

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ВЕЙВЛЕТНО-ПАКЕТНОГО ПРЕДСТАВЛЕНИЯ ЗА СЧЕТ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ

Кириллов С.Н., Зорин С.В.

Рязанская государственная радиотехническая академия.
390024 Рязань, ул. Гагарина, д.59, e-mail: snk@rinf.ryazan.ru

Роль алгоритмов сжатия речевых сигналов (РС) приобретает особенное значение в связи с развитием пакетной передачи речи в сетях телекоммуникаций. Известны различные методы, обеспечивающие эффективное представление РС [1], однако, в последнее время все большее внимание уделяется использованию алгоритмов вейвлетного и вейвлетно-пакетного анализа [2].

Основным достоинством вейвлетных и вейвлетно-пакетных разложений является локализация используемых базисных функций не только в частотной, но и во временной области, что необходимо при обработке нестационарных случайных процессов. При этом существуют быстрые алгоритмы вейвлетно-пакетных разложений, что определяет возможность их реализации на базе современных цифровых устройств.

Обычно [3] кодирование заключается в разбиении исходного РС $x(i)$ на блоки отсчетов $\mathbf{x}^j = [x(Mj), x(Mj+1), \dots, x(Mj+M-1)]^T$ длины M , дальнейшем вейвлетно-пакетном преобразовании, описываемом матрицей \mathbf{W}^j , векторов \mathbf{x}^j : $\mathbf{y}^j = \mathbf{W}^j \mathbf{x}^j$, с последующем кодированием спектральных коэффициентов \mathbf{y}^j , которые вместе с вспомогательной информацией о структуре матрицы \mathbf{W}^j поступают в линию передачи. Матрица вейвлетно-пакетного преобразования \mathbf{W}^j может иметь различную структуру в зависимости от корреляционных характеристик вектора \mathbf{x}^j . При этом известны [4] алгоритмы, позволяющие находить оптимальную по какому-либо критерию матрицу разложения \mathbf{W}^j .

В качестве порождающего вейвлета в рассматриваемом алгоритме использовался вейвлет Добеши 4-го порядка [5], а критерием оптимальности базиса выступал критерий минимума энтропии. Библиотека вейвлетно-пакетных базисов была ограничена глубинами разложения 3 и 4, т.е. минимальное число полос, на которые разбивался сигнал составляло 8, а максимальное 16. Такое ограничение не приводит к значительному ухудшению качества восстановленного РС, но зато резко снижает объем информации о структуре вейвлетно-пакетного дерева, которую необходимо передавать вместе со спектральными отсчетами. При этом эксперименты показали [2], что достаточно передавать только четверть спектральных коэффициентов. Из передаваемых спектральных коэффициентов 50% кодировалось с помощью трех бит, а оставшиеся с помощью двух. Длительность сегмента была выбрана равной 32 секундам, что соответствует 256 отсчетам при частоте дискретизации 8 кГц. Можно показать, что для кодирования информации о структуре дерева и передаваемых спектральных отсчетах достаточно 12 бит. В этом случае скорость передачи будет составлять 5.4 кбит/с, обеспечивая качество речи 3.5 балла в соответствии с ГОСТ Р50840-95.

Стандартные алгоритмы обеспечивают достаточно полное использование корреляционной структуры сигнала внутри сегмента, однако при этом не используются корреляционные связи между отсчетами сигнала, принадлежащими различным сегментам.

В работе предлагается использовать предварительную обработки сигнала, уменьшающую корреляционные взаимосвязи между отсчетами РС, принадлежащими различным сегментам, и только после этого осуществлять стандартное вейвлетно-пакетное преобразование.

Обозначим через \mathbf{x}_{right} и \mathbf{x}_{left} векторы, состоящие соответственно из первой и второй половины элементов произвольного вектора \mathbf{x} . Также для предварительно обработанного сигнала введем обозначение \mathbf{d}^j .

