

А.В. КОРОЛЕВ, д-р техн. наук,
Ю.В. ДАНИЮК

АДАПТИВНЫЙ МЕТОД СЖАТИЯ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ДИСКРЕТНОГО ПРЕОБРАЗОВАНИЯ УОЛША

Запропоновано адаптивний метод стиснення голосових сигналів на основі дискретного перетворення Уолша.

The adaptive method of compression of speech signals is offered on basis of discrete Walsh transform.

Постановка проблемы. В современных сетях связи существенная часть пропускной способности каналов используется для передачи данных и речи. Поэтому важным является эффективное использование полосы пропускания, используемой речевыми сигналами. Для обеспечения такой эффективности применяются различные методы сжатия.

Обзор литературы свидетельствует о наличии нескольких направлений методов сжатия в этой области [1 – 8]. Как основные следует считать:

- 1) методы кодирования с линейным и нелинейным предсказаниями;
- 2) вокодерные методы кодирования;
- 3) методы кодирования на основе психоакустических особенностей слуха человека;
- 4) методы кодирования с использованием ортогональных преобразований;
- 5) гибридные методы.

В каждом из направлений есть свои достоинства и недостатки. С учетом их следует отметить, что требование малой вычислительной сложности и простоты реализации методов является существенным. Это обусловлено широким их использованием в сетях передачи данных и IP-телефонии, в которых помимо речевых сообщений передается и другая информация.

Целью данной статьи является разработка адаптивного метода сжатия речевых сигналов на основе дискретного преобразования Уолша.

Преобразование Уолша раскладывает исходный сигнал $x(t)$ в ряд [2]:

$$x(t) = \sum_{k=0}^{N-1} d_k wal_w(k, t), \quad (1)$$

$$\text{где } wal_w(k, t) = \begin{cases} cal(s_k, t), & \text{если } k - \text{четное} ; \\ sal(s_k, t), & \text{если } k - \text{нечетное}. \end{cases} \quad (2)$$

При этом частота s_k функции $wal_w(k, t)$ определяется как

$$s_k = \begin{cases} 0, & k = 0; \\ k/2, & k - \text{четное}; \\ (k+1)/2, & k - \text{нечетное}. \end{cases}$$

Таким образом, выразив $wal_w(k, t)$ через составляющие sal и cal , получим

$$x(t) = a_0 wal_w(0, t) + \sum_{k=1}^N [a_k cal(k, t) + b_k sal(k, t)]. \quad (3)$$

В [1] предлагается заменить выражение (3) рядом

$$x[t] = a_0 wal_w(0, t) + \sum_{k=1}^N a_k cal(k, t). \quad (4)$$

Такая замена обусловлена распределением энергии между коэффициентами трансформанты Уолша. Это позволяет сжать исходный сигнал в 2 раза с небольшими потерями энергии. Дальнейшее, более подробное изучение свойств преобразования Уолша и речевых сигналов показало, что существует возможность дополнительного сжатия за счет использования их свойств.

В речевых сигналах имеются участки с различной степенью изменений его значений на длине рассматриваемого блока. С целью упорядочивания блоков проведем их классификацию: 1 – не меняющие своего значения; 2 – меняющие не значительно; 3 – меняющие значительно. Определение принадлежности блока к тому или иному классу будет осуществляться по алгоритму, приведенному на рис. 1.

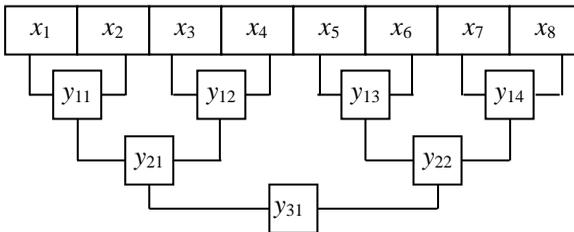


Рис. 1. Пример алгоритма определения характера изменений сигнала ($N = 8$)

Суть работы алгоритма состоит в том, что с его помощью определяется класс текущего блока. Введем специальный коэффициент изменчивости сигнала и обозначим его w . Указанный коэффициент находится следующим образом. Сначала сравниваются попарно по два соседних отсчета исходного сигнала (длина блока всегда равна некоторой целой степени двойки – 4, 8, ...)

