

Субполосная обработка сигналов: эффективность и применение в речевых технологиях



А.А. Петровский,
кандидат технических наук, доцент

Приводится краткое введение в банки цифровых фильтров, даются основные определения и понятия: полное восстановление или совершенная реконструкция, параунитарный банк, субполосное кодирование, скорость передачи. Рассматривается общий случай схемы субполосного кодера. Выводятся оценки ошибки реконструкции сигнала, а также оптимальное распределение бит по каналам. Доказывается эффективность субполосного кодирования по отношению к полнополосному кодированию. Показано, что если входной сигнал имеет нормальный закон распределения, коэффициенты децимации в каналах равны, банк фильтров — ортогональный, то субполосный кодер обеспечивает равное или лучшее качество по сравнению с полнополосным кодером при любом входном сигнале. Эффективность применения субполосной обработки речи показана на примере системы редактирования шума и кодирования речевого сигнала в вэйвлет области.

1. Введение в банки фильтров

1.1. Определения

Банк фильтров — цифровая система, состоящая из секций анализа и синтеза, называемых банком фильтров анализа и банком фильтров синтеза (рис. 1). Входной сигнал $x(n)$, представленный последовательностью отсчетов, разбивается при помощи фильтров секции анализа $H_k(z)$ ($k=0, 1, \dots, M-1$) на M субполосных составляющих, которые в идеальном случае в частотной области не перекрываются. Операции, выполняемые секцией синтеза, являются обратными операциями секции анализа. Подобранным соответствующим образом набор фильтров секции синтеза $Fk(z)$ ($k=0, 1, \dots, M-1$), можно восстановить исходный сигнал $y(n)$ из его субполосных компонент [1].

Банк фильтров относится к классу многоскоростных систем цифровой обработки сигналов [1–4], в которых частота дискретизации различна в разных точках системы.

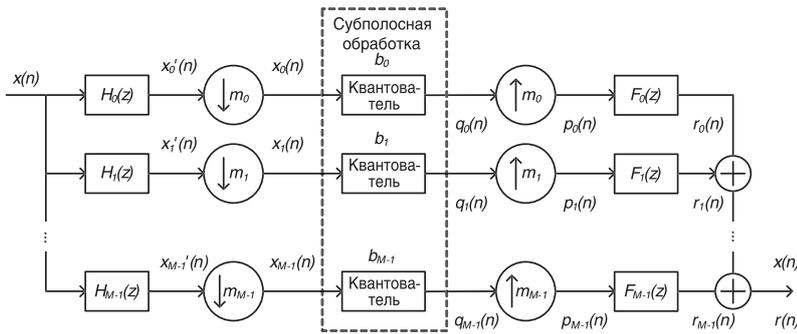


Рис. 1. Банк фильтров: система анализа/синтеза сигнала

Значение коэффициента темпа поступления отсчетов в канале (коэффициента децимации m_k) задаётся утверждением о дискретизации сигнала в зависимости от ширины частотной полосы канала B_k и его положения в банке фильтров. Оригинальная частота дискретизации f_s сигнала в k -м канале теоретически может быть уменьшена в $m_k \leq f_s/B_k$ раз. Равенство в данном случае означает, что канал максимально децимирован:

$$\sum_{k=0}^{M-1} \frac{1}{m_k} = 1. \quad (1.1)$$

Таким образом, в частотно-временном представлении сигнала исключена избыточность, т.е. сумма отсчётов во всех каналах соответствует количеству отсчётов в исходном сигнале. Банк фильтров считается передецимированным, если $\sum_{k=0}^{M-1} \frac{1}{m_k} > 1$, т.е. достаточно, чтобы хотя бы в одном канале коэффициент децимации не был равен числу каналов. Более сложные случаи позиционирования субполос каналов в банках фильтров рассмотрены в [4–6].

Соотношение между z -преобразованием сигналов на входе и выходе M -канального банка фильтров, изображённого на рис. 1, следующее [1]:

$$Y(z) = \sum_{k=0}^{M-1} F_k(z) \frac{1}{m_k} \sum_{l=0}^{m_k-1} H_k(zW_{m_k}^l) X(zW_{m_k}^l) \quad (1.2)$$

где $W_{m_k} = e^{-j2\pi/m_k}$. Анализ данного выражения показывает, что в банке возможны искажения входного сигнала: амплитудные, фазовые и отражения частотных характеристик (элайзинг), обусловленные наличием дециматоров и интерполяторов. Величина искажений зависит как от частотных характеристик канальных фильтров, так и выбора коэффициентов децимации m_k . Анализ искажений, возникающих в структуре банка фильтров, проще анализировать в максимально децимированном банке фильтров, для которого соотношение (1.2) значительно упрощается:

$$Y(z) = \sum_{k=0}^{M-1} T_k(z) X(zW_M^k), \quad (1.3)$$

где $T_k(z)$ — передаточная функция k -го канала:

$$T_k(z) = \frac{1}{M} \sum_{l=0}^{M-1} F_l(z) H_l(zW_M^k). \quad (1.4)$$

Выходной сигнал $y(n)$ системы анализа-синтеза банка фильтров будет свободен от элайзинговой составляющей $X(zW_M^k)$, $k > 0$ в случае, если

$$T_k(z) = 0, \text{ для } 1 \leq k \leq M. \quad (1.5)$$

В банке фильтров, для которого справедливо условие (1.5), остаются только амплитудные и фазовые искажения, которые определяются согласно следующему выражению:

$$\frac{Y(z)}{X(z)} = T_0(z) = \frac{1}{M} \sum_{l=0}^{M-1} F_k(z) H_k(z). \quad (1.6)$$

