УДК 681.3.01:621.391:517.988

ПЕРЦЕПТУАЛЬНОЕ КОДИРОВАНИЕ АУДИО И РЕЧЕВЫХ СИГНАЛОВ

А.А. ПЕТРОВСКИЙ, К. БЕЛЯВСКИЙ, АЛ.А. ПЕТРОВСКИЙ

Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь

Белостокский технический университет, кафедра систем реального времени

Поступила в редакцию 15 декабря 2003

В статье предлагается новое решение в построении перцептуальных аудио и речевых кодеров на основе пакета дискретного вэйвлет преобразования (ПДВП), а также комбинированная система редактирования слышимых шумов и кодирования речевых сигналов на основе ПДВП, согласованного с психоакустической шкалой барков и перцептуального взвешивания.

Ключевые слова: кодирование, пакет дискретного вэйвлет-преобразования, психоакустика.

Введение

Непрерывное увеличение передач в системах мультимедиа через Интернет обусловливает поиск новых решений эффективной обработки в реальном масштабе времени аудио и речевых данных (их компрессию и декомпрессию) [1, 2]. Запись в память компьютера и передача высокого качества музыки (аудиосигналов) требует изучения и разработки новых, соответствующих особенностям данного сигнала методов компрессии-кодирования и архитектурных решений процессоров [1]. Компрессия речевых сигналов традиционно базируется на определенных моделях речеобразования [3], в то время как в методах компрессии высококачественного аудиосигнала пытаются использовать свойства шумового маскирования человеческого слуха [4].

Общая философия перцептуального кодера взаимосвязана с выбором метода частотно-временного анализа [5]. В настоящей статье исследуется построение перцептуальных кодеров сигналов как на основе адаптированного под сигнал и заданный вычислительный ресурс пакета дискретного вэйвлет преобразования (ПДВП) (аудио кодеры), так и на фиксированной структуре дерева ПДВП, согласованной со шкалой критических частот восприятия акустической информации человеком (широкополосные кодеры речи). Структура кодера базируется на подходе динамической трансформации алгоритма, вычислении перцептуальной энтропии в области вэйвлет-коэффициентов, эффективном распределении битов кодирования, учитывающего неидеальность преобразования.

Статистическая и перцептуальная избыточность

Основной идеей кодеров является разделение сигнала на частотные компоненты с помощью некого банка фильтров. Далее компоненты сигнала квантуются в частотной области и общее количество бит динамически распределяется в зависимости от энергии каждого спектрального компонента и его значимости. Пусть в какой-то момент времени спектральные ком-

поненты сигнала обладают одинаковой энергией и занимают весь спектр, а также предполагается отсутствие модуля психоакустического анализа информации. Таким образом, все действия сконцентрированы на устранении статистической избыточности (далее просто избыточности). В данном случае увеличение степени компрессии за счет перераспределения общего количества бит между всеми спектральными компонентами не осуществится в силу того, что для кодирования каждого компонента потребуется одно и тоже количество бит. С другой стороны, если допустить, что спектр сигнала "окрашенный", например, основные спектральные компоненты сконцентрированы в области нижних частот, то произойдет перераспределение общего количества бит между всеми спектральными компонентами и значение степени компрессии увеличится. Здесь сигнал содержит избыточность и соответственно в большей или в меньшей степени ее можно устранить. Эффективность этой операции зависит от характеристик применяемого банка фильтров.

Пусть x_k — k-я спектральная компонента сигнала, а $Q(x_k)$ — ее R_k битный квантованный аналог, Q – операция квантования, тогда ошибка реконструкции k-й компоненты равна $q_k = x_k - Q^{-1}(Q(x_k))$. Другими словами, q_k – внесенное искажение в сигнал в результате его кодирования. Среднее число бит на одну спектральную компоненту равно:

$$R = \frac{1}{N}\sum_{k=0}^{N-1} R_k \,, \tag{1}$$

где N — количество спектральных компонент (каналов в банке фильтров). Принимая во внимание, что шум квантователя является белым [6], дисперсия внесенных искажений в сигнал в результате кодирования для ИКМ квантователя равна [6]:

$$q^2 = \frac{1}{N}\sum_{k=0}^{N-1}\left(\frac{x_k^2}{3\cdot 2^{2R_k}}\right). \tag{2}$$

Целью оптимизации является минимизация дисперсии ошибок реконструкции q^2 при ограничении на общее распределение бит. Число уровней реконструкции для квантования компоненты k-го канала банка фильтров $L_k = 2^{R_k}$, тогда

$$R = \frac{1}{N}\sum_{k=0}^{N-1} log_2 L_k = \frac{1}{N}\log_2\prod_{k=0}^{N-1} L_k \,. \tag{3}$$

Далее

$$2^R = \prod_{k=0}^{N-1} L_k = L_g^N \text{, где } L_g = \left(\prod_{k=0}^{N-1} L_k\right)^{\frac{1}{N}} \tag{4}$$

является средним геометрическим значением уровней реконструкции квантователя. Минимизация дисперсии внесенных искажений при кодировании сигнала основывается на методе множителей Лагранжа λ:

$$\frac{d}{dL_k}\left\{\frac{1}{N}\sum_{k=0}^{N-1}\frac{x_k^2}{3\cdot L_k^2} + \lambda\prod_{k=0}^{N-1} L_k\right\} = 0 \,. \tag{5}$$

После дифференцирования и некоторых преобразований формула оптимального распределения бит по каналам банка фильтров примет вид

$$R_k = R + \frac{1}{2}\log_2\left(x_k^2\right) - \frac{1}{2}\log_2\left(\prod_{k=0}^{N-1} x_k^2\right)^{\frac{1}{N}} . \tag{6}$$

Из выражения (6) следует, что минимальное число бит в каждом *k*-м канале определяется распределением спектральной энергии в сигнале и выигрыш в количестве бит по сравнению с однополосным банком фильтров будет только в том случае, когда среднегеометрическое значение спектральной плотности мощности сигнала будет много меньше ее среднеарифметического значения. Отношение среднегеометрического значения спектральной плотности мощности сигнала к ее среднеарифметическому значению есть мера пологости спектра сигнала (Spectral Flatness Measure — *SFM*) [5]:

$$SFM = \frac{\left(\prod_{k=0}^{N-1} x_k^2\right)^{\frac{1}{N}}}{\frac{1}{N}\sum_{k=0}^{N-1} x_k^2}. \qquad (7)$$

Из формулы (7) видно, что значения *SFM* варьируются от 0 до 1. Если $SFM = 1$, то подразумевается, что входной сигнал с пологим спектром и соответственно никакого увеличения компрессии нельзя получить. Пусть $SMF = 1$, тогда, согласно (6), получается, что $R_k = R$. Следует отметить, что *SFM* зависит не только от распределения спектральной энергии сигнала, но также и от разрешающей способности банка фильтров, т.е. от общего числа *N* каналов в банке фильтров.

Таким образом, мерой избыточности в сигнале является мера пологости спектра *SFM* [5]: чем более пологий спектр сигнала, тем меньше избыточности в сигнале. Малое значение *SFM* подразумевает потенциально высокую степень компрессии сигнала, которую естественно можно оценить числом бит, необходимых для кодирования сигнала без артефактов. Из приведенных выше формул может быть получено выражение, показывающее уменьшение энтропии входного сигнала за счет его разбиения банком фильтров.

