
0.84	38

Так как собственные числа численно равны энергии собственного вектора, падающей в заданный частотный интервал, то при осуществлении субполосного преобразования можно пренебречь теми собственными векторами, собственные числа которых близки к нулю. Таким образом, можно уменьшить длину вектора субполосного преобразования, а следовательно, и объем данных, необходимых для хранения или передачи в 2-4 .

Субполосное преобразование осуществляется по формуле:



$$\hat{y}_{1\Sigma} = \sqrt{L_{\Sigma}} Q_{1\Sigma}^T \hat{x}, \quad (13)$$

где $\hat{x} = (x_1, \dots, x_N)^T$ – анализируемый отрезок речевых данных;

$Q_{1\Sigma}^T$ – матрица собственных векторов, собственные числа которых больше нуля.

Обратное субполосное преобразование будет осуществляться по формуле:

$$\hat{\hat{x}} = Q_{1\Sigma}^T \hat{y}_{1\Sigma}. \quad (14)$$

Ниже, в табл. 3, приведены результаты вычислительных экспериментов по оценке погрешности восстановления исходного отрезка речевого сигнала (12) с использованием составной субполосной матрицы при различных значениях параметра m . Номера частотных интервалов упорядочены в соответствии с величиной доли энергии, содержащейся в них.

Таблица 3

Погрешность восстановления отрезка речевого сигнала, соответствующего звуку «а», при различных значениях m

Номер частотного интервала (R)	Задаваемая доля общей энергии, которая сосредоточена в минимальном количестве частотных интервалов (m)						
	1	0.98	0.96	0.94	0.92	0.9	0.84
2	1,5425*10 ⁻¹⁵	0,013274	0,02046 5	0,02046 5	0,02699 3	0,02699 3	0,02699 3
3	1,2771*10 ⁻¹⁵	0,036811	0,04726	0,04726	0,04687 1	0,04687 1	0,046871
6	1,1038*10 ⁻¹⁵	0,010284	0,12446	0,12446	0,12458	0,12934	0,17951
7	1,6473*10 ⁻¹⁵	0,037149	0,04260 2	0,04260 2	0,04982 5	0,05754 5	–
1	2,9896*10 ⁻¹⁵	0,041878	0,06581 4	0,06581 4	0,07026 2	–	–
10	2,1645*10 ⁻¹⁴	0,13488	0,138	0,138	–	–	–
5	4,9288*10 ⁻¹⁵	0,20619	–	–	–	–	–
4	4,3957*10 ⁻¹⁵						

Результаты, приведенные в табл. 3, свидетельствуют о том, что избирательное воздействие на выбранные частотные интервалы с использованием составной субполосной матрицы позволяет достигать меньших величин погрешности восстановления исходных речевых данных по сравнению с КИХ-фильтром, даже при уменьшении параметра m .

Как видно из рис. 4, форма отрезка речевого сигнала, восстановленного с использованием составной матрицы, даже при величине параметра $m=0,92$ достаточно близка к форме исходного отрезка речевого сигнала. Этот факт позволяет говорить о целесообразности использования такого подхода в задачах обработки речевых данных и конкретно в задаче сжатия, т.к. близость формы восстановленного сигнала к исходной позволяет говорить о достаточно высокой степени разборчивости и узнаваемости диктора.

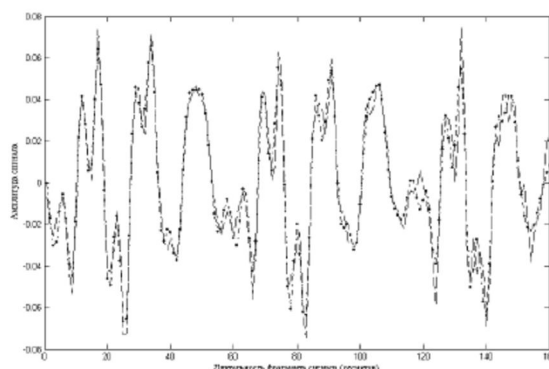


Рис. 4. Отрезок речевого сигнала, соответствующий звуку «А»: пунктирная линия – исходный сигнал; пунктир (маркер точка) – восстановленный сигнал с использованием составной матрицы (СМ) при $m = 0,92$

Выводы

Проведенные вычислительные эксперименты показали высокую эффективность разработанного метода с позиции сжатия речевых данных при сохранении заданного качества воспроизведения. Экспериментально было установлено, что предлагаемый метод избирательного воздействия на частотные компоненты речевого сигнала на основе использования суммы субполосных матриц обладает малой погрешностью восстановления. Таким образом, использование этого метода позволяет со значительной степенью точности восстановить исходный отрезок речевых данных. Этот факт позволяет утверждать о сохранении высокой степени разборчивости и узнаваемости в восстановленных речевых сообщениях.

Работа выполнена в рамках ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы (мероприятие 1.3.2 шифр лота: 2010-1.3.2-214-005, шифр заявки: 2010-1.3.2-214-005-020).

Литература

1. Ковалгин, Ю.А. Цифровое кодирование звуковых сигналов: учеб. пособие / Ю. А. Ковалгин, Э. И. Вологдин // СПб. : КОРОНА-принт, 2004. – 240 с.
2. Жилияков, Е.Г. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений: моногр. / Е.Г. Жилияков, С.П.Белов, Е.И. Прохоренко // Белгород, 2007. – 136 с.
3. Шелухин, О.И. Цифровая обработка и передача речи / О.И. Шелухин, Н.Ф. Лукьянцев; под ред. О.И. Шелухина // М.: Радио и связь, 2000. – 456с.: ил.
4. Болдышев А.В. О различиях распределения энергии звуков русской речи и шума / А.В. Болдышев, А.А. Фирсова // Материалы 12-й Междунар. конф. и выставки «ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ и ЕЁ ПРИМЕНЕНИЕ – DSPA'2010» 31 марта – 02 апреля 2010 года, г. Москва.
5. Прохоренко Е.И. Новый метод оптимального субполосного преобразования в задаче сжатия речевых данных / Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова, А.В. Эсауленко // Журнал «Вопросы Радиоэлектроники», серия электроника и вычислительная техника (ЭВТ). Выпуск №1 Москва 2010. с. 49-55.
6. Жилияков, Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. / Е.Г. Жилияков. – Белгород: Изд-во, 2007. – БелГУ, 2007. – 160.



7. Жилияков, Е.Г. Вариационные методы частотного анализа звуковых сигналов / Е.Г. Жилияков, С.П. Белов, Е.И. Прохоренко // Труды учебных заведений связи. – СПб, 2006. – № 174. – С.163-170.
8. Гантмахер, Ф.Р. Теория матриц / Ф.Р. Гантмахер. – М.: Физматлит, 2004. – 560 с.
9. Сизиков, В.С. Математические методы обработки результатов измерений: учебник для вузов / В.С. Сизиков. – СПб.: Политехника, 2001.

ABOUT ELECTORAL AFFECTING TO FREQUENCY COMPONENTS OF SPEECH SIGNALS IN RESOLVE OF COMPRESSION

A. V. BOLDYSHEV

Belgorod State University

e-mail:
boldyshev@bsu.edu.ru

In the article expounded the new method of the electoral affecting on frequency components of speech signals, based on the use of sub-band transformations on component frequency intervals. The results of calculable experiments are resulted as evaluated by efficiency of the developed method.

Key words: informatively-telecommunication technologies, compression of speech data, sub-band transformation, frequency interval, set parts of energy.