Предлагается использовать предварительное преобразование вида: $\mathbf{d}_{left}^j = \mathbf{x}_{left}^j - \hat{\mathbf{x}}_{left}^j = \mathbf{x}_{left}^j - \mathbf{P}^j \mathbf{x}_{right}^{j-1}$, где \mathbf{P}^j – матрица предсказания размерности $M/2 \times M/2$, \mathbf{d}_{left}^j – разностный сигнал, $\hat{\mathbf{x}}_{left}^j$ – предсказанное значение сигнала. Необходимо выбрать матрицу \mathbf{P} таким образом, чтобы обеспечить отсутствие корреляционных взаимосвязей между векторами \mathbf{d}_{left}^j и $\mathbf{x}_{right}^{j-1} = \mathbf{d}_{right}^{j-1}$, т.е. $E[\mathbf{d}_{left}^j \mathbf{x}_{right}^{j-1T}] = 0$. Можно показать, что такая матрица может быть

найдена как: $\mathbf{P}^j = \mathbf{R}_2^j \mathbf{R}_1^{j-1}$, где $\mathbf{R}_1^j = E[\mathbf{x}_{right}^j \mathbf{x}_{right}^{jT}]$, $\mathbf{R}_2^j = E[\mathbf{x}_{right}^{j-1} \mathbf{x}_{left}^j T]$. Такая матрица \mathbf{P}^j является симметричной и обеспечивает минимум дисперсии $E[\mathbf{d}^{jT} \mathbf{d}^j]$ разностного сигнала \mathbf{d}^j .

Несмотря на то, что использование метода предварительной обработки, основанного только на предсказании, позволяет декоррелировать векторы \mathbf{x}_{right}^j и \mathbf{x}_{left}^j , его недостатком является то, что корреляционная структура подвекторов \mathbf{d}_{left}^j и $\mathbf{x}_{right}^j = \mathbf{d}_{right}^j$ сильно отличается друг от друга.

По этой причине предлагается вектор \mathbf{d}^j подвергнуть дополнительной обработке $\mathbf{d}_{left}^j = \mathbf{H}^j \mathbf{d}_{left}^j$, которая заключается в обеспечении равенства корреляционных матриц векторов \mathbf{d}_{left}^j и \mathbf{d}_{right}^j .

Такая матрица может быть найдена как $\mathbf{H}^j = \mathbf{T}_1^j \mathbf{T}_d^{j-1}$, где \mathbf{T}_1^j и \mathbf{T}_d^j – нижние треугольные матрицы в разложении Холецкого [6] матриц $\mathbf{R}_1^j = E[\mathbf{d}_{right}^j \mathbf{d}_{right}^{jT}] = \mathbf{T}_1^j \mathbf{T}_1^{jT}$ и $\mathbf{R}_d^j = E[\mathbf{d}_{left}^j \mathbf{d}_{left}^{jT}] = \mathbf{T}_d^j \mathbf{T}_d^{jT}$.

Экспериментальные исследования проводились для фонетически сбалансированных фраз, рекомендованных ГОСТ Р50840-95. При этом, если классический вариант вейвлетно-пакетного преобразования обеспечивал качество восстановленной речи 3.5 балла, то при той же скорости передачи, предварительная обработка РС на основе алгоритма предсказания позволила повысить качество речи до 3.9 балла. Дальнейшая обработка разностного сигнала дополнительно привела к улучшению качества восстановленного сигнала до 4.3 балла.

Библиографический список

1. Назаров М.В., Прохоров Ю.Н. Методы цифровой обработки и передачи речевых сигналов. М.: Радио и связь, 1985. 176 с.
2. Кириллов С.Н., Зорин С.В. Применение алгоритмов вейвлет-анализа для сжатия речевых сигналов в IP-телефонии// Электросвязь. 2001. №4. С. 40-42.
3. Coifman, Y. Meyer, S. Quake, AND M.V. Wickerhauser. Signal processing and compression with wave packets, in Proceedings of the International Conference on Wavelets, Marseille, 1989, Y. Meyer, ed., Masson, Paris.
4. Wickerhauser M.V. Lectures on wavelet packet algorithms, Dept. of Mathematics, Washington University, St. Louis, Nov. 1991.
5. Daubechies I. Orthonormal bases of compactly supported wavelets – Commun. On Pure and Appl. Math/ 1988 XLI. P. 909.
6. Гантмахер Ф.Р. Теория матриц. М.: Государственное издательство технико-теоретической литературы. 1954. 576 с.

THE INCREASING OF EFFECTIVENESS OF WAVELET PACKET REPRESENTATION AT A EXPENSE OF PRELIMINARY PROCESSING OF SPEECH SIGNAL

Kirillov S., Zorin S.

The role of speech signal compression get a particular importance in connection with development of telecommunication technologies. Now especial attention is given to algorithms on compression on a basis wavelet and wavelet packet expansion. The main advantage of wavelet and wavelet packet expansion consists in localization of basis functions in both time and frequency. Furthermore there are fast algorithms of wavelet packed transformation, so there is a possibility of realization these algorithms on a basis of contemporary digital devises.

