УДК 004.934.2

Д.И. ТРУБИЦ^1НА
D.I. TRUBITSYNA

О СУБПОЛОСНОМ МЕТОДЕ СЖАТИЯ И ВОССТАНОВЛЕНИЯ РЕЧЕВЫХ ДАННЫХ В ОБЛАСТИ ОПРЕДЕЛЕНИЯ КОСИНУС-ПРЕОБРАЗОВАНИЯ

ABOUT SUBBAND METHOD OF COMPRESSION AND RECOVERY OF VOICE DATA IN THE FIELD OF DETERMINING COSINUS TRANSFORMATION

В работе рассмотрен метод сжатия и восстановления речевых данных на основе субполосного анализа/синтеза в области косинус-преобразования. Представлены алгоритмы субполосного преобразования в области определения косинус-преобразования для выполнения операции субполосного кодирования и декодирования речевых сигналов.

Ключевые слова: субполосный анализ, сжатие речевых данных, субполосная матрица, косинус-преобразование.

The paper considers a method for speech data compression and recovery based on subband analysis/synthesis in the cosine transform domain. Algorithms of sub-band transformation in the field of determining the cosine transformation for performing operations of sub-band encoding and decoding of speech signals are presented.

Keywords: subband analysis, speech data compression, subband matrix, cosine transform.

Введение

Информационный обмен на основе устной речи является наиболее естественным для человека, наблюдается устойчивый рост объемов речевых данных (звукозаписи лекций, аудиокниги, системы массового звукового оповещения, голосовые помощники, системы навигации, подкасты) в информационно-телекоммуникационных системах, в которых осуществляется их хранение и передача. Однако необходимость хранения длительных записей делает целесообразным их сжатие, что сводится к процедуре перекодирования речевых данных исходного файла, при котором объем будет меньше, чем у исходного. Известными способами сокращения объемов речевых данных являются кодирование пауз (VAD), использование методов линейного предсказания, кодирование данных на основе кодеков MP-3, OGG и т. д. Современные методы кодирования пауз являются недостаточно эффективными, так как не позволяют с высокой точностью определить начало и конец паузы. Методы линейного предсказания нашли применение в системах IP-телефонии и мобильной связи. Данный метод позволяет стандартизовать количество коэффициентов линейного предсказания p=8-12 и создать кодовые книги. Однако в связи с этим теряется узнаваемость диктора и возникают сложности с адекватным определением возбуждающего воздействия. Речевые сигналы (вокализованные) состоят из узкополосных компонент, поэтому всегда есть опасность подбора таких коэффициентов, которые могут полностью подавить эти компоненты. Вот почему VAD на основе ЛП могут пропускать паузы. Если нет выраженных участков спектра с подъемом энергии, то выбираться порядок фильтра будет очень большим. Методы на основе кодеков MP-3, OGG в основном используются для обработки музыкальных данных, что делает использование данных кодеков не эффективным для обработки речевых данных, так как не будет учитываться присутствие пауз в речевых сообщениях [3, 4].

Для достижения высокой эффективности сжатия необходимо использовать математический аппарат, адекватно отражающий свойства речевых сигналов, обусловленные нестационарностью речевых сигналов и высокой концентрацией энергии их отрезков в малой доле частотной полосы. Последнее свойство соответствует представлению о том, что слуховой аппарат человека функционирует как набор фильтров с пересекающимися частотными характеристиками, что позволяет выделять информационные компоненты, отражающие свойства звуков речи.

В данной статье предлагается для реализации субполосного подхода использовать косинус-преобразование следующего вида:

$$ад = \sum_{t=i}^{N} Xj \cos(zi), \quad (1)$$

которое является периодическим, и поэтому в качестве области определения рассматривается следующий интервал оси z:

(2)

где X = (X]_ ,X2 ,. . .,Xj,. . .,x_N) (T- символ транспонирования) - отсчеты анализируемого речевого сигнала.

- трансформанта косинус-преобразования.

Ввиду справедливости тригонометрического тождества:

$$\cos((2n - z) i) = \cos(z i), \quad (3)$$

имеет место симметрия:

$$X(z) = X(2n - z), \; O < z < п. \quad (4)$$

Данное свойство позволяет рассматривать только половину области определения (2).

