

Е.Г. ЖИЛЯКОВ, д.т.н., проф. НИУ "БелГУ", Белгород,
А.В. БОЛДЫШЕВ, аспирант НИУ "БелГУ", Белгород,
А.В. КУРЛОВ, аспирант НИУ БелГУ, Белгород,
А.А. ФИРСОВА, аспирант НИУ "БелГУ", Белгород,
А.В. ЭСАУЛЕНКО, магистр НИУ "БелГУ", Белгород

ОБ ИЗБИРАТЕЛЬНОМ ПРЕОБРАЗОВАНИИ ЧАСТОТНЫХ КОМПОНЕНТ РЕЧЕВЫХ СИГНАЛОВ В ЗАДАЧЕ СЖАТИЯ

В статье приведены результаты вычислительных экспериментов по апробации метода сжатия речевых данных на основе избирательного преобразования частотных компонент речевых сигналов, полученных с помощью нового метода субполосного частотного анализа/синтеза. Ил.: 1. Табл.: 2. Библиогр.: 8 назв.

Ключевые слова: частотные компоненты речевых сигналов, сжатие речевых данных, субполосный частотный анализ/синтез.

Постановка проблемы и анализ литературы. Одной из особенностей звуков русской речи является сосредоточенность энергии в достаточно узких частотных диапазонах, суммарная ширина которых гораздо меньше частоты дискретизации [1 – 5]. Эта особенность может быть использована в различных направлениях области обработки речевых сообщений, в том числе, в задаче сжатия речевых данных, для этого необходимо точно определить, в каком количестве частотных интервалов сосредоточена основная доля энергии [4, 6]. Чтобы определить количество частотных интервалов, в которых сосредоточена основная доля энергии, можно воспользоваться понятием частотной концентрации, которая определяется минимальным количеством частотных интервалов, в которых сосредоточена заданная доля энергии отрезка речевого сигнала [6]:

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

где l_{NR}^{tm} – минимальное количество частотных интервалов, в которых сосредоточена заданная доля энергии отрезка речевого сигнала, так что имеет место

$$l_{NR}^{tm} = \min d_{NR}^{tm}; \quad (2)$$

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

t – обозначает один из анализируемых речевых отрезков, порождаемых звуком русской речи;

N – значение длины анализируемого отрезка;

m – доля общей энергии, задаваемая для определения минимального количества частотных интервалов, в которых она сосредоточена.

Для правых частей (2) выполняется неравенство

$$\sum_{k=1}^{d_{NR}^{tm}} P_{(k)N} \geq m \|\bar{x}_N\|^2 = m \sum_{i=1}^N x_i^2, \quad (3)$$

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

T – операция транспонирования.

Индекс в скобках, у слагаемых суммы слева соотношения (3) означает, что части энергий P_{kN} упорядочиваются по убыванию:

$$P_{(k)N} \in \{P_{rN}, r = 1, \dots, R\}; P_{(k+1)N} \leq P_{(k)N}, k = 1, \dots, R. \quad (4)$$

В качестве примера в таблице 1 приведено минимальное количество частотных интервалов, составляющих заданную долю энергии, для звука русской речи "И".

Таблица 1. Минимальное количество частотных интервалов и частотная концентрация при различных значениях параметра m (звук "И")

m	I_{NR}^{tm}	W_{NR}^t
0,98	5	0,3175
0,96	3	0,1875
0,9 – 0,94	2	0,125
0,86 – 0,88	1	0,0625

При заданной доле энергии 0,98 частотная концентрация для приведенного примера составляет 0,3125. Для большинства звуков русской речи величина частотной концентрации составляет порядка 0,35 и только для шумоподобных звуков – порядка 0,55 – 0,60 [6]. На основании сведений 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.

Целью статьи является анализ эффективности сжатия речевых данных за счет использования предлагаемого метода субполосного частотного анализа/синтеза.