В перцептуальном кодере сигналов цель не только устранения информационной избыточности, но и изоляции перцептуальной избыточности акустической информации в сигнале. Это желание расположить в спектре сигнала внесенные искажения в реконструированный сигнал в результате кодирования ниже порога маскирования, т.е. порога восприятия акустической информации слушателем. Соотношение сигнал шум *SNR* для квантования компонент каналов банка фильтров равно:

$$SNR = 10\log\frac{x^2}{q^2}, \qquad (8)$$

а соотношение сигнал к порогу маскирования *T SMR* определяется следующим образом:

$$SMR = 10\log\frac{x^2}{T^2}. \qquad (9)$$

Далее для компонент сигнала *k*-го канала, значения которых больше порога маскирования T_k, хотелось бы максимизировать разность *SNR–SMR* или что эквивалентно, минимизировать разность *SMR–SNR*. Для соотношения *SMR–SNR* с учетом дисперсии q^2 (2) дисперсия внесенных искажений кодированием, взвешенная маскирующим фактором, равна:

$$\frac{q^2}{T^2} = \frac{1}{N}\sum_{k=0}^{N-1}\left(\frac{x_k^2 \big/ T_k^2}{3\cdot 2^{2R_k}}\right), \qquad (10)$$

где T_k — уровень порога маскирования в *k*-м канале банка фильтров. Минимизация данной взвешенной ошибки (10), аналогично варианту минимизации дисперсии ошибки реконструкции

q^2 (2), приводит к следующей формуле оптимального распределения бит по каналам банка фильтров:

$$R_k = R + \frac{1}{2}\log_2\left(x_k^2 \Big/ T_k^2\right) - \frac{1}{2}\log_2\left(\prod_{k=0}^{N-1} x_k^2 \Big/ T_k^2\right)^{\frac{1}{N}}. \qquad (11)$$

Из формулы (11) следует, что мера перцептуальной избыточности [7] определяется как отношение:

$$PSFM = \frac{\left(\prod_{k=0}^{N-1} x_k^2 \Big/ T_k^2\right)^{\frac{1}{N}}}{\frac{1}{N}\sum_{k=0}^{N-1} x_k^2 \Big/ T_k^2}. \qquad (12)$$

Как видно из (12), *PSFM* зависит от распределения по частотному диапазону спектральной энергии взвешенной энергией порога маскирования. В данном случае необходимо построить частотно-временное преобразование, характеристики которого зависят от временных изменений сигнала, т.е. обеспечивается требуемое разрешение как по частоте, так по времени, а не только по частоте. Характеристика информационной емкости сигнала в частотной области и область его эффективного кодирования схематически показаны на рис. 1.

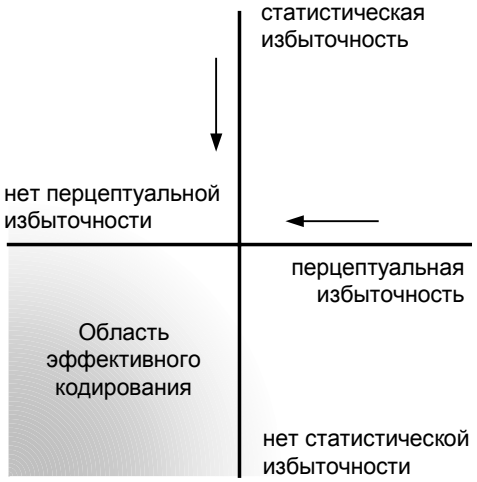

Рис. 1 Характеристика информационной емкости аудиосигнала в частотной области и область его эффективного кодирования (стрелками показаны направления уменьшения избыточности соответсвенно статистической и перцептуальной)

Общая структура перцептуального кодера

Ключевая концепция кодирования аудиосигналов на основе восприятия акустической информации человеком (перцептуальное кодирование [8, 9]) базируется на так называемом пороге едва различимых искажений, который является функцией спектра входного сигнала и параметров психоакустической модели [4], а минимальное число бит, необходимое для кодирования аудиосигнала, оценивается "перцептуальной энтропией" (*PE*) [10]:

$$PE = \frac{1}{N}\sum_{f=f_l}^{f_h} \max\left(0, \log_2 \frac{|signal(f)|}{threshold(f)}\right), \qquad (13)$$

где N — число частотных компонент в частотном диапазоне f_l и f_h; f_l — нижняя частота (например, f_l = 0 Гц) диапазона; f_h — верхняя частота (например, f_h = 22050 Гц) диапазона; $|signal(f)|$ — амплитуда частотной компоненты f; $threshold(f)$ — оценка порога маскирования на частоте f. На практике PE часто называют функцией Джонстона (Johnston, 1988) и вычисляют на основе полосового анализа аудиосигнала [7]:

$$PE = \sum_{i=1}^{25} \sum_{\omega=bl_i}^{bh_i} \log_2 \left(2 \left| n\mathrm{int} \left(\frac{\mathrm{Re}(\omega)}{\sqrt{6T_i / k_i}} \right) \right| + 1 \right) + \log_2 \left(2 \left| n\mathrm{int} \left(\frac{\mathrm{Im}(\omega)}{\sqrt{6T_i / k_i}} \right) \right| + 1 \right), \qquad (14)$$

где i — индекс критической полосы; bl_i и bh_i — нижнее и верхнее значение частоты i-й критической полосы; k_i — количество компонентов преобразования в i-й критической полосе; T_i — значение порога маскирования в критической полосе i; $n$int — операция округления до ближайшего целого значения. Следовательно, та часть сигнала, которая может быть изменена (в общем случае отброшена) и при этом не вносится дополнительных искажений при его восстановлении, является перцептуально избыточной, а часть сигнала, отражающая слышимую акустическую информацию человеком, измеряется и кодируется.

Структуры большинства кодеров сигналов на основе психоакустики сходны и могут быть представлены обобщенной схемой [8] (рис. 2).

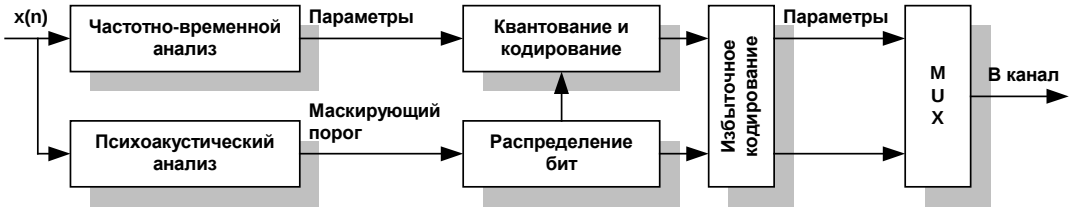

Рис. 2. Общая структура перцептуального аудиокодера

Входной аудиосигнал разбивается на квазистационарные фреймы длительностью от 2 до 50 мс в зависимости от алгоритмов обработки и методов кодирования. Блок частотно-временного анализа аппроксимирует временные и спектральные параметры аудиосигнала для каждого анализируемого фрейма с учетом шкалы критических частотных полос. В блоке психоакустического анализа оценивается энергия маскирующего сигнала (уровни маскирующих порогов) на базе психоакустической модели. При этом определяются максимальные искажения, возникающие в каждой точке частотно-временной плоскости в процессе квантования и кодирования частотно-временных оценок без введения искусственного артефакта слышимости при восстановлении сигнала. Следовательно, психоакустический анализатор вычисляет частотно-временной параметр не восприятия акустической информации слушателем, который затем передается в блок квантования и кодирования. Таким образом, в процессе психоакустического кодирования необходимо, во-первых, установить вид маскирующего сигнала, во-вторых, вычислить соответствующие пороги. Затем полученную информацию использовать для того, чтобы расположить спектр шума кодирования ниже так называемого порога едва различимых искажений JND (just noticeable distortion) [4].

ПДВП-кодеры с адаптивной структурой дерева преобразования

ПДВП есть обобщение диадического вэйвлет-преобразования, которое позволяет получить множество структур путем его соответствующих декомпозиций [11]. ПДВП впервые было предложено в [12] для обработки нестационарных сигналов. Структура ПДВП больше согласуется с сигналом, чем вэйвлет-преобразование [11], и характеризуется следующими свойствами: малая вычислительная сложность процедуры декомпозиции аудиосигнала в выбранном базисе (процедура анализа); малая вычислительная сложность процедуры суперпозиции в выбранном базисе (структура реконструкции сигнала (синтеза)); конвейерность вычислительного процесса процедур анализа и синтеза, что способствует организации поточных и параллельно-поточных структур процессоров реального времени; гибкое изменение временного разрешения, что по-

зволяет выбирать определенной длины фреймы сигнала; гибкое изменение частотного разрешения, обеспечивающее локализацию нестационарностей в сигнале; единственность преобразования, т.е. в ограниченном числе структур ПДВП имеется одна, идентифицирующая соответствующие компоненты сигнала.

Структуру перцептуальных аудио кодеров на основе адаптивного ПДВП [13] укрупненно можно представить в виде следующей схемы (рис. 3).