В свою очередь, справедливо свойство ортогональности:

$$2 \cos(z i) \cos(zm)\, dz/п = б^{\wedge\wedge\wedge}, \quad (5)$$

где $6ifn$ - символ Кронекера.

$$A \quad \begin{cases} fl, I = ш \\ “ I O, I \wedge m ■ \end{cases}$$

Поэтому справедливо и обратное преобразование следующего вида:

$$x^{\wedge} = 2 I \int_{Jo}^{Г} X(z) \cos(zm)\, dzjn, \quad (6)$$

и равенство Парсеваля относительно эвклидовых норм (энергии):

$$\sum_{1=1}^{N} xf = 2 \int_{0}^{гт} X^{\wedge}(z) dzln, \quad (7)$$

N - количество отсчетов в окне анализа.

Соотношение (2.7) можно представить в субполосном виде:

$$|2 = \sum_{i=l}^{N} \quad = \sum_{r=1}^{R} \quad (x), \quad (8)$$

$$P\text{-}m(x) = 2 \int_{Zj\text{-}\backslash}^{rZ'Z\,г} {}^{\wedge}X^2(z)\, dz/n, \; г = 1,\ldots,R, \quad (9)$$

$$■2'l, о = {}^{O}, '{}^{\wedge}2Д - 1 = {}^{n},$$

$$^{\wedge}1,i < {}^{\wedge}2,j.$$

В дальнейшем характеристику вида (9) будем именовать частью энергии отрезка сигнала, попадающей в соответствующий частотный интервал.

Подставив в (9) определение (1), нетрудно получить представление для части энергии:

$$Pj. (x) = x'^Pj\text{-}x, r = 1,. . ,,R \qquad (10)$$

где By. - субполосная матрица вида:

$$Bf = A y + Cy, \qquad (11)$$

где - субполосная матрица , элементы которой определяются в соответствии с соотношениями [14, 71]:

$$a_{ik} = \sin (z_2,r\text{-}(i - \kappa) - \sin (z\text{^}y\text{^}i - \quad - \kappa), \qquad (12)$$

- элементы матрицы вычисляются в соответствии с соотношениями:

$$Cii\text{^} = \sin (z_2,r\text{-}(i + k) - \sin (z\text{^}y\text{^}i + k)))/n(i + k), \qquad (13)$$

Соотношение (10) позволяет вычислить части энергии непосредственно в области значений речевых сигналов, не вычисляя при этом трансформанту косинус-преобразования.

Предлагается использовать субполосное преобразование в области определения косинус-преобразования для выполнения операции субполосного кодирования и декодирования речевых сигналов.

В качестве исходных данных для проведения вычислительных экспериментов использовались речевые сигналы, которые записывались разными дикторами при следующих параметрах записи: частота дискретизации ^3=8 кГц, разрядность 8 бит. Фрагменты речи, которые соответствуют определенному звуку русской речи, выделялись из записи на слух. Длительность окна анализа в современных алгоритмах кодирования выбирается порядка 16-20 мс, что соответствует 128-160 отсчетам при частоте дискретизации 8 кГц. Выбор окна анализа N = 128 отсчетов обусловлен тем, что в среднем длительность гласных звуков 15 мс, а согласных - от 8 до 30 мс, и это поможет избежать захвата фрагментов, соответствующих другим звукам.

Алгоритм кодирования.

1. Задать длину окна анализа N;
2. Вычислить границы частотных интервалов:

$$.Zij. - {}_0,.Z_2y\text{-} - Ti;$$

$$z_1 y = (4тг/N) \cdot г, \; z_1 y = (4тг/N) \cdot (г \quad 1)$$

3. Вычислить количество частотных интервалов

$$R = 7Г/(47Г/A)$$

4. Вычислить субполосные матрицы косинус-преобразования B_y для частотных интервалов:

$$By - Ay + Cy$$

где - субполосная матрица , элементы которой определяются в соответствии с соотношениями [2]:

$$a[д. = \sin (z_2,r\text{-}(i - \kappa) - \sin (z\text{^}y\text{^}i - k)))/n(i - \kappa),$$

- элементы матрицы вычисляются в соответствии с соотношениями:

$$Cii\text{^} = \sin (z_2,r\text{-}(i + k) - \sin (z\text{^}y\text{^}i + k)))/n(i + k).$$

5. Вычислить значение долей энергии речевого сигнала:

$$Pj\text{-}(x) = ByX, г = 1, \ldots, R$$

6. Определить пороговое значение:

$$p = ||x| \quad \frac{}{n}, г = 1 \ldots R,$$

где Af - ширина анализируемого частотного интервала, который определяется согласно соотношению

$$— \text{^}2,г \text{^}1,г$$

7. Определить информационные частотные интервалы, при этом предполагается:

$$Pг(x) > p,$$

8. Сформировать $Mask(R)$:

$$Mask(R) = ;1, Pг(Ю > P \quad .0, Pj.(x) < p$$

9. Для субполосной матрицы By определить собственные числа и собственные вектора:

$Gy = (p]\_y, \ldots, gj\backslash iy\text{^}$ - ортонормированная матрица собственных векторов, причем имеет место свойство:

$$By\ Gy = HyGy\text{^}$$

$$Gy\text{^}Gy = GyGy\text{^} = / =$$

- диагональная матрица собственных чисел,

10. Вычислить $/_y]\_ = [iV (z_{2},_{y} \_ {}_1 — Z]\_,_y\_]\_) /тг] — 1,$

[] - целая часть числа.