путем нахождения модуля их разности – y_{ij} , где i – номер уровня сравнения, а j – номер пары на уровне. Если значение модуля разности отвечает заранее определенным условиям, то производится операция вычитания, но уже со значениями сравниваемых модулей разности на предыдущем шаге и т.д. Остановка алгоритма происходит в случае перехода алгоритма на последний уровень или, если на каком-то уровне значение модуля разности не соответствует заранее заданному значению.

В дальнейшем целесообразно присвоить значение коэффициенту w в зависимости от значения модуля разности y_{ij} . При этом следует определить правила работы алгоритма при сравнении модулей разности: если $y_{ij} = 0$, то $w = 0$ и сигнал не меняется на всей длине блока; если $y_{ij} \in [1; 2]$, то $w = 1$ и сигнал практически не меняется; если $y_{ij} \in [3; \infty]$, то $w = 2$ и сигнал меняется значительно.

После того, как определен класс текущего блока, принимается решение о том, сколько коэффициентов трансформанты Уолша необходимо оставить для восстановления исходного сигнала. Такой подход возможен из-за наличия следующего свойства преобразования Уолша. При разложении сигнала в ряд вида (4), а затем его восстановлении оказалось, что в полученном блоке отсчетов восстановленного сигнала усредненными окажутся по два соседних отсчета. А именно

$$\bar{x}_i = \bar{x}_{i+1} = \frac{(x_i + x_{i+1})}{2}, \quad (5)$$

где \bar{x}_i – отсчет восстановленного сигнала; x_i – отсчет исходного сигнала.

Данное свойство работает и в том случае, если ряд (4) содержит не все составляющие, т.е. при исключении $3/4$ коэффициентов трансформанты получим по четыре усредненных отсчета. Если оставить только один коэффициент трансформанты, то получим N усредненных значений исходного сигнала и т.д.

В общем виде выражение (5) принимает вид:

$$\bar{x}_i = \begin{cases} \frac{1}{N} \sum_{j=1}^Z x_j, & \text{если } w = 0; \\ \left\{ \begin{array}{l} 1 \leq i \leq N/2; \frac{1}{Z} \sum_{j=1}^Z x_j \\ N/2 < i \leq N; \frac{1}{Z} \sum_{j=Z}^N x_j \end{array} \right. & \text{если } w = 1; \\ \frac{1}{Z} \sum_{j=1}^Z x_j, & \text{если } w = 2, \end{cases} \quad (6)$$

где $Z = 2^{\log_2(N)-w}$ – коэффициент нормировки; N – длина обрабатываемого блока; w – коэффициент изменчивости сигнала.

В выражении (6) Z представляет собой коэффициент нормировки, значение которого зависит от класса, к которому принадлежит блок и его длины. Следует также отметить, что обязательным для использования указанного свойства, является наличие постоянной составляющей $a_0 wal_w(0, t)$, которая несет основную энергию сигнала.

Алгоритм останавливается если, достигнут последний уровень, либо если $w = 2$. Сразу после остановки алгоритма фиксируется уровень, на котором произошел останов и определяется значение коэффициента w .

Таким образом, в результате работы алгоритма по значению коэффициента w возможно определить количество коэффициентов трансформанты Уолша, необходимых для восстановления исходного сигнала.

Среднее процентное содержание блоков по предложенной классификации в речевых сигналах приведено на рис. 2.