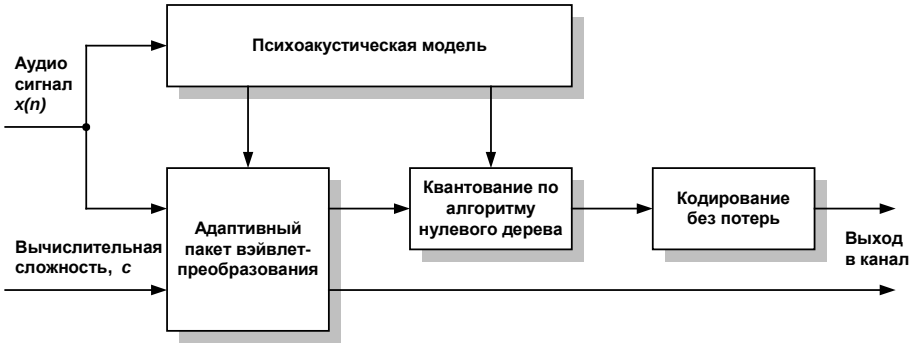

Рис. 3 Структура перцептуального ПДВП-кодера с адаптивным деревом преобразования

На основе перцептуальной энтропии дерево преобразования адаптируется к сигналу, т.е. структура дерева — сигналозависимая, банк анализирующих фильтров инвариантный во времени и базируется на семействе биортогональных сплайн вэйвлет-функций. Основная задача процесса адаптации структуры дерева ПДВП заключается в построении такой декомпозиции полос, которая обеспечивает минимальную скорость передачи при максимизации минимального порога маскирования в каждой полосе. Ширина полос, определяемая текущей структурой дерева ПДВП, может совпадать или нет с частотным разбиением порогов. Для каждой конкретной структуры дерева ПДВП скорость передачи определяется следующим образом: ищутся минимальные значения порогов маскирования в каждой полосе, далее выполняется размещение необходимого количества бит таким образом, чтобы шум квантователя не превышал значения минимального порога маскирования в полосе. Таким образом, "точная" психоакустическая модель, вычисленная в частотной области, на первых уровнях анализа ПДВП сильно загрубляется. Это обусловленно тем, что частотные полосы, "нарезанные" ПДВП на данных уровнях, значительно шире, чем в психоакустической модели. Другой подход адаптации ПДВП предложен в [13]. Здесь используется инвариантная во времени структура дерева преобразования, но адаптируется к сигналу вэйвлет-базис, в частности число коэффициентов фильтра. Схема "глобальной адаптации" рассматривается в [14]. Эффективность перцептуального кодера здесь ассоциируется с оптимизацией анализирующих фильтров для каждого узла дерева ПДВП, а также на основе перцептуальной энтропии оптимизируется структура дерева преобразования: разбиение частотного диапазона на полосы как можно ближе к критической шкале частот.

Основные недостатки данных подходов следующие: во-первых, большая алгоритмическая задержка, обусловленная вычислением на всем фрейме сигнала психоакустической модели и оптимизации структуры дерева преобразования на основе вычисления полного дерева; во-вторых, невозможность организации поточного режима вычисления в кодере из-за того, что психоакустическая модель вычисляется в частотной области на основе дискретного преобразования Фурье; в-третьих, как следствие первого и второго недостатков, кодер работает не в реальном масштабе времени.

Декомпозиция пакета дискретного вэйвлет преобразования

Пусть $\{\varphi_n(t): n \in Z\}$ определяет множество структур деревьев ПДВП и пусть $E \subset \{(l,n): 0 \le l \le L,\ 0 \le n \le 2^l\}$ представляет собой узлы дерева ПДВП, тогда отрезок [0,1) разделяется на диадические интервалы:

$$I_{l,n} = [n2^{-l}, (n+1)2^{-l}], \qquad (15)$$

которые соответствуют специфическому множеству узлов E. В частности,

$$\left\{\varphi_{l,n,k}(t):\ (l,n)\in E, k\in Z\right\}, \tag{16}$$

где $\varphi_{l,n,k}(t)\overset{\Delta}{=}2^{-l/2}\varphi_n\left(2^{-l}t-k\right)$ является базовой формой в пространстве сигнала $\overline{span}\left\{\varphi_0(t-k):\ k\in Z\right\}$. Узел $(l,n)\in E$ дерева ПДВП ассоциируется с частотной полосой, у которой центральная частота и полоса пропускания приблизительно задаются следующими соотношениями:

$$f_{l,n}=2^{-l}\left(GC^{-1}(n)+0{,}5\right)f_s/2, \tag{17}$$

$$\Delta f_{l,n}=2^{-l}\cdot f_s/2, \tag{18}$$

где GC^{-1} — обратный код перестановок Грея; f_s — частота дискретизации сигнала.

Аппроксимация критической шкалы частот на основе ПДВП осуществляется таким образом, чтобы расстояние между центральными частотами $z(f)$ полос пропускания было размером в один барк [15]. На рис. 4 показано дерево ПДВП (Critical Band Wavelet Packet Decomposition (*CB-WPD*)), полученное эмпирически, которое осуществляет разделение частотного интервала аудиосигнала на полосы согласно критической шкале частот [16]:

$$CB\text{-}WPD\text{: }(l,n)\in E_{CB},\ l=\overline{0,8}, \tag{19}$$

где E_{CB} обозначает множество узлов дерева ПДВП, соответствующего *CB-WPD*.

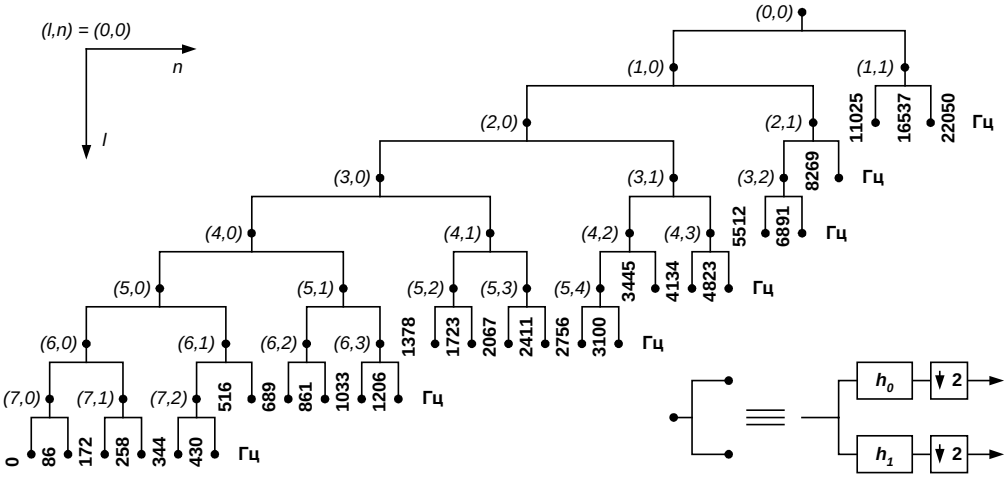

Рис. 4. Дерево ПДВП $(l,n)\in E_{CB}$

Дерево *CB-WPD* делит частотный диапазон, например аудиосигнала [0 – 22,05 кГц], на 25 неравномерных полос *CBW*(*f*), т.е. на 25 барков. Корневой узел $(l,n)=(0,0)$ данного дерева соответствует всему частотному диапазону сигнала. Каждый внутренний узел дерева $(l,n)\in E$, названный узлом предка, делится на два потомка: 1-й потомок и 2-й потомок, ассоциируемые соответственно с высокочастотной и низкочастотной фильтрацией, выходные сигналы (вэйвлет-коэффициенты) которых децимируются в соотношении 2:1:

$$X_{l,n,k}(t)=\left\langle x(t),\varphi_{l,n,k}(t)\right\rangle,\ (l,n)\in E_{CB},\ k\in Z. \tag{20}$$

Следовательно, банк вэйвлет-фильтров (*CB-WPD*: $(l,n)\in E_{CB}$), согласованный с критической шкалой частот восприятия акустической информации человеком [15], является предельной структурой для метода перцептуального кодирования аудиосигнала. Процедура расчета

порогов маскирования в вэйвлет-области для аудиокодера на базе ПДВП, согласованного с критической шкалой частот, приведена в приложении 1 [16].