В ходе экспериментов было определено, что для восстановления речевого сигнала с достаточной степенью разборчивости могут быть использованы три собственных числа значения, которых близки к единице.

11. Сформировать матрицу из близких к единице собственных чисел и соответствующих им собственных векторов:

$$P\text{^}1\_m(Haд\text{^}h\text{-^}yf\ h2yr\ '"\text{^^}jy\text{-}iT\text{^}$$

$$Gir — \{\text{^}gir> \cdots ' д]y\text{^}r)$$

12. Вычислить проекции исходного сигнала на отображение собственных векторов:

$$Pir \sim gir\backslash\text{^} \sim 1. \blacksquare\blacksquare\blacksquare./ri$$

13. Сформировать файл, содержащий значения проекций $/?j_y$ и $Mask(R)$, который будет использован при декодировании.

Алгоритм декодирования:

1. Задать длину окна анализа $N;$
2. Вычислить границы частотных интервалов:

$$Ziy — 0, Z2y — {}^{г},$$

$$z_{1y} = (4г / V) \cdot г,\ z_{1y} = (4тг / V) \cdot (г — 1)$$

3. Вычислить количество частотных интервалов

$$R = 7Г/(47Г/A)$$

4. Вычислить субполосные матрицы косинус-преобразования B_y для частотных интервалов:

$$B_y = i4_y + G_y,$$

где - субполосная матрица , элементы которой определяются в соответствии с соотношениями [2]:

$$a[д. = \sin(z_{2},r\text{-}(i — к) — \sin(z\text{^}y\text{^}i — k)))/n(i — к),$$

Cr = { с[д.} - элементы матрицы вычисляются в соответствии с соотношениями:

с[д. = sin (z2,r-(i + к) — sin (zir-(i + k)))/n(i + к).

G_r = (,^i$_r$,. . ■, Qn -T) - ортонормированная матрица собственных векторов, причем:

$$ByGy = Hj\text{-}Gj\text{-}\wedge$$

$$Gf\wedge Gf — Gf\text{-}Gf\wedge = / =$$

H_r = $diag$ ($_r$, /12 $_r$,..., h_{Nr}) - диагональная матрица собственных чисел.

5. Вычислить

$$Jr_1 = [N(z2, r_{-1} - Zi\, r - i)/n] - 1,$$

[] - целая часть числа

6. Сформировать матрицу из близких к единице собственных чисел и соответствующих им собственных векторов:

$$H1\,r \quad diag\,(h\, 1r*h2\, r* \dots *h]r\backslash r),$$

$$G\, ir = (d\, ir' \, .. \, .'d]rl\, r).$$

7. Загрузить файл, содержащий значения проекций i_r и $M\ as\ к\ \{R\}$.

8. Вычислить субполосное преобразование информационных частотных компонент:

$$Уг \sim \sum_{1=1}^{Jrl} Pirdir$$

9. Восстановить речевой сигнал:

$$\wedge Уг \sim \wedge Уг \quad \sum_{Г=1}^{R} \quad \sum_{Г=1}^{Hi}$$

где - количество информационных частотных интервалов, в которых сосредоточена подавляющая доля энергии.

10. Сохранить восстановленное речевое сообщение.

На рисунке представлен фрагмент восстановленного и исходного речевого сигнала при одинаковом количестве уровней квантования.