Очевидно, что для получения перфективной реконструкции входного сигнала $x(n)$ банком фильтров, передаточная функция искажений $T_0(z)$ должна принять форму простого звена задержки с некоторым масштабированием амплитуды:

$$T_0(z) = cz^{-\Delta}, c \neq 0, \Delta \in \mathbb{Z}. \quad (1.7)$$

Полное восстановление или перфективная реконструкция — свойство банка цифровых фильтров, заключающееся в том, что сигнал, прошедший через схему анализа-синтеза, идентичен входному с точностью до задержки. Для этого фильтры синтеза должны подавлять наложение частотных характеристик (элайзинг) и устранять амплитудные и фазовые искажения [1].

Параунитарный (ортогональный) банк фильтров (ПУБФ) — банк фильтров, у которого передаточные функции анализирующих и синтезирующих фильтров и их соответственно смещенные версии ортогональны друг другу. Фильтры синтеза в параунитарных банках являются транспонированными версиями фильтров анализа [1]:

$$F_k(z) = H_k^T(z^{-1}). \quad (1.8)$$

При соблюдении этого условия обеспечивается возможность перфективной реконструкции банком фильтров входного сигнала $x(n)$ пусть $x[n] = [x_0[n] \dots x_{M-1}[n]]^T$ будет входным вектором и $y[n] = [y_0[n] \dots y_{M-1}[n]]^T$ будет соответствующим выходным вектором с $M \times N$ параунитарной передаточной матрицей $A(z)$. Пусть $S_{xx}(e^{jw})$ будет $M \times M$ СПМ-матрица входного вектора $x[n]$. Заметим, что СПМ i -го входного компонента $x_i[n]$ является i -й элемент $S_{xx}(e^{jw})$. Поэтому дисперсия $x_i[n]$ составит

$$\int_{-\pi}^{\pi} (S_{xx}(e^{jw}))_i \frac{dw}{2\pi}. \quad (1.9)$$

Усредненная дисперсия входного сигнала будет

$$\frac{1}{N} \sum_{i=0}^{N-1} \sigma_i^2 = \frac{1}{N} \int_{-\pi}^{\pi} \text{tr} (S_{xx}(e^{jw})) \frac{dw}{2\pi}, \quad (1.10)$$

где $\text{tr}(\mathbf{B})$ — след матрицы \mathbf{B} .

СПМ вектора на выходе определяется как

$$S_{yy}(e^{jw}) = A(e^{jw}) S_{xx}(e^{jw}) A^H(e^{jw}), \quad (1.11)$$



где A^H — матрица, эрмитово транспонированная к матрице A .

Усредненная дисперсия выхода составит

$$\frac{1}{N} \int_{-\pi}^{\pi} \text{tr}(S_{yy}(e^{jw})) \frac{dw}{2\pi} = \frac{1}{N} \int_{-\pi}^{\pi} \text{tr}(A(e^{jw})S_{xx}(e^{jw})A^H(e^{jw})) \frac{dw}{2\pi}. \quad (1.12)$$

Так как $\text{tr}(AB) = \text{tr}(BA)$, это упрощает выражение (1.12)

$$\frac{1}{N} \int_{-\pi}^{\pi} \text{tr}(S_{xx}(e^{jw})A^H(e^{jw})A(e^{jw})) \frac{dw}{2\pi}.$$

Так как матрица $A(z)$ — параунитарная, $A^H(e^{jw})A(e^{jw}) = I$, следовательно

$$\frac{1}{N} \int_{-\pi}^{\pi} \text{tr}(S_{xx}(e^{jw})) \frac{dw}{2\pi} \quad (1.13)$$

или усреднённая дисперсия выхода равна усреднённой дисперсии входа. Этот факт показывает, что в параунитарной системе энергия сохраняется.

Банк фильтров можно представить в полифазной форме, если передаточные функции секций анализа и синтеза записать в виде соответствующих векторов:

$$\begin{aligned} \mathbf{H}(z) &= [H_0(z) H_1(z) \dots H_{M-1}(z)]^T, \\ \mathbf{F}(z) &= [F_0(z) F_1(z) \dots F_{M-1}(z)]^T, \end{aligned} \quad (1.14)$$

то тогда можно выбрать такие полифазные матрицы анализа

$$\mathbf{E}(z) = \begin{bmatrix} E_{0,0}(z) & \dots & E_{0,M-1}(z) \\ \vdots & \ddots & \vdots \\ E_{M-1,0}(z) & \dots & E_{M-1,M-1}(z) \end{bmatrix} \quad (1.15)$$

и синтеза

$$\mathbf{D}(z) = \begin{bmatrix} D_{0,0}(z) & \dots & D_{0,M-1}(z) \\ \vdots & \ddots & \vdots \\ D_{M-1,0}(z) & \dots & D_{M-1,M-1}(z) \end{bmatrix}, \quad (1.16)$$