Поиск структур деревьев ПДВП базируется на известном утверждении [11]: любая комбинация целых индексов $(l,n,k) \in Z$, для которых вэйвлеты сконцентрированы на двоичных интервалах $[n2^{-l}, (n+1)2^{-l}]$ из диапазона $[0, \infty)$ соответствует ортогональным базисам $\psi_{l,n,k}(t)$, $\varphi_{l,n,k}(t)$ из пространства $L^2(k)$. Утверждение доказывает существование множества структур ПДВП. Следовательно, стоит задача поиска такой структуры дерева преобразования из библиотеки, при которой обеспечивается максимальная компрессия без воспринимаемых на слух вносимых искажений в реконструированный сигнал при заданном временном разрешении. Таким образом, оптимизация — это итеративный процесс, и декомпозиция "лучшего" дерева преобразования выбирается как можно ближе к шкале барков [14, 17].

Предположим, что необходимый вычислительный ресурс, под которым понимается производительность процессора и емкость его памяти на фрейме входных данных, есть c_j. Предположим также, что имеющийся в распоряжении вычислительный ресурс равен C. Таким образом, проблема заключается в адаптивном построении структуры дерева ПДВП $(l,n) \in E_j$, $j = 1,2,3,\ldots$, при которой достигается минимум стоимостной функции

$$\min J\left(X_{l,n,k}(t), (l,n) \in E_j, k \in Z\right) \text{ для } c_j < C. \qquad (21)$$

Ее величина ограничивается перцептуальной энтропией *PE* (13)) для заданного вычислительного ресурса C и временного сегмента сигнала длинной кратной степени двух. В перцептуальном кодировании по мере обработки входного фрейма сигнала решение задачи (21) предлагается разбить на два этапа: на основании стоимостных функций для каждого узла дерева преобразования $(l,n) \in E$, определяемых перцептуальной энтропией, осуществляется декомпозиция данных узлов, при которой будет минимизировано требуемое количество бит кодирования вэйвлет-коэффициентов $X_{l,n}$, а шумы квантования не воспринимаются слухом человека. Таким образом, осуществляется рост дерева преобразования; оценивается информативность новой структуры дерева преобразования, т.е. как точно новое частотно-временное разрешение банка фильтров анализа локализовало нестационарности сигнала, а также требуемый вычислительный ресурс.

Стоимостная функция декомпозиции узлов $(l,n) \in E$ дерева ПДВП выбирается следующим образом:

$$J_{l,n} = \sum_{\forall X_{l,n,k}} \log_2\left(2\left[n\mathrm{int}\left(SRM_{l,n,k}\right)\right]+1\right), \text{ (l,n)} \in \text{E, k} \in \text{Z}, \qquad (22)$$

где $SMR_{l,n,k}$ — отношение сигнала к порогу маскирования в полосе узла (l,n) дерева E. Другими словами, отношение (22) вычисляется для каждой частотной полосы (узла дерева ПДВП) и представляет собой отношение среднеквадратического значения сигнала в узле (l,n) к среднеквадратическому значению шума квантования, который будет внесен в реконструированный сигнал. Максимально возможный уровень шума, не воспринимаемый на слух человеком, очевидно, является значением порога маскирования. Таким образом, весовая функция $J_{l,n}$ есть индикатор необходимого числа битов для кодирования аудиосигнала.

Итак, отношение среднеквадратического значения вэйвлет-коэффициентов $X_{l,n,k}$ в полосе узла (l,n) дерева E_j к соответствующему маскирующему порогу $T_{l,n}$, равномерно распределенному между $K_{l,n}$ коэффициентами $X_{l,n,k}$, $k = \overline{1, K_{l,n}}$, узла (l,n), определяется следующим образом [18]:

$$SMR_{l,n,k} = \frac{\left|X_{l,n,k}\right|}{\sqrt{12\dfrac{T_{l,n}}{K_{l,n}}}}, \qquad (23)$$

где знаменатель $\sqrt{12\,T_{l,n}/K_{l,n}}$ — максимальный шаг квантователя $\Delta_{l,n}$ вэйвлет-коэффициентов в узле $(l,n) \in E_j$, а величина $SMR_{l,n,k}$ задает минимальное количество уровней квантования. Следовательно, стоимостная функция $J_{l,n}$ (22) декомпозиции узлов $(l,n) \in E_j$ дерева ПДВП (роста структуры ПДВП) определяется как перцептуальная энтропия узла $(l,n) \in E_j$ и показывает требуемое число двоичных разрядов для кодирования аудиосигнала в частотной полосе, определяемой узлом (l,n):

$$PE_{l,n} = \sum_{k=1}^{K_{l,n}-1} \log_2\left(2\left[n\,\mathrm{int}\left(SRM_{l,n,k}\right)\right]+1\right),\ [\text{бит}/(l,n)],\ (l,n) \in E_j,\ k \in Z. \tag{24}$$

Функция $PE_{l,n}$ $(l,n) \in E_j$ представляет собой функцию перцептуальной энтропии Джонстона (13), однако вычисляемую для действительных коэффициентов и в вэйвлет-области для текущего дерева E_j ПДВП.

В качестве меры информативности дерева ПДВП может быть выбрана энтропия [11]:

$$H(u) = \sum_k p(n) \log \frac{1}{p(n)}, \tag{25}$$

где $p(k) = |u(n)|^2 / \|u\|^2$ — нормализованная энергия k-го элемента вектора $u = \{u(n)\}$, $n = 1,2,3,\ldots$, представленная функцией распределения вероятности p, причем $p\log(1/p) = 0$ для $p = 0$. Исходя из свойств энтропии [11], в частности, характеризующего среднюю неопределенность выбора, применительно в ПДВП-кодеру предлагается конструировать меру количества информативности ПДВП (неопределенности) некой структуры дерева преобразования из множества структур в виде следующей стоимостной функции:

$$J_{l,n} = \sum_{\forall (l,n) \in E_i} \left|X_{l,n,k}\right|^2 \log \frac{1}{\left|X_{l,n,k}\right|^2},\ (l,n) \in E_i,\ k \in Z. \tag{26}$$

С учетом определения энтропии (25) не сложно показать, что

$$H\left(X_{l,n,k}\right) = \frac{J_{l,n}}{\left\|X_{l,n,k}\right\|^2} + \log\left\|X_{l,n,k}\right\|^2,\ (l,n) \in E_i,\ k \in Z, \tag{27}$$

т.е. минимизация стоимостной функции $J_{l,n}$ ведет к минимизации энтропии $H(X_{l,n,k})$, $(l,n) \in E_i$. Меру информативности структуры дерева ПДВП в соответствии с (26) и (27) предлагается конструировать следующим образом [19]:

$$WTE_{E_i} = \sum_{\substack{\text{для} \\ \forall (l,n) \in E_i}} \sum_k \frac{\left|X_{l,n,k}\right|}{\sum\limits_{\substack{\text{для} \\ \forall (l,n) \in E_i}} \left|X_{l,n,k}\right|} \ln\left(\frac{\left|X_{l,n,k}\right|}{\sum\limits_{\substack{\text{для} \\ \forall (l,n) \in E_i}} \left|X_{l,n,k}\right|}\right),\ (l,n) \in E_i,\ k \in Z,\ i = \overline{1,8}, \tag{28}$$

где $X_{l,n,k} \in (l,n)$ — коэффициенты узла (l,n) дерева E_i. Данная стоимостная функция характеризует энтропию вэйвлет-коэффициентов $X_{l,n,k}$ в узлах (l,n) дерева E_i и отражает изменение во времени информативности ПДВП, отсюда и название — временная энтропия вэйвлет-коэффициентов (*WTE* — wavelet time entropy).

Декомпозиция ПДВП, т.е. "рост" дерева преобразования, может осуществляться на основании следующего алгоритма [20].

Алгоритм. Рост дерева ПДВП

Пусть решение о декомпозиции узла (l,n) дерева E_j ПДВП будет обозначаться как $split(l,n)$, где l – уровень декомпозиции, т.е. масштабный уровень преобразования, а n есть n-й узел на уровне l. Пусть текущий узел (предок) будет (l,n), а его потомки определяются как $(l+1,2n)$ и $(l+1,2n+1)$, где $l = 0,1,2,3,\ldots$, $n = 0,1,2,3,\ldots$

Шаг 1. Пусть $l = 0$, $split(l,n)$ = YES, т.е. задан корневой узел (0,0) дерева преобразования E_0 – входной фрейм аудиосигнала, перцептуальная энтропия которого равна $PE_{0,0}$.