Одним из способов получения составляющих речевого сигнала, соответствующих выбранным частотным интервалам, является субполосный частотный анализ/синтез. В настоящее время наибольшее

распространение получил метод на основе банка КИХ-фильтров, однако, этот метод обладает рядом недостатков, которые приводят к увеличению погрешностей восстановления данных [3]. В ряде публикаций [1 – 3] описывается метод субполосного анализа/синтеза, оптимальный с точки зрения минимума среднеквадратической погрешности аппроксимации трансформант Фурье исходного отрезка речевого сигнала в заданном частотном интервале, а также показаны преимущества этого метода перед современными аналогами. В основе предлагаемого метода лежит математический аппарат с использованием субполосных матриц вида:

$$A_r = \{a_{ik}^r\} = \{\sin(v_r(i-k)) - \sin(v_{r-1}(i-k))\} / \pi(i-k), \quad i, k = 1, \dots, N, \quad (5)$$

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 [3, 7]. На основе этих матриц можно вычислять точные значения долей энергий отрезков речевых сигналов в выбранных частотных интервалах

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

Данный математический аппарат можно использовать для получения частотных компонент исходного анализируемого отрезка речевого сигнала, которые отражают заданную долю энергии, сосредоточенную в выбранных частотных интервалах l . Для этого будет использован метод субполосного преобразования, основанный на формировании составной матрицы, которая вычисляется как сумма субполосных матриц, соответствующих выбранным частотным интервалам, составляющих заданную долю энергии m

$$A_\Sigma = \sum_{r=1}^{d_{NR}^m} A_{(r)}, \quad (7)$$

где $A_{(r)}$ – субполосные матрицы, соответствующие тем частотным интервалам, которые составляют заданную долю энергии m .

Составная матрица обладает полной системой ортонормальных собственных векторов (8), соответствующих неотрицательным собственным числам (9):

$$Q_\Sigma = \{\bar{q}_{\Sigma 1}, \bar{q}_{\Sigma 2}, \dots, \bar{q}_{\Sigma N}\}, \quad (8)$$

$$L_{\Sigma} = \text{diag}(\lambda_{\Sigma 1}, \dots, \lambda_{\Sigma N}). \quad (9)$$

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

$$0 \leq \lambda_{\Sigma k} \leq 1, \quad k = 1, \dots, N. \quad (10)$$

Субполосное преобразование осуществляется следующим образом

$$\vec{y}_{\Sigma} = \sqrt{L_{\Sigma}} Q_{\Sigma}^T \vec{x}, \quad (11)$$

где $\sqrt{L_{\Sigma}}$ – корень из диагонального элемента, соответствующего определенному собственному вектору.

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 субполосного преобразования, 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. Выражение (6) можно представить как

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

где $\sum_{i=1}^{J_{\Sigma}} y_{\Sigma i}^2$ – первое слагаемое, в котором λ_i – собственные значения суммарной матрицы, величина которых достаточно большая;

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

Доля энергии второго слагаемого, настолько мала, что ею 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

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

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

Вычислительные эксперименты. В качестве исходных данных использовано большое количество речевого материала, который был получен в результате записи естественной речи различных дикторов. Для 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 $m = 0,86 \div 0,98$. В качестве примера в табл. 2 приведено минимальное количество собственных значений J_{Σ} для звука русской речи "И".

Таблица 2. Звук "И", $N = 160$, $R = 16$

$c \backslash m$	0.86 – 0.88	0.9 – 0.94	0.96	0.98
0.89 – 0.9	10	18	28	46
0.91	10	19	29	47
0.92 – 0.93	10	19	29	48
0.94 – 0.95	11	20	30	49
0.96	11	20	31	51
0.97	12	21	32	52
0.98	12	22	33	54
0.99	13	22	34	56
0.995	14	23	36	58