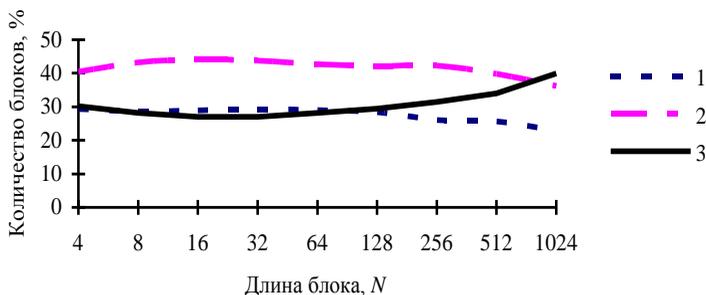


Рис. 2. Процентное содержание блоков вида 1, 2, 3 в зависимости от его длины

Данные, представленные на рис. 2 получены путем статистического подсчета наличия блоков согласно предложенной классификации в реальных речевых сигналах. Подсчет производился на наборе речевых фрагментов порядка 120 человек с параметрами: частота дискретизации 8 кГц, 8 бит на один отсчет, один канал (моно). Согласно полученных данных видно, что наибольшее количество блоков принадлежит к 2 классу ($w = 2$). Естественно, что чем ниже w , тем меньше количество коэффициентов трансформанты необходимо оставить для восстановления исходного сигнала. Помимо этого, значение коэффициента w можно использовать в качестве показателя качества в зависимости от требований к сжимаемому сигналу.

Наиболее выгодными с точки зрения сжатия являются блоки большой длины. Это обусловлено тем, что при большой длине блока и принадлежности его к классам 1 или 2 его сжатие более эффективно, чем для меньших блоков. Статистические данные показали, что наиболее выгодными являются блоки длиной 32, 64 и 128 отсчетов. При таких длинах блока наблюдается

наибольшее количество блоков вида 1 и 2, которые наиболее пригодны для сжатия, к тому же большая длина блока также способствует эффективному сжатию.

Применение указанного метода к реальным речевым сигналам показало, что коэффициент сжатия в среднем составляет 3,8. По сравнению с методом, предложенным в [8], полученное значение в 1,85 раза больше.

Кроме коэффициента сжатия были проведены исследования по оценке среднеквадратичного отклонения (СКО) в зависимости от длины блока. В этом случае использовался для вычислений тот же набор речевых фрагментов.

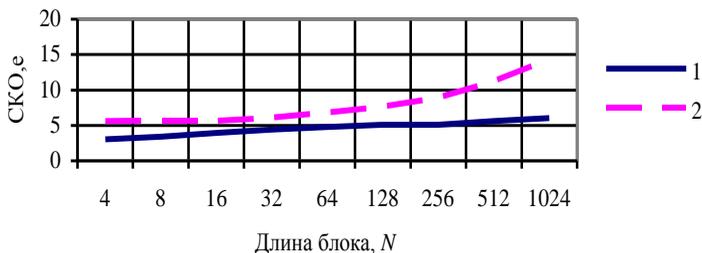


Рис. 3. Зависимость СКО от длины блока

На рис. 3 приведены значения СКО в зависимости от длины блока: 1 – предлагаемый адаптивный метод, 2 – метод, предложенный в [8]. При этом предлагаемый метод позволяет уменьшить СКО в среднем на 40% по сравнению с [8].

Выводы. Таким образом, предложенный адаптивный метод сжатия речевого сигнала на основе дискретного преобразования Уолша позволяет сжать речевой сигнал в среднем в 3,8 раза, что по сравнению с методом [8] больше в 1,85 раза. При этом СКО в среднем составляет 4,59 (1,8% от динамического диапазона сигнала), что на 40 % меньше.

Список литературы: 1. Звездин В.С. Речевое общение человека и ЭВМ. – М.: Знание, 1980. – 64 с. 2. Ахмед Н., Рао К.Р. Ортогональные преобразования при обработке цифровых сигналов. – М.: Связь, 1980. – 248 с. 3. Залманзон Л.А. Преобразование Фурье, Уолша, Хаара и их применение в управлении, связи и других областях. – М.: Наука, 1989. – 496 с. 4. Бондарев В.Н., Трестер Г. и др. Цифровая обработка сигналов: методы и средства. – Харьков: Конус, 2001. – 398 с. 5. Назаров М.В., Прохоров Ю.Н. Методы цифровой обработки и передачи речевых сигналов. – М.: Радио и связь, 1985. – 176 с. 6. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. – 495 с. 7. Полонников Р.И., Костюк В.И. и др. Матричные методы обработки сигналов. – К.: Техніка, 1977. – 136 с. 8. Данюк Ю.В., Бохан К.А. Способ сжатия речевых сигналов, основанный на дискретном преобразовании Уолша // Системы обробки інформації. – Х.: НАНУ, ПАНМ, ХВУ. – 2004. – Вип. 5(11). – С. 181 – 184.

Поступила в редакцию 14.05.04