Шаг 2. Осуществляется декомпозиция входного сигнала на основе ячейки – банка из двух ортонормальных вэйвлет-фильтров.

Шаг 3. Вычисляется перцептуальная энтропия в узлах декомпозиции.

Шаг 4. $l = l+1$.

ЕСЛИ $l-1$ > максимального масштабного уровня предельного дерева *CB-WPD*,

ТОГДА STOP – конец роста дерева ПДВП.

Шаг 5. Для каждого узла n уровня l рост дерева E_l ПДВП осуществляется следующим образом:

выполняется декомпозиция узла предка (l,n);

вычисляется перцептуальная энтропия в узлах потомках: $PE_{l+1,2n}$ и $PE_{l+1,2n+1}$.

ЕСЛИ $PE_{l,n} \geq PE_{l+1,2n} + PE_{l+1,2n+1}$.

ТОГДА $split(l,n)$ = YES.

ИНАЧЕ $split(l,n)$ = NO.

Шаг 6. Переход к шагу 4.

Таким образом, для каждого входного фрейма сигнала каждый узел-предок (l,n) дерева E_j разделяется на два узла-потомка $(l+1,2n)$ и $(l+1,2n+1)$ тогда и только тогда, когда сумма перцептуальной энтропии в узлах-потомках $(l+1,2n)$ и $(l+1,2n+1)$ меньше, чем значение перцептуальной энтропии в узле–предке (l,n). Данный алгоритм роста дерева ПДВП позволяет определить субоптимальную структуру декомпозиции ПДВП при минимальном числе бит на отсчет аудиосигнала без воспринимаемых на слух искажений, вносимых в процессе кодирования входного сигнала. Достоинством данного алгоритма является то, что рост дерева осуществляется сверху вниз, без возвратов на меньшие масштабные уровни преобразования и необходимости построения полного дерева ПДВП. Применительно к ПДВП-кодеру рост дерева преобразования в большей мере будет наблюдаться в области низких частот. Поэтому "грубая" оценка порогов маскирования $T_{l,n}$ в узлах $(l,n) \in E_j$ по мере построения структуры дерева ПДВП, т.е. с увеличением разрешающей способности по частоте, будет уточняться.

Структура перцептуального кодер-декодера на основе ПДВП

Методы динамической декомпозиции ПДВП, расчет психоакустической модели в вэйвлет-области позволяют построить новую структуру перцептуального ПДВП-кодера аудиосигналов, ориентированную на обработку сигналов только в вэйвлет-области и работу в реальном времени [22]. На рис. 5 показана новая структура перцептуального ПДВП-кодера, ядром которой является адаптивный пакет дискретного вэйвлет-преобразования "адаптивный ПДВП".

Для каждого текущего дерева E_i в темпе обработки сигналов в блоке "адаптивный ПДВП" вычисляются пороги маскирования $T_{l,n}$ в соответствии с процедурой (см. приложение 1); значения перцептуальной энтропии $PE_{l,n}$; энтропия структуры дерева E_i ПДВП WTE_{E_i}. На основании данной информации в блоке "Формирователь структуры дерева ПДВП" рассчитываются параметры реконфигурации дерева ПДВП. Данный процесс осуществляется поступательно, без возвратов на меньшие масштабные уровни преобразования. Следовательно, весь вычислительный процесс идеально ложится на архитектуру параллельно-поточных процессоров [22]. Обработка аудиосигнала выполняется фреймами. Ввиду того что ПДВП осуществляется над каждым последующим фреймом с новой структурой дерева преобразования, то для устранения фазовых разрывов обработка аудиосигнала выполняется с перекрывающимися фреймами, предварительно взвешенными временным окном Хеннинга. Следующим этапом работы кодера является квантование и кодирование коэффициентов оптимального дерева ПДВП

$X_{l,n,k}$, $(l,n) \in E$, $k \in Z$. Управление данным процессом осуществляется алгоритмом размещения бит на основе кодовых книг Хаффмана. Наконец, необходимо неким образом закодировать структуру дерева ПДВП E. Принимая во внимание факт, что рост дерева осуществляется поступательно и изменчивость сигнала во времени более инерционна, чем время обработки, то кодируются только изменения в структуре дерева от фрейма к фрейму.

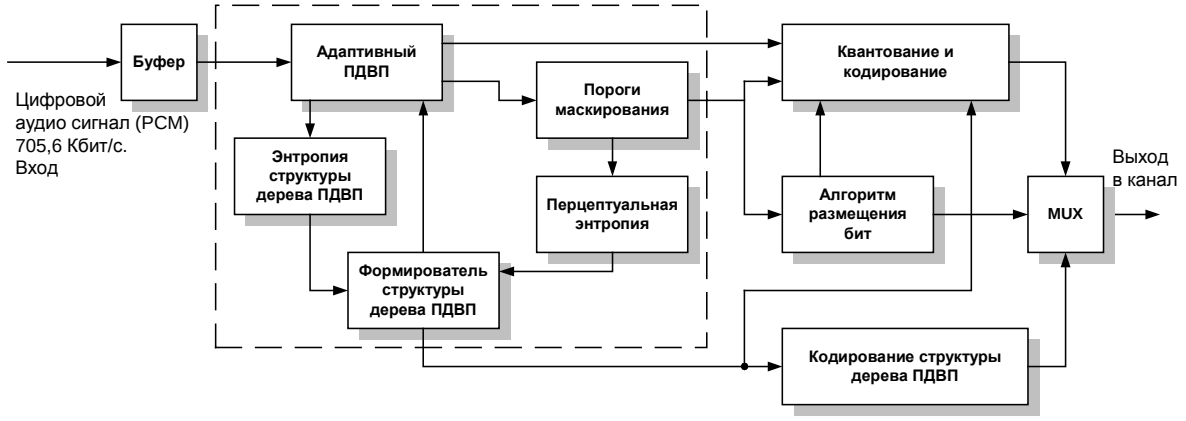

Рис. 5. Структура перцептуального кодера аудиосигналов на базе адаптивного ПДВП

Структура декодера ПДВП-кодера аудиосигналов схематически показана на рис. 6.

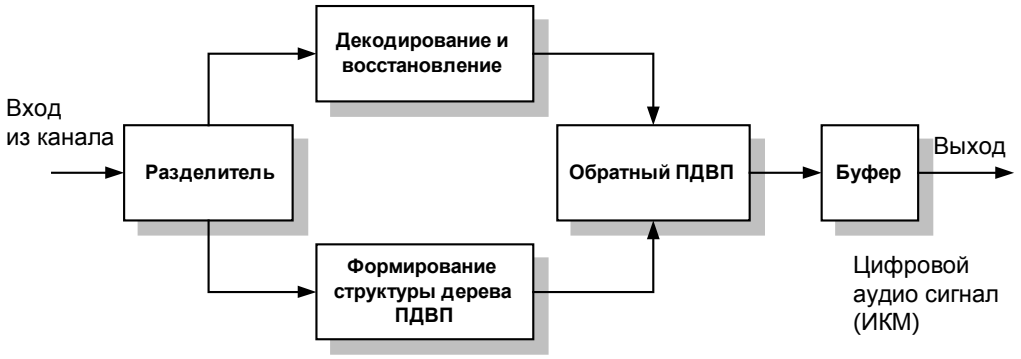

Рис. 6. Структура декодера аудиосигналов

Работа декодера выполняется в следующем порядке: разделяется входная информация на два потока данных: один содержит кодированные вэйвлет-коэффициенты, а второй — код структуры дерева ПДВП; формируется новая структура дерева ПДВП путем суммирования кодовой последовательности, описывающей изменение структуры дерева ПДВП, с текущей кодовой последовательностью структуры дерева ПДВП; выполняется реконструкция аудиосигнала синтезирующим банком цифровых фильтров, реализованным как обратное адаптивное ПДВП.