Вектора передаточных функций секций анализа и синтеза можно представить следующим образом [1]:

$$\begin{aligned} \mathbf{H}(z) &= \mathbf{E}(z^M) [1 \ z^1 \ \dots \ z^{-(M-1)}]^T, \\ \mathbf{F}(z) &= [z^{-(M-1)} \ z^{-(M-2)} \ \dots \ 1]^T \mathbf{D}(z^M). \end{aligned} \quad (1.17)$$

Для получения перфективной реконструкции на компоненты полифазных матриц накладывается дополнительное ограничение:

$$\mathbf{D}(z)\mathbf{E}(z) = cz^{-\Delta}\mathbf{I}, \quad c \neq 0, \Delta \geq 0, \quad (1.18)$$

где c — ненулевая константа; Δ — задержка, выраженная целым числом интервалов дискретизации, вносимая секциями анализа-синтеза; \mathbf{I} — единичная матрица. На рис. 2 показана полифазная структура банка фильтров.

Традиционно банки фильтров разделяют на банки с равнополосными и неравнополосными каналами, ортогональные, биортогональные, двухканальные и многоканальные и т.д. Каждый фильтр банка цифровых фильтров образует канал. Поэтому говорят об M -канальном банке фильтров. Сигнал в канале называется субполосой, отсюда название «субполосная фильтрация» или «субполосное кодирование» [3, 4].

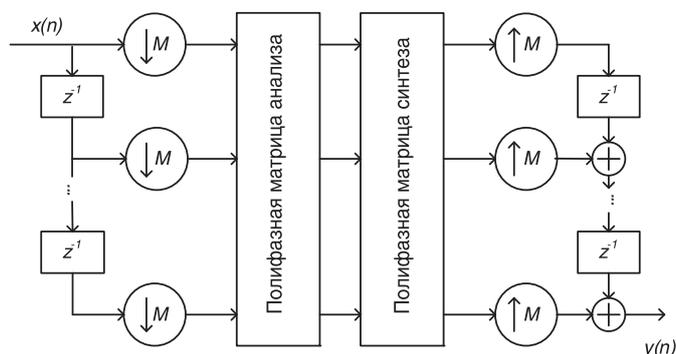


Рис. 2. Полифазная структура банка фильтров

1.2. Ошибка реконструкции сигнала

Рассматривается субполосный кодер, представленный на рис. 1.1, состоящий из $k=0, \dots, M-1$ каналов. Каждый канал имеет анализирующий фильтр $H_k(z)$, синтезирующий фильтр $F_k(z)$, дециматор/интерполятор с коэффициентом передискретизации m_k и b_k -битный квантователь. Входной сигнал $x(n)$ после фильтрации анализирующими фильтрами на рис. 1.1 обозначен как $x'(n)$, а каналные сигналы — $x_k(n)$. Субполосный кодер максимально децимирован, если выполняется условие (1.1).

Пусть входной сигнал $x(n)$ стационарный в широком смысле и имеет среднее значение, равное нулю. Следовательно, все последующие (производные) сигналы (включая шум квантования) также будут иметь среднее значение, равное нулю. Это предположение не является причиной возникновения каких-либо трудностей для применения субполосного кодирования к сигналам со средним значением, отличным от нуля, таких как изображения, так как система в этом случае эквивалентна системе с нулевым средним, полученной путем вычитания среднего значения из входного сигнала [7].

Следовательно, если $S_{xx}(e^{j\omega})$ является спектральной плотностью мощности (СПМ) сигнала $x(n)$, то его дисперсия определяется как [1]

$$\sigma_x^2 = \int_{-\pi}^{\pi} S_{xx}(e^{j\omega}) \frac{\partial \omega}{2\pi}.$$

Дисперсия сигнала $x'_k(n)$ на выходе анализирующего фильтра $H_k(z)$ будет определяться следующим выражением [1]

$$\sigma_k^2 = \int_{-\pi}^{\pi} S_{xx}(e^{j\omega}) |H_k(e^{j\omega})|^2 \frac{\partial \omega}{2\pi}$$

для $k=0, 1, \dots, M-1$. Так как сигнал стационарен в широком смысле, то дисперсия не изменится после децимации [1,7], т.е. дисперсия $x_k(n)$ субполосы k равна σ_k^2 . Квантователь представляется как модель с аддитивным шумом [8], т.е. выходом квантователя является сигнал $x_k(n) + q_k(n)$, где $x_k(n)$ — входной сигнал и $q_k(n)$ — шум квантования. Дисперсия шума квантования, как для равномерных квантователей, так и для квантователей, адаптированных под сигнал, определяется как [1,8]



$$\sigma_{q_k}^2 = \varepsilon_k 2^{-2b_k} \sigma_k^2, \quad (1.19)$$

где ε_k — константа, характеризующая квантователь, которая зависит от функции плотности вероятности k -го субполосного сигнала.