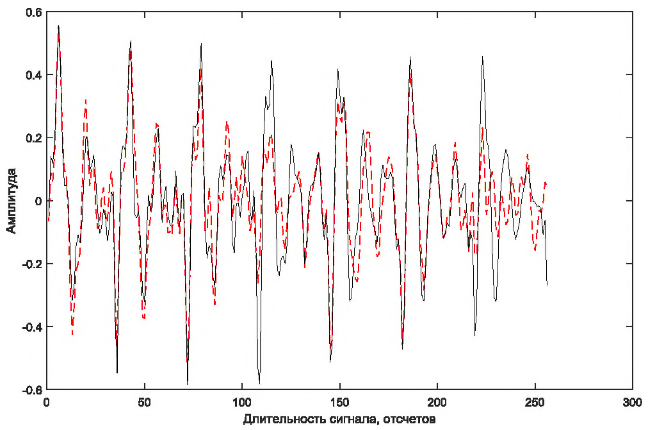

Рисунок 1 - Фрагмент сигнала, соответствующий звуку «А» (пунктирная линия - исходный сигнал, сплошная линий - восстановленный сигнал)

Результаты вычислений по оценке степени сжатия представлены в таблице 1.

Степень сжатия определяется как отношение объема исходного файла речевых данных $V_{исх}$ (бит) к объему файла, полученному в результате преобразований ¥$_{сж}$ (бит) с учетом служебной информации :

„ _ KlCX

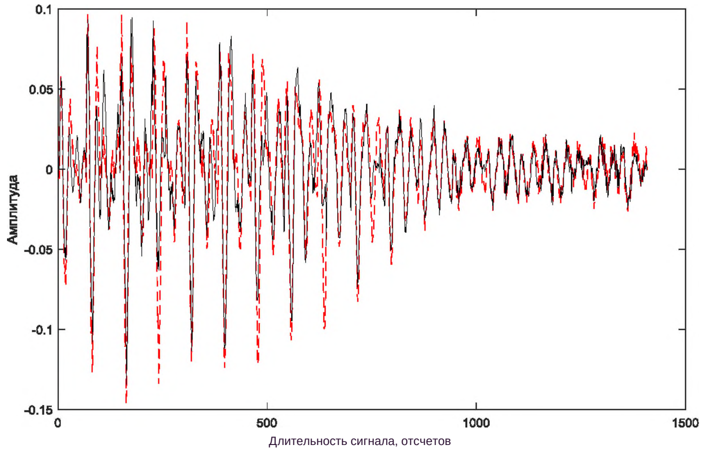

Рисунок 2 - Фрагмент сигнала, соответствующий звуку «З» (пунктирная линия - исходный сигнал, сплошная линий - восстановленный сигнал)

Таблица 1 - Результаты оценки сжатия речевых данных

Фрагмент речевого сообщения	Умсх, бит	Усж 4" Ум as к >	бит Степень сжатия,
Слово «Слушай»	434 176	59 328	27,3
Слово «Четыре»	401 408	43 008	29, 5
Слово «Назад»	212 992	18 432	21,5

Можно отметить, что использование данного метода позволяет добиться высокой степени сжатия. Однако для сохранения компонент векторов субполосного представления используется 8 бит. Уменьшение разрядности позволит увеличить степень сжатия. Таким образом, целесообразно использовать процедуру квантования по уровню компонент субполосного преобразования.

СПИСОК ЛИТЕРАТУРЫ

1.Жиляков Е.Г. Субполосн^гй анализ и синтез сигналов в рамках косинусного преобразования [Текст] / С.П. Белов, Е.Г. Жиляков, А.В. Коськин, Д.И. Трубицына // Научно-технический журнал Информационн^ге системы и технологии. № 4 (114) июль-август 2019. С. 13-22

2. Жиляков Е.Г. О сжатии речевых сигналов [Текст] / Е.Г. Жиляков, С.П. Белов, Е.И. Прохоренко // Вестник Национального технического университета «ХПИ». - 2005. - вып.56. - С. 32-41.

3. Жиляков Е.Г. Методы обработки речевых данных в информационно - телекоммуникационных системах на основе частотных представлений [Текст] / Е.Г. Жиляков, С.П. Белов, Е.И. Прохоренко. - Белгород: Изд-во БелГУ, 2007. - 136 с.

4. Рабинер Л.Р. Теория и применение цифровой обработки сигналов [Текст] / Л.Р. Рабинер, Б. Гоулд. - М.: Мир, 1975. - 835с.: ил.

5. Болдышев А.В. Метод сжатия речевых данных на основе оптимального субполосного преобразования по составным частотным интервалам [Текст] / А.В. Болдышев // Научные Ведомости БелГУ, серия «История. Политология. Экономика. Информатика». - Белгород: 2011. - № 1 (96). - Вып. 17/1 - С. 217-222.

Трубицына Диана Игоревна
Белгородский государственный национальный исследовательский университет, г. Белгород
Старший преподаватель кафедры информационно-телекоммуникационных систем и технологий
Тел.: +7 (4722) 301392
E-mail: trubitsyna@bsu.edu.ru