На рис. 7 приводится сравнение реконструированных аудиосигналов предложенного кодера с известным стандартом MPEG-1, уровень III. Объективные оценки ПДВП-кодера аудиосигналов показывают, что реконфигурированный сигнал имеет достаточно хорошее качество, соответствующее требованиям стандарта ITU-R PEAQ при высокой степени компрессии в 15 раз и более, или, что соответственно, при минимальной скорости передачи от 36 до 45 кбит/с. Восстановленный сигнал не содержит никаких артефактов при оценке отношения шума к порогу восприятия $NMR_{total} \approx -9$ дБ (рис. 8), в то время как у MPEG-1: $NMR_{total} \approx 3$ дБ, скорость передачи — 64 кбит/с.

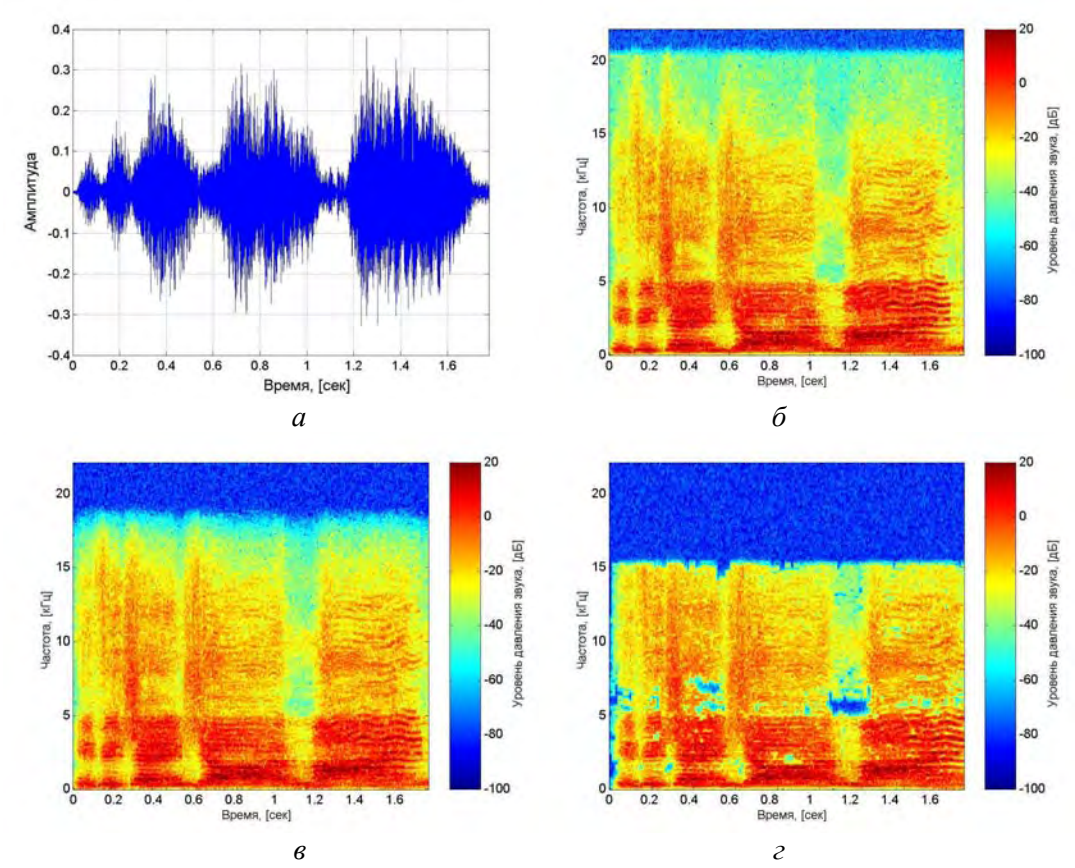

Рис. 7. Аудиосигнал "ABBA "Take a chance on me": *а* — оригинальный сигнал; *б* — его спектрограмма; *в* — спектрограмма реконструированного сигнала (ПДВП-кодер); *г* — спектрограмма реконструированного сигнала (MPEG-1 уровень III)

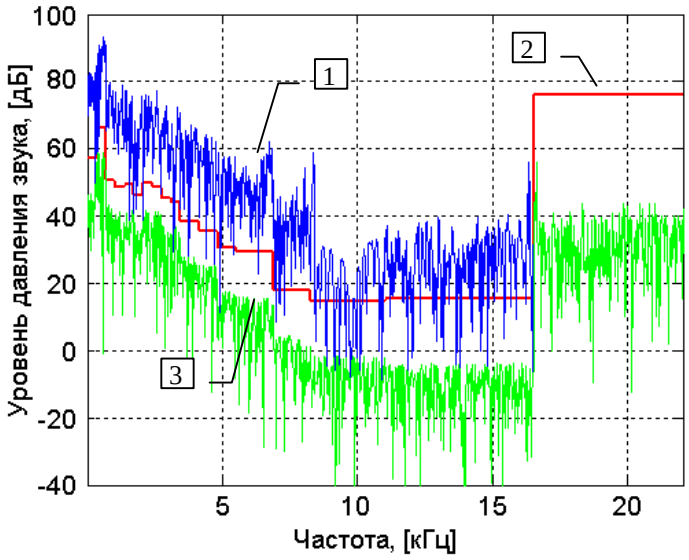

Рис. 8. Спектральная плотность мощности входного сигнала (*1*), порог маскирования (*2*) и шум квантователя (*3*)

Комбинированная система редактирования шумов и кодирования речи

Предлагается комбинированная система редактирования шумов и кодирования речевого сигнала без специального процессора повышения качества речи на основе критического дерева ПДВП *CB-WPD*: $(l,n) \in E_{CB}$, $l = \overline{0,6}$ (рис. 9) и вычисления порога восприятия речевого сигнала человеком. Разработка ориентирована на частоту дискретизации 16 кГц, и обработка введется в 24 барках.

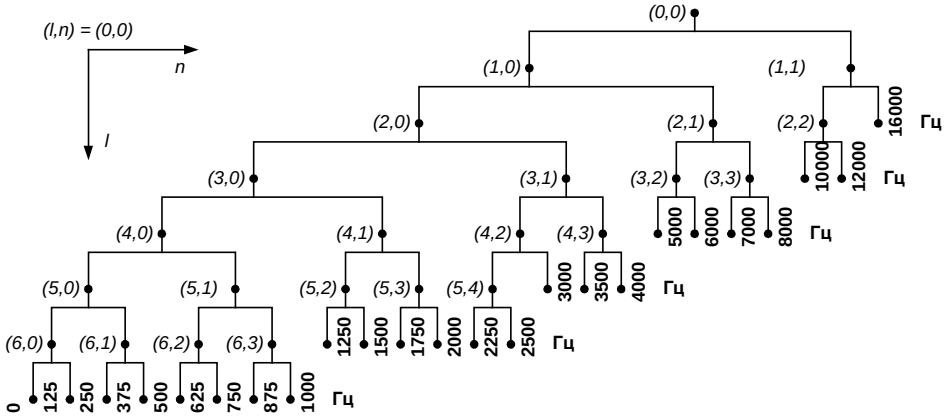

Рис. 9. Дерево ПДВП *CB-WPD* $(l,n) \in E_{CB}$

Пусть $P_{y,m}(k)$, $P_{s,m}(k)$, $P_{n,m}(k)$ – оценки мощности вэйвлет коэффициентов речевого сигнала с аддитивным шумом, чистой речи и шума в обрабатываемом фрейме длинны W *m*-ветви (частотной полосы) дерева ПДВП:

$$P_{y,m}(k_b) = \frac{1}{W}\sum_{i=0}^{W-1} P'_{y,m}(k_b W + i), \tag{29}$$

где k_b — индекс обрабатываемого блока, тогда как оценка мощности рассчитывается на основе экспоненциального усреднения

$$P'_{y,m}(k) = \alpha |y_m(k)|^2 + (1-\alpha) P'_{y,m}(k-1). \tag{30}$$

Различимые на слух фрагменты сигнала по частотным полосам могут быть определены как

$$S_{x,m}(k_b) = \begin{cases} P_{s,m}(k_b), & \text{åñëè} \quad P_{s,m}(k_b) \ge T_m(k_b) \\ T_m(k_b), & \text{åñëè} \quad P_{s,m}(k_b) \le T_m(k_b) \end{cases}, \; 0 \le m \le M-1, \tag{31}$$

$$S_{y,m}(k_b) = \begin{cases} P_{y,m}(k_b), & \text{åñëè} \quad P_{y,m}(k_b) \ge T_m(k_b) \\ T_m(k_b), & \text{åñëè} \quad P_{y,m}(k_b) \le T_m(k_b) \end{cases}, \; 0 \le m \le M-1, \tag{32}$$