Предположим, используются равномерные (многобитные) квантователи, тогда шумы квантования $q_k(n)$ различных каналов некоррелированы между собой и являются «белым» шумом [9]. СПМ шума $q_k(n)$ является константой $\sigma_{q_k}^2$, так как это «белый» шум, и имеют нулевое среднее. Сигналы $p_k(n)$, полученные после интерполяции $q_k(n)$, определяются как

$$p_k(n) = \begin{cases} q_k(n/m_k), & \text{если } n \bmod m_k = 0 \\ 0 & \text{иначе} \end{cases},$$

и больше не являются стационарными в широком смысле, но имеют циклическую стационарность с периодом m_k [7]. Другими словами, если $n \bmod m_k = 0$, то СПМ сигнала $p_k(n)$ равна $\sigma_{q_k}^2$, а СПМ оставшихся отсчетов равна нулю. Отсчеты шума $p_k(n)$ поступают на вход синтезирующего фильтра $F_k(n)$, на выходе которого получается сигнал $r_k(n)$, стационарный в широком смысле с периодом m_k . Усредненная на периоде m_k дисперсия сигнала $r_k(n)$ определяется как

$$\frac{1}{m_k} \left[\int_{-\pi}^{\pi} \sigma_{q_k}^2 |F_k(e^{j\omega})|^2 \frac{\partial \omega}{2\pi} + 0 + \dots + 0 \right],$$

где $m_k - 1$ — нули, полученные из нулевых отсчетов сигнала шума $p_k(n)$. Ошибка реконструкции $r(n)$ равна сумме всех ошибок $r_k(n)$, что обусловлено линейностью секции синтеза банка фильтров. Так как $r_k(n)$ во всех каналах некоррелированы, то дисперсия их суммы равна сумме их дисперсий [10]

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\sigma_{q_k}^2}{m_k} \int_{-\pi}^{\pi} |F_k(e^{j\omega})|^2 \frac{\partial \omega}{2\pi}, \quad (1.20)$$

где σ_r^2 — дисперсия ошибки реконструкции. Пусть $n_k = \int_{-\pi}^{\pi} |F_k(e^{j\omega})|^2 \frac{\partial \omega}{2\pi}$ означает нормы синтезирующих фильтров (заметим, что для КИХ-фильтра с коэффициентами импульсной характеристики (h_0, h_1, \dots, h_L) норма равна $(h_0^2 + h_1^2 + \dots + h_L^2)$). Затем, используя уравнение (1.19), выражение (1.20) преобразуется к следующему виду

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\varepsilon_k 2^{-2b_k} \sigma_k^2 n_k}{m_k}. \quad (1.21)$$

Для параунитарных систем [1] свойство отсутствия потерь (или энергетический баланс) говорит о том, что энергия выхода равна энергии входа (1.13). Вследствие этого факта, дисперсия ошибки реконструкции сигнала банком фильтров определяется следующим выражением:

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\varepsilon_k 2^{-2b_k} \sigma_k^2}{m_k}. \quad (1.22)$$

Сравнивая уравнения (1.21) и (1.22) для ортогонального случая можем заметить, что в (1.22) отсутствует норма синтезирующего фильтра n_k . Это является следствием свойства отсутствия потерь, подразумевающего фильтры с единичной энергией или $n_k = 1$ для всех k [1]. Следует отметить также, что допущение об ортогональном банке фильтров заменяет допущение о равномерных (многобитных) квантователях и не обязательно для получения данного результата. Поэтому, в смысле квантователя, результат более общий для ортогонального случая.

1.3. Скорость передачи субполосного кодера

Для субполосного кодера на рис. 1 b_k определяет количество бит на отсчет для k -го квантователя. Однако благодаря коэффициенту децимации m_k данным квантователем квантуется один отсчет для каждого из m входных отсчетов. Поэтому скорость передачи квантователя равна b_k/m_k бит на входной отсчет. Следовательно, усредненная скорость передачи субполосной системы можно определить как

$$b = \sum_{k=0}^{M-1} \frac{b_k}{m_k} \text{ бит/отсчет.} \quad (1.23)$$

Полнополосный кодер просто квантует вход $x(n)$, используя b -битный квантователь. Поэтому из уравнения (1.22) его дисперсия ошибки квантования (которая также является дисперсией ошибки реконструкции) определяется как

$$\sigma_q^2 = \varepsilon 2^{-2b} \sigma_x^2, \quad (1.24)$$

где ε определяется как и ранее.

Эффективность субполосного кодирования вычисляется как отношение дисперсии ошибки реконструкции полнополосного кодера (или импульсно-кодовой модуляции — ИКМ) σ_q^2 к дисперсии ошибки реконструкции субполосного кодера σ_r^2 с аналогичной скоростью передачи данных [8].