In the proposed algorithm original speech signal $x(i)$ first is chopped into frames $\mathbf{x}^j = [x(Mj), x(Mj+1), \dots, x(Mj+M-1)]^T$, containing 256 samples. Further we get vector of wavelet packet coefficients using a matrix $\mathbf{y}^j = \mathbf{W}^j \mathbf{x}^j$. After this we code spectral coefficients. Encoded coefficients with some additional information about structure of transformation matrix are transmitted into line. Matrix \mathbf{W}^j can have a different structure depending on a correlation characteristics of treated frame.

The proposed algorithm used Daubechie's wavelet IV-order. The library of wavelet packets is limited by third and fourth levels. That is the original signal can be splitted into minimum eight and maximum sixteen frequency bands. The quarter of spectral coefficients is transmitted into line, one half of which is coding by three bits and other half by two bits. To transmit the additional information about structure of matrix \mathbf{W}^j we need twelve bits. The duration of segment is chosen equal 30 ms, this correspond to 256 samples in frame on the sampling rate 8 kHz. In this case the bit rate is equal to 5.4 bits par second.

We propose the additional processing of original signal before a wavelet packet expansion. The additional processing consists in a diminution of correlation between samples of speech signal, which belong to adjacent blocs of samples.

Denote by \mathbf{x}_{right} and \mathbf{x}_{left} vectors, which contain respectively the first and second half of vectors \mathbf{x} elements. Also denote by \mathbf{d}^j differential vectors signals. We propose the preliminary processing as $\mathbf{d}_{left}^j = \mathbf{x}_{left}^j - \hat{\mathbf{x}}_{left}^j = \mathbf{x}_{left}^j - \mathbf{P}^j \mathbf{x}_{right}^{j-1}$, where $\hat{\mathbf{x}}_{left}^j$ is a predicted signal, \mathbf{P}^j is a prediction matrix. If we introduce correlation matrix \mathbf{R}_1^j and \mathbf{R}_2^j as $\mathbf{R}_1^j = E[\mathbf{x}_{right}^j \mathbf{x}_{right}^{jT}]$ and $\mathbf{R}_2^j = E[\mathbf{x}_{right}^{j-1} \mathbf{x}_{left}^j]^T$, then prediction matrix that minimizes a variance $E[\mathbf{d}^{jT} \mathbf{d}^j]$ of differential vector \mathbf{d}^j can be found as $\mathbf{P}^j = \mathbf{R}_2^j \mathbf{R}_1^{j-1}$. Let's mark, that the prediction matrix ensures the absence of correlation between vectors \mathbf{d}_{left}^j and $\mathbf{x}_{right}^{j-1} = \mathbf{d}_{right}^{j-1}$, that is $E[\mathbf{d}_{left}^j \mathbf{x}_{right}^{j-1T}] = 0$.

The method of prediction has a disadvantage consisting in great difference between correlation structure of subvectors \mathbf{d}_{left}^j and $\mathbf{x}_{right}^j = \mathbf{d}_{right}^j$. We propose also an additional processing $\mathbf{d}_{left}^j = \mathbf{H}^j \mathbf{d}_{left}^j$, insuring the equality $E[\mathbf{d}_{left}^j \mathbf{d}_{left}^{jT}] = E[\mathbf{d}_{right}^j \mathbf{d}_{right}^{jT}]$. A matrix \mathbf{H}^j can be found as $\mathbf{H}^j = \mathbf{T}_1^j \mathbf{T}_d^{j-1}$, where \mathbf{T}_1^j and \mathbf{T}_d^j low triangular matrixes from Cholesky decomposition of matrixes $\mathbf{R}_1^j = E[\mathbf{d}_{right}^j \mathbf{d}_{right}^{jT}] = \mathbf{T}_1^j \mathbf{T}_1^{jT}$ and $\mathbf{R}_d^j = E[\mathbf{d}_{left}^j \mathbf{d}_{left}^{jT}] = \mathbf{T}_d^j \mathbf{T}_d^{jT}$.

Summing up, the algorithm of speech signals coding on the bases of wavelet packet representation with preliminary treatment is offered. This preliminary processing ensures a diminution of correlation between signal samples, which belong to adjacent blocs of samples. Thus, the classical variant of wavelet packet transformation ensured quality of restored speech about 3.5 number on a bite rate 5.4 kbit/s. The offered preliminary processing on the bases of prediction algorithm has allowed to increase quality of speech up to 3.9 numbers. The further treatment of a difference signal improved a quality of restored signal up to 4.3 numbers.