где порог едва заметных искажений (порог восприятия) $T_m(k_b)$ частотной полосы *m* блока k_b вычисляется в соответствии с процедурой расчета порогов маскирования в вэйвлет области (см. приложение 1). Слышимый шум рассчитывается согласно формуле

$$S_{n,m}(k_b) = S_{y,m}(k_b) - S_{s,m}(k_b). \tag{33}$$

На основании выражений (31)–(33) слышимый шум для полос $0 \le m \le M-1$ равен

$$S_{n,m}(k_b)=\begin{cases} P_{y,m}(k_b)-P_{s,m}(k_b), & \text{åñëè} \quad P_{y,m}(k_b)\ge T_m(k_b) \quad \text{è} \quad P_{s,m}(k_b)\ge T_m(k_b) \quad (I), \\ P_{y,m}(k_b)-T_m(k_b), & \text{åñëè} \quad P_{y,m}(k_b)\ge T_m(k_b) \quad \text{è} \quad P_{s,m}(k_b)< T_m(k_b) \quad (II), \\ T_m(k_b)-P_{s,m}(k_b), & \text{åñëè} \quad P_{y,m}(k_b)< T_m(k_b) \quad \text{è} \quad P_{s,m}(k_b)\ge T_m(k_b) \quad (III), \\ 0, & \text{åñëè} \quad P_{y,m}(k_b)< T_m(k_b) \quad \text{è} \quad P_{s,m}(k_b)< T_m(k_b) \quad (IV). \end{cases} \tag{34}$$

Как видно из (34), компоненты слышимого шума зависят от энергий сигнала чистой речи $P_{s,m}(k_b)$, зашумленного сигнала $P_{y,m}(k_b)$ и порога восприятия $T_m(k_b)$ для чистой речи, оценки которых вычисляются для блока k_b. Базируясь на концепции правила редактирования шума работы [23, 24], уровень подавления слышимых шумов определяется неравенством

$$S_{n,m}(k_b)\le 0\,,\ 0\le m\le M-1\,, \tag{35}$$

а правило взвешивания имеет следующий вид:

$$S_m(k_b)=G_m(k_b)\cdot y_m(k_b),\ 0\le m\le M-1\,, \tag{36}$$

где $S_m(k_b)$ — отредактированный речевой сигнал. Коэффициент взвешивания для каждого блока обработки k_b рассчитывается по формуле

$$G_m(k)=\frac{1}{\left(\dfrac{a_m(k_b)}{P_{y,m}(k_b)}\right)^{\nu_m}+1},\ k_b\le k\le k_b+W \ \text{ и } \ 0\le m\le M-1 \tag{37}$$

где переменные $a_m(k)$ и $\nu_m\in\Re^+\le 1$ зависят от времени и определяют степень подавления слышимого шума. Коэффициент $a_m(k_b)$ определяет порог, выше которого все компоненты шума подавляются, а параметр ν_m показывает степень подавления и зависит от соотношения $\left[\dfrac{P_{y,m}(k_b)}{a_m(k_b)}\right]$:

$$a_m(k_b)=\left[T_m(k_b)+P_{n,m}(k_b)\right]\left[\frac{P_{n,m}(k_b)}{T_m(k_b)}\right]^{\frac{1}{\nu_m}}. \tag{38}$$

В приложении 2 приводится вывод данного утверждения.

На рис. 10 показана схема обработки речевого сигнала в одной из ветвей *CB-WPD*: $(l,n)\in E_{CB}$, $l=\overline{0,6}$ (соответствующей ей частотной полосе (см. рис. 9)) комбинированной системы редактирования шума и кодирования речевого сигнала.

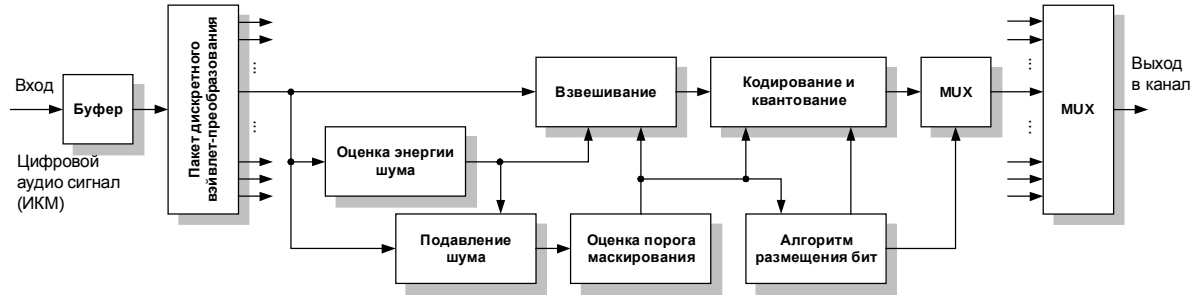

Рис. 10. Структура кодера-редактора шумов речевого сигнала на базе ПДВП

Представленное выше правило (36) повышения качества речи (модификация вэйвлет-коэффициентов в соответствующей полосе) базируется на оценке порога восприятия для чистой речи, в то время как в системе имеется только зашумленный сигнал (одномикрофонный

вариант системы). Грубая оценка $P_{s,m}(k)$ осуществляется на основе метода спектрального вычитания (блок "Подавление шума)"). Тем не менее, повышение качества речи строго зависит от слежения за оценками мощности шума $P_{n,m}(k)$ и порога маскирования $T_m(k_b)$. После удаления из входного речевого сигнала слышимого шума, аналогично как в кодере аудиосигналов (см. рис. 5), осуществляется квантование и кодирование модифицированных вэйвлет-коэффициентов. Структура блока декодера соответствует рис. 6, за исключением блока реконструкции дерева ПДВП, так как здесь дерево фиксировано.

Результаты, представленные на рис. 11, позволяют судить о достаточно хорошем качестве восстановленного речевого сигнала для скорости передачи 17–25 кбит/с. Еще одним достоинством данного широкополосного кодера является то, что как аудио, так и речевые сигналы могут кодироваться.

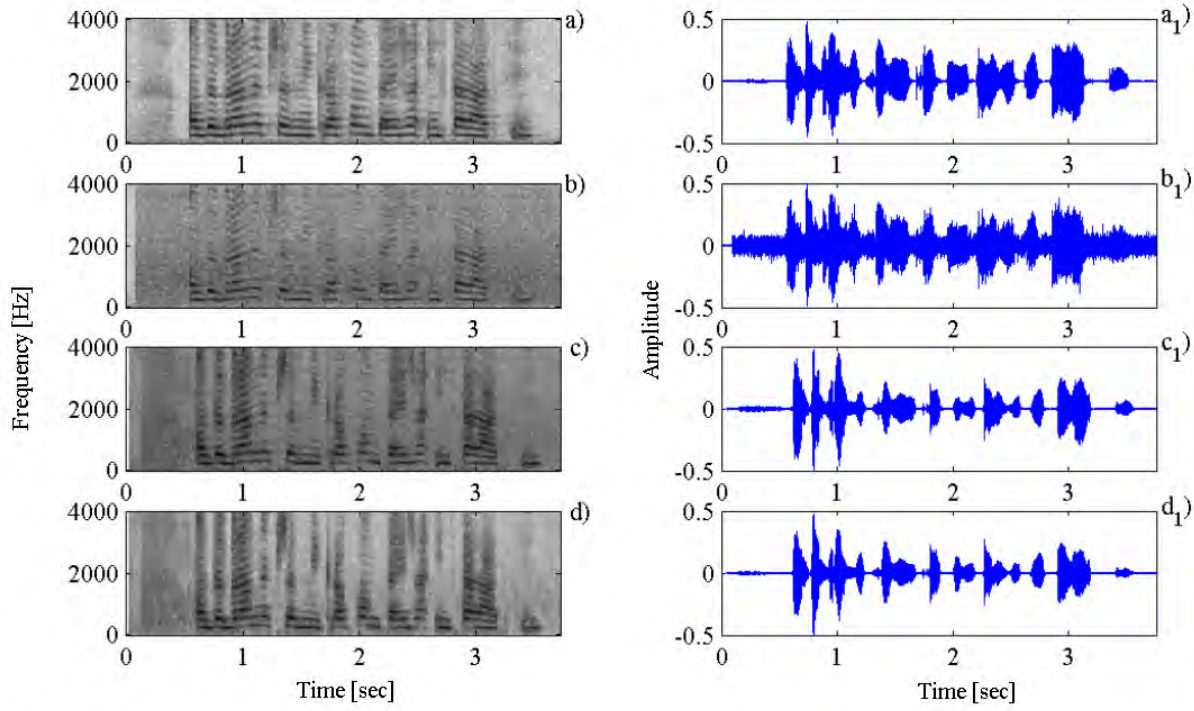

Рис. 11. Результаты обработки речевого сигнала в кодере-редакторе шумов: *a* — чистый речевой сигнал, *b* — зашумленный речевой сигнал, *c* — отредактированный речевой сигнал от шума, *d* — реконструированный речевой сигнал декодером

Приложение 1

Процедура. Расчет порогов маскирования в вэйвлет-области.