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 возможности четырехкратного сокращения объема памяти, требуемого для хранения данного звука. Однако, выбираемое количество собственных значений будет сказываться на качестве воспроизведения речевого сообщения. Ниже на рис. 1 приведена полученная степень сжатия для всех звуков русской речи. Степень сжатия определялась следующим образом [8]:

$$K = N / J_{\Sigma} .$$

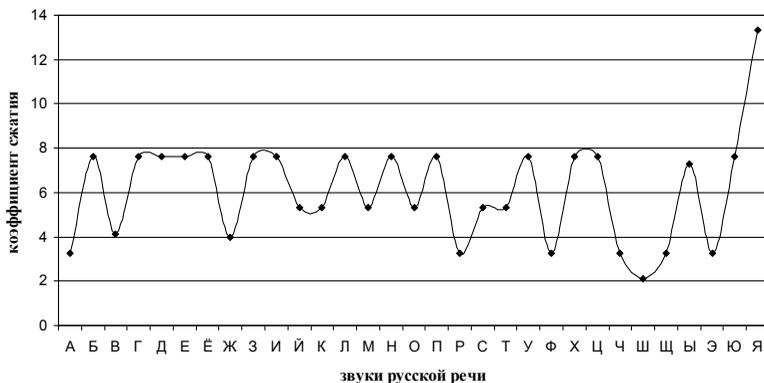


Рис. Степень сжатия для различных звуков русской речи
($c = 0.92$, $m = 0.92$)

Выводы. Проведенные вычислительные эксперименты показали высокую эффективность разработанного метода с позиции сжатия речевых данных при сохранении приемлемого для пользователя качества воспроизведения. Предлагаемый подход к сжатию речевых сигналов в среднем позволяет достичь сжатия в 6 – 7 раз без существенной потери качества воспроизведения. Этот показатель можно увеличить за счет использования известных алгоритмов обнаружения и удаления пауз.

Список литературы: 1. Жиликов Е.Г. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений: монография / Е.Г. Жиликов, С.П. Белов, Е.И. Прохоренко // Белгород: Изд-во БелГУ, 2007. – 136 с. 2. Шелухин О.И. Цифровая обработка и передача речи / О.И. Шелухин, Н.Ф. Лукьянцев. – М.: Радио и связь, 2000. – 456 с. 3. Жиликов Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: монография / Е.Г. Жиликов. – Белгород: Изд-во БелГУ, 2007. – 160 с. 4. 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 ЭВТ. – Вып. №1. – М.: 2010. – С. 49 – 55. 5. Ковалгин Ю.А. Цифровое кодирование звуковых сигналов / Ю.А. Ковалгин, Э.И. Вологдин. – Изд-во: Корона Принт, 2004. – 240 с. 6. Болдышев А.В. О различиях распределения энергии звуков русской речи и шума / А.В. Болдышев, А.А. Фирсова // Материалы 12-ой Международной конференции и выставки "Цифровая обработка сигналов и её применение. – "DSP'2010". – М.: 2010. – С. 204 – 207. 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. – 240 с.

УДК 621.391

Про виборче перетворення частотних компонент мовних сигналів в завданні стиснення / Жіляков Е.Г., Болдышев А.В., Курлов А.В., Фірсова А.А., Есауленко А.В. // Вісник НТУ "ХПІ". Тематичний випуск: Інформатика і моделювання. – Харків: НТУ "ХПІ". – 2011. – № 17. – С. 37 – 43.

У статті приведені результати обчислювальних експериментів по апробації методу стиснення мовних даних на основі виборчого перетворення частотних компонент мовних сигналів, отриманих за допомогою нового методу субсмугового частотного аналізу/синтезу. Іл.: 1. Табл.: 2. Бібліогр.: 8 назв.

Ключові слова: частотні компоненти мовних сигналів, стиснення мовних даних, субполосний частотний аналіз/синтез.

UDK 621.391

About electoral transformation frequency component of vocal signals in the task of compression / Zhilyakov E.G., Boldyshev A.V., Kurlov A.V., Firsova A.A., Esaulenko A.V. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2011. – №. 17. – P. 37 – 43.

In this article the results of calculable experiments are resulted on approbation of method of compression of vocal data on the basis of electoral transformation frequency component of vocal signals, got by the new method of subband frequency analysis/synthesis. Figs.: 1. Tabl.: 2. Refs.: 8 titles.

Keywords: frequency components of speech signals, compress speech data, sub-band frequency analysis/synthesis.

Поступила в редакцію 03.02.2011