2. Оценка эффективности субполосного кодирования

2.1. Оптимальное распределение бит по каналам субполосного кодера

Проблема оптимального распределения бит в каналах состоит в нахождении b_0, \dots, b_{M-1} , которые минимизируют дисперсию ошибки реконструкции σ_r^2 в уравнении (1.21), удовлетворяя ограничению (1.23). Данная минимизация с ограничением (1.23) может быть решена с использованием метода множителей Лагранжа. Для этого определяется целевая функция

$$C = \sum_{k=0}^{M-1} \frac{\varepsilon_k 2^{-2b_k} \sigma_k^2 n_k}{m_k} + \lambda \left(\sum_{k=0}^{M-1} \frac{b_k}{m_k} - b \right),$$

где λ — множитель Лагранжа. Дифференцируя C по b_k и приравнявая к нулю, ($dC/db_k = 0$) или

$$\frac{1}{m_k} \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k (-2 \ln 2) + \lambda \frac{1}{m_k} = 0,$$



из которого следует, что

$$\frac{\partial}{\partial b_k} 2^{-2b_k} = \frac{\partial}{\partial b_k} e^{-2 \ln 2 b_k} = -2 \ln 2 e^{-2 \ln 2 b_k}.$$

Таким образом,

$$2 \ln 2 \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k = \lambda,$$

или

$$2^{-2b_k} = \frac{\lambda}{2 \ln 2 \varepsilon_k \sigma_k^2 n_k}. \quad (2.1)$$

После логарифмирования по основанию 2 обеих частей (1.25) получается, что

$$-2b_k = \log_2 \frac{\lambda}{2 \ln 2 \varepsilon_k \sigma_k^2 n_k},$$

или

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\varepsilon_k 2^{-2b_k} \sigma_k^2}{m_k}. \quad (2.2)$$

Данное выражение действительно для всех значений $k = 0, \dots, M - 1$. Подставляя уравнение (2.2) в ограничение (1.23), получается, что

$$\begin{aligned} b &= \sum_{k=0}^{M-1} \frac{1}{m_k} \left[\frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} + \frac{1}{2} \log_2 (\varepsilon_k \sigma_k^2 n_k) \right] = \\ &= \frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} \sum_{k=0}^{M-1} \frac{1}{m_k} + \frac{1}{2} \sum_{k=0}^{M-1} \log_2 (\varepsilon_k \sigma_k^2 n_k)^{\frac{1}{m_k}}. \end{aligned} \quad (2.3)$$

Первая сумма здесь равна 1 на основании уравнения (1.16). Вторая сумма логарифмов может быть записана как логарифм произведения. Используя эти упрощения, средняя скорость передачи в субполосной системе равна

$$b = \frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} + \frac{1}{2} \log_2 \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}, \quad (2.4)$$

или

$$\frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} = b - \frac{1}{2} \log_2 \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}. \quad (2.5)$$

Подставляя данный результат в уравнение (2.2), получается следующее распределение бит по каналам:

$$\begin{aligned} b_k &= b - \frac{1}{2} \log_2 \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}} + \frac{1}{2} \log_2 (\varepsilon_k \sigma_k^2 n_k) = \\ &= b + \frac{1}{2} \log_2 \frac{\varepsilon_k \sigma_k^2 n_k}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}}, \end{aligned} \quad (2.6)$$

для всех значений $k = 0, \dots, M - 1$.

Это и есть оптимальное распределение бит по каналам субполосного кодера. Заметим, что для ортогонального случая оптимальное распределение бит получается путем подстановки $n_k = 1$ для всех k . Результирующая скорость передачи — действительная величина и может быть отрицательной.

2.2. Дисперсия минимальной ошибки реконструкции сигнала

На основании оптимального распределения бит (уравнение (2.6)) следует, что

$$2^{-2b_k} = 2^{-2b} \frac{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}}{\varepsilon_k \sigma_k^2 n_k},$$

для всех $k = 0, \dots, M-1$. Подставляя это выражение в уравнение (1.22) и принимая во внимание (1.1), получается, что дисперсия минимальной ошибки реконструкции сигнала равна

$$\begin{aligned} \sigma_r^2 &= \sum_{k=0}^{M-1} \frac{1}{m_k} 2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}} = 2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}} \sum_{k=0}^{M-1} \frac{1}{m_k} = \\ &= 2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}, \end{aligned} \quad (2.7)$$

2.3. Оценки эффективности субполосного кодирования

Эффективность субполосного кодирования можно определить как отношение дисперсии ошибки реконструкции полнополосного кодера σ_q^2 (1.24) к дисперсии ошибки реконструкции субполосного кодера σ_r^2 (1.31) с аналогичной скоростью передачи данных

$$G = \frac{\varepsilon 2^{-2b} \sigma_x^2}{2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}} = \frac{\varepsilon \sigma_x^2}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}}. \quad (2.8)$$

Некоторые специальные случаи оценки эффективности субполосного кодирования рассматриваются ниже.

Предположим, что входной сигнал $x(n)$ является гауссовым (отсчеты имеют гауссову функцию плотности вероятности). Тогда известно, что выход линейной системы с гауссовым входом также является гауссовым [10]. Таким образом, все субполосные сигналы будут гауссовыми, что приведет к $\varepsilon = \varepsilon_o = \varepsilon_r = \dots \varepsilon_{M-1}$ (при условии многобитного квантователя). Следовательно, эффективность кодирования составит

$$G = \frac{\varepsilon \sigma_x^2}{\left(\varepsilon^{\frac{1}{m_0} + \dots + \frac{1}{m_{M-1}}} \right) \prod_{i=0}^{M-1} (\sigma_i^2 n_i)^{\frac{1}{m_i}}} = \frac{\sigma_x^2}{\prod_{i=0}^{M-1} (\sigma_i^2 n_i)^{\frac{1}{m_i}}}.$$



Пусть коэффициенты децимации будут равны $m_0 = m_1 = \dots = m_{M-1}$. Из уравнения (1.1) следует, что каждый коэффициент равен M и эффективность субполосного кодирования составит

$$G = \frac{\varepsilon \sigma_x^2}{[\prod_{i=0}^{M-1} (\sigma_i^2 n_i)]^{\frac{1}{M}}}.$$

В случае ортогонального банка фильтров $n_k = 1$ как было обозначено ранее, эффективность субполосного кодера будет определяться как

$$G = \frac{\varepsilon \sigma_x^2}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2)^{\frac{1}{m_i}}}. \quad (2.9)$$

Далее, свойство отсутствия потерь, обозначенное ранее, может быть применено к матрице полифазного анализа (1.15). Так как входом полифазного анализа является $x(n)$, то усреднённая дисперсия входа будет равна σ_x^2 . Пусть σ_k^2 обозначает дисперсию выхода $k = 0, \dots, M - 1$ субполос, тогда на основании свойства отсутствия потерь следует, что

$$\sigma_x^2 = \sum_{k=0}^{M-1} \frac{\sigma_k^2}{m_k}. \quad (2.10)$$

Подставляя равенство (2.10) в соотношение (2.9), получаем, что эффективность субполосного кодирования может быть выражена следующим образом:

$$G = \frac{\varepsilon \sum_{i=0}^{M-1} \frac{\sigma_i^2}{m_i}}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2)^{\frac{1}{m_i}}}. \quad (2.11)$$

Анализ (2.11) показывает, что более простое выражение оценки эффективности субполосного кодирования получается при одновременном рассмотрении следующих допущений [11–15]: входной сигнал имеет нормальный закон распределения, коэффициенты децимации в каналах равны, банк фильтров — ортогональный. При этом эффективность субполосного кодирования определяется отношением среднего арифметического к среднему геометрическому неотрицательных величин σ_i^2 :

$$G = \frac{\frac{1}{M} \sum_{i=0}^{M-1} \sigma_i^2}{(\prod_{i=0}^{M-1} \sigma_i^2)^{\frac{1}{M}}}. \quad (2.12)$$

Так как среднее арифметическое больше или равно среднему геометрическому, следовательно, эффективность субполосного кодирования $G \geq 1$, т.е. субполосный кодер обеспечивает равное или лучшее качество по сравнению с полнополосным кодером при любом входном сигнале. Величина, обратная выражению (2.12), также известна как мера пологости спектра (SFM) [16,17]. Часто в задачах кодирования речи используется взвешенная на порог маскирования величина SFM (перцептуально взвешенная SFM — PSFM) [17, 18].

3. Применение в речевых технологиях субполосной обработки сигналов

3.1. Структура кодера-редактора шумов речевого сигнала

Предлагается комбинированная система редактирования шумов и кодирования речевого сигнала без специального процессора повышения качества речи на основе критического дерева пакета дискретного вейвлет преобразования (ПДВП) $CB - WPD: (l, n) \in E_{cv}, l = \overline{0, 5}$ (рис. 3) и вычисления порога восприятия речевого сигнала человеком. Разработка ориентирована на частоту дискретизации 8кГц и обработка введётся в 17 барках [19].

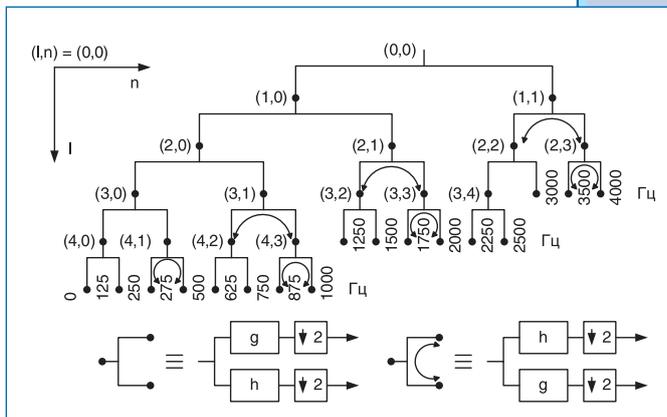


Рис. 3. Структура критического дерева ПДВП

Предполагается, что чистый речевой сигнал $x(t)$ и окружающий его шум $n(t)$ статистически независимы и стационарны в широком смысле (где t обозначает временной индекс). Зашумленный временной сигнал $y(t) = x(t) + n(t)$ преобразуется в вейвлет область на основе ПДВП. Вейвлет коэффициенты зашумленного речевого сигнала определяются следующим образом:

$$Y_{l,n}(k) = \langle y, \psi_{l,n,k} \rangle, (l, n) \in E_{CB}, k \in Z. \quad (3.1)$$

где k временной индекса вейвлет коэффициента в субполосе обработки (l, n) .

На рис. 4 показана схема обработки речевого сигнала в одной из ветвей $CB - WPD: (l, n) \in E_{cv}, l = \overline{0, 5}$ (соответствующей ей частотной полосе (рис. 3)) комбинированной системы редактирования шума и кодирования речевого сигнала. Оценка порогов маскирования выполняется в вейвлет области в соответствии с алгоритмом, показанным в [20]. В данной работе используется наиболее общее психоакустически мотивированное правило спектрального взвешивания [21]. Вейвлет коэффициенты $\hat{X}_{l,n}(k)$, отредактированные от шума входного сигнала, поступают на схему кодирования и квантования и далее формируется пакет данных

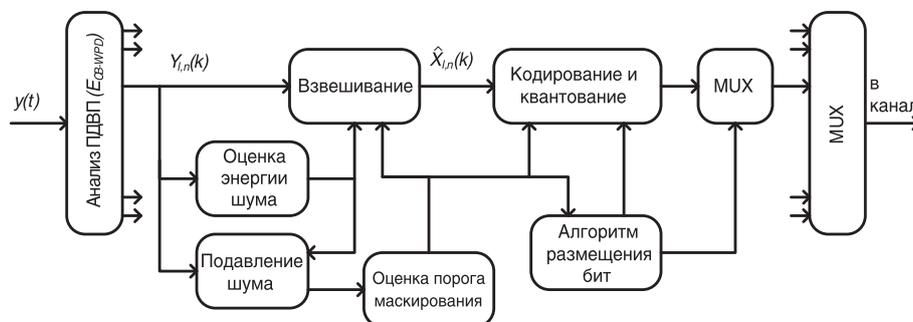


Рис. 4. Структура кодера-редактора шумов речевого сигнала на базе ПДВП $CB - WPD: (l, n) \in E_{cv}, l = \overline{0, 5}$

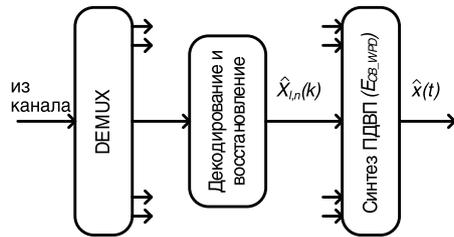


Рис. 5. Структура декодера речевого сигнала на базе ПДВП

для передачи в канал. Схема кодирования и квантования реализована в соответствии с [22].

Структура декодера показана на рис. 5, где закодированные вейвлет коэффициенты $\hat{X}_{l,n}(k)$ декодируются и восстанавливаются в каждой субполосе (l,n) . Синтез сигнала выполняется на основе обратного ПДВП в соответствии со структурой дерева $E_{SB} WPD$:

$$\hat{x}(t) = \sum_{(l,n) \in E, k \in Z} \hat{X}_{l,n}(k) \hat{\psi}_{l,n,k}(t), \quad (3.2)$$

где $\{\hat{\psi}_n(e): n \in Z_+\}$ — множество ортогональных вейвлет функций ПДВП $\{\psi_n(e): n \in Z_+\}$, где $\{\hat{\psi}_{l,n,k}(t) = 2^{-l/2} \hat{\psi}_n(2^{-l}t - k)$.

3.2. Эксперимент

Для кодера со скоростью передачи 4–6 кбит/с экспериментальные результаты приведены на рис. 6: (а) чистый речевой сигнал с частотой дискретизации $f_s = 8 \text{ кГц}$ и его спектрограмма, (б) зашумленный речевой сигнал с $SNR = 5 \text{ дБ}$ и его спектрограмма, (в) отредактированный речевой сигнал от шума и его спектрограмма, (г) реконструированный речевой сигнал декодером и его спектрограмма.

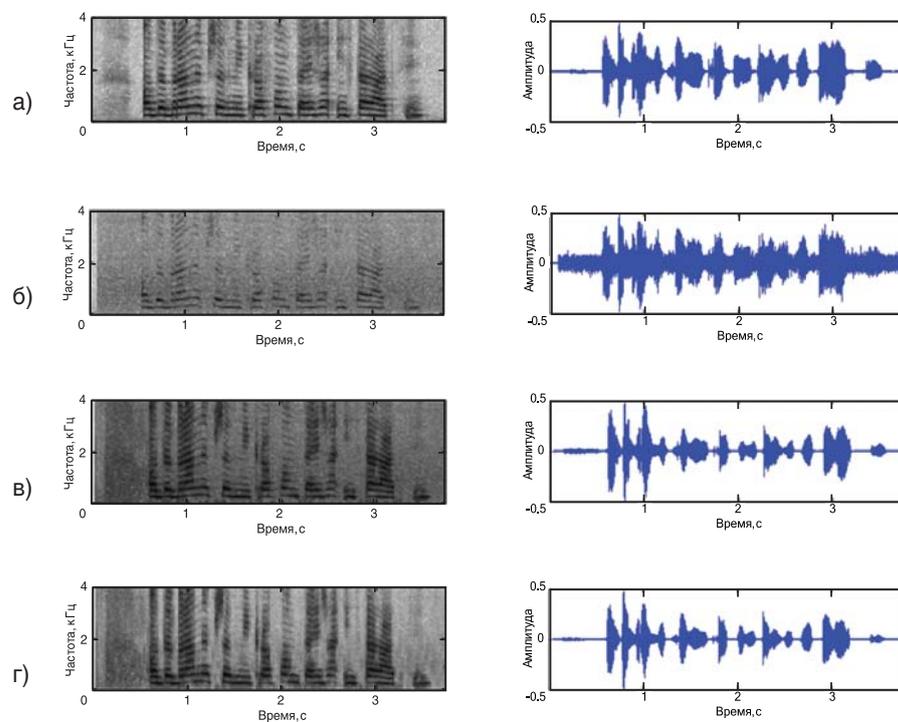


Рис. 6. Результаты обработки речевого сигнала в кодере-редакторе шумов: а) чистый речевой сигнал; б) зашумленный речевой сигнал; в) отредактированный речевой сигнал от шума; г) реконструированный речевой сигнал декодером

Достоинством данной системы субполосной обработки речевых сигналов является возможность комбинирования двух процессов перцептуальной обработки в субполосах: редактирование акустического шума в речевом сигнале и последующая его компрессия. Обработка ведётся в области вейвлет коэффициентов, причём порог маскирования рассчитывается один раз для обоих процессов.

Литература

1. Vaidynathan P.P. Multirate systems and filter banks, Prentice Hall: Englewood Cliffs, 1993.
2. Crochiere R.E., Rabiner L. Multirate digital signal processing, Prentice-Hall, Englewood Cliffs, NJ, USA, 1983.
3. Витязев В.В. Цифровая частотная селекция сигналов // Радио и связь. М., 1993.
4. Piotrowski A., Parfieniuk M. Cyfrowe banki filtrów: analiza, synteza i implementacja dla systemów multimedialnych, Politechnika Białostocka, Białystok, 2006.
5. Vaughan R.G., Scott N.L., White D.R. The theory of bandpass sampling, IEEE Trans. Signal processing, 1991. V. 39. № 9. P. 1973–1984.
6. Parfieniuk M., Petrovsky A. Simple rule of selection of subsampling ratios for warped filter banks, in Proc. VIII Int. conf. «Modern communication systems», Naroch, Belarus, 2003. P. 130–134. Special Issue of Trans. Belarussian Engineer Academy, № 1(15)/3.
7. Sathe V.P., Vaidynathan P.P. Effects of multirate systems on the statistical properties of random signals. IEEE Transactions Signal Processing, 41(1), January 1993. P. 131–146.
8. Jaynt N.S., Noll P. Digital coding of waveforms, Prentice Hall: Englewood Cliffs, 1984.
9. Uzun N., Haddad R.A. Cyclostationary modeling, analysis, and optimal compensation of quantization errors in subband coders. IEEE Transactions Signal Processing, 43(9), September 1995. P. 2109–2119.
10. Papoulis A. Probability, random variables, and stochastic process, McGraw-Hill: Tokyo, 1984.
11. Soman A.K., Vaidynathan P.P. Coding gain in paraunitary analysis/synthesis systems. IEEE Transactions Signal Processing, 41(5), May 1993. P. 1824–1835.
12. Djokovic I., Vaidynathan P.P. On optimal analysis/synthesis filters for coding gain maximization. IEEE Transactions Signal Processing, 44(5), May 1996. P. 1276–1279.
13. Calvagno G., Mian G.A., Rinaldo R. Computation of the coding gain for subband coders. IEEE Transactions Communication, 44(4), April 1996. P. 475–487.
14. Kok C.W., Nguyen T.Q. Multirate filter banks and transform coding gain. IEEE Transactions Signal Processing, 46(7), July 1998. P. 2041–2044.
15. Gosse K., Duhamel P. Perfect reconstruction versus MMSE filter banks in source coding. IEEE Transactions Signal Processing, 45(9), September 1997. P. 2188–2202.
16. Spanias A., Painter T., Atti V. Audio signal processing and coding, Wiley-Interscience, NJ, USA, 2007.
17. Bosi M., Goldberg R.E. Introduction to digital audio coding and standards, Springer Science+Business Media, USA, 2003.
18. Петровский А.А., Белявский К., Петровский Ал.А. Перцептуальное кодирование аудио и речевых сигналов: Доклады БГУИР. 2004. № 1(5). С. 73–91.
19. Petrovsky A.A., Bielawski K., Petrovsky Al.A. Combined system for acoustic echo and Noise reduction based on the psychoacoustically motivated multirate filter bank // Mittweida, Germany: IWKM, 2000, Journal of the University of Applied Sciences Mittweida. P. 33–41.
20. Петровский А. Построение психоакустической модули в области вейвлет-коэффициентов для перцептуальной обработки звуковых и речевых сигналов // Речевые технологии. 2008. № 4. С. 61–71.
21. Петровский Ал.А., Борович А., Парфенюк М. Дискретное преобразование Фурье с неравномерным частотным разрешением в перцептуальных системах редактирования шума в речи // Речевые технологии. 2008. № 3. С. 16–26.
22. Петровский Ал. Перцептуальный кодер звука на базе вейвлет преобразования с динамической трансформацией частотно-временного плана // Цифровая обработка сигналов. 2009. № 4. С. 48–58.



Петровский Алексей Александрович —

кандидат технических наук, доцент. Работает в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники», кафедра Электронных вычислительных машин.

Закончил учреждение образования «Белорусский государственный университет информатики и радиоэлектроники», специальность — «Проектирование и технология электронных вычислительных средств». Сфера интересов — цифровая обработка сигналов: многоскоростная обработка, анализ/синтез банков фильтров, проектирование проблемно-ориентированных средств вычислительной техники реального времени для систем мультимедиа. Член общества AES.