Дано: дерево ПДВП, согласованное с критической шкалой частот $(l,n) \in E_{CB}$;
карта частотно-временного разрешения дерева ПДВП;
коэффициенты ПДВП $X_{l,n,k}$.

Вычислить спектральную энергию барка:

$$A_{CB}(z) = \sum_{k=0}^{K-1} X_{z,k}^2 , \qquad (П.1.1)$$

где $z = \overline{1,25}$ — номер критической полосы; K — количество вэйвлет-коэффициентов преобразования в каждой критической полосе z.

Оценить тональность сигнала в каждой критической полосе и значения индексов $a_{tmn}(z)$ и $a_{nmn}(z)$ уменьшения спектральной энергии барка соответственно для тоновых и шумовых маскеров:

индекс $a_{tmn}(z)$, который оценивает отношение маскирования тоном шума, задается так:

$$a_{tmn}(z) = -0{,}275z - 15{,}025\text{, дБ, } z = \overline{1{,}25}; \quad \text{(П.1.2)}$$

индекс маскирования шумом шума a_{nmn} оценивается как константа

$$a_{nmn} = -25\text{, дБ,} \quad \text{(П.1.3)}$$

так как ПДВП уже внесло некоторое спектральное перекрытие;

среднее значение тональности маскеров в каждой критической полосе определяется маскирующим индексом:

$$a_{CB}(z) = \eta a_{tmn}(z) + (1-\eta)a_{nmn}(z)\text{, дБ, } z = \overline{1{,}25}, \quad \text{(П.1.4)}$$

где η — тональный коэффициент:

$$\eta = \min(SFM_{äÁ} / SFM_{äÁmax}, 1), \quad \text{(П.1.5)}$$

где $SFM_{дБ}$ — мера спектральной пологости [11]; $SFM_{дБmax}$ — максимальное значение меры пологости спектра. Для заданного фильтра прототипа $SFM_{дБmax} = -25$ дБ;

Спектральная энергия барка с учетом тональности сигнала равна:

$$D_{CB}(z) = 10 \cdot \log\left(A_{CB}(z) \cdot 10^{\frac{a_{CB}(z)}{10}} \right)\text{, дБ, } z = \overline{1{,}25}. \quad \text{(П.1.6)}$$

Вычислить разброс энергии барка $C_{CB}(z)$ как свертку $D_{CB}(z)$ с функцией разброса $B(z)$ в каждой критической полосе z (значение параметров для функции $B(z)$ определены в первой строке таблицы):

$$C_{CB}(z) = 10\log\left(\frac{1}{K} \sum_{k=1}^{25} 10^{\frac{D_{CB}(k)}{10}} \cdot 10^{\frac{B(z-k)}{10}} \right)\text{, дБ, } z = \overline{1{,}25}. \quad \text{(П.1.7)}$$

Найти временные маскирующие пороги:

аналогично, как и в частотном маскировании, во временном маскировании уже присутствуют некоторые элементы перекрытия, обусловленные ПДВП;

предполагается, что временное маскирование аддитивно сигналу;

временное маскирование определяется через коэффициенты ПДВП в каждой критической полосе z (на рис. одна строка) с учетом временной функции разброса $B(k)$:

$$B(k) = a + \frac{v+u}{2}(k+c) - \frac{v-u}{2}\sqrt{\left(d + (k+c)^2\right)}\text{, дБ;} \quad \text{(П.1.8)}$$

максимальное временное разрешение для ПДВП имеет место в критических полосах верхних частот с минимальной протяженностью по времени $F_{min} = 2$ отсчета или 0.0454 мс;

Таблица.

Функция разброса	v	u	d	c	a
Барк шкала	30 *дБ/барк*	–25 *дБ/барк*	0,3	0,05	15
Временная шкала	0,0825 *дБ/*F_{min}^{*}	–0,0412 *дБ/*F_{min}^{*}	0,3	0,157	0,032/F_{min}

Примечание. F_{min} — минимальная длина анализируемого фрейма.

параметры функции разброса вдоль оси времени определяются как v = 40 *дБ/мс* = 0,0825 *дБ/*F_{min} и u = –20 *дБ/мс* = –0,0412 *дБ/*F_{min} (см. таблицу, строка 2);

вычислить энергию вэйвлет-коэффициента в каждой критической полосе z :

$$E_z(k) = X_{z,k}^2,\ k = \overline{0, K-1}, z = \overline{1{,}25}; \quad \text{(П.1.9)}$$

определить временную функцию разброса энергии в каждой критической полосе *z* как свертку $E_z(k)$ и функции разброса $B(k)$:

$$F_z(m) = \frac{1}{K}\sum_{k=0}^{K-1} E_z(k)\cdot 10^{\frac{B(K-k)}{10}},\ m = \overline{0,K-1}; \qquad (П.1.10)$$

временной фактор маскирования в полосе *z* находится как результат сравнения величин:

$$F_z(k) \geq E_z(k),\ k = \overline{0,K-1},\ z = \overline{1,25}. \qquad (П.1.11)$$

Если данное соотношение выполняется, то в соответствующей критической полосе имеет место временное маскирование, в противном случае нет.

Оценить частотно-временной маскирующий порог $M_{CB}(z)$ в каждой критической полосе:

$$M_{CB}(z) = C_{CB}(z)\,max\left(\frac{F_z(k)}{E_z(k)},1\right), \text{ дБ},\ k = \overline{0,K-1},\ z = \overline{1,25}. \qquad (П.1.12)$$

Результирующее значение маскирующего порога $T_{CB}(z)$ в соответствующей критической полосе частот получается из сравнения временно-частотного маскирующего порога $M_{CB}(z)$ с минимальным значением абсолютного порога слышимости $ATH(z)$ (см. приложение 2):

$$T_{CB}(z) = \max\left(ATH(z), M_{CB}(z)\right), \text{ дБ}. \qquad (П.1.13)$$

Конец процедуры.

Приложение 2

В системе с одним микрофоном в наличии имеется только зашумленный сигнал. Следовательно, все оценки энергии должны быть вычислены базируясь на данном сигнале, а редактирование шума базируется на переходах *I* и *II* выражения (34). Пусть

$$P_{y,m}(k_b) = P_{s,m}(k_b) + P_{n,m}(k_b), \qquad (П.2.1)$$

а оценка мощности "чистой речи" равна

$$P_{\hat{s},m}(k_b) = G_m(k)P_{y,m}(k_b). \qquad (П.2.2)$$

После подстановки (П.2.1) и (П.2.2) в (34) получается, что

$$\begin{aligned} &G_m(k_b)P_{y,m}(k_b) - P_{s,m}(k_b) \leq 0, \quad \text{åñëè} \quad P_{y,m}(k_b) \geq T_m(k_b) \quad \text{è} \quad P_{s,m}(k_b) \geq T_m(k_b) \quad (I), \\ &G_m(k_b)P_{y,m}(k_b) - T_m(k_b) \leq 0, \quad \text{åñëè} \quad P_{y,m}(k_b) \geq T_m(k_b) \quad \text{è} \quad P_{s,m}(k_b) \leq T_m(k_b) \quad (II). \end{aligned} \qquad (П.2.3))$$

И правило повышения качества речи формулируется при решении (П.2.3) относительно $a_m(k_b)$:

$$\begin{aligned} &a_m(k) \geq P_{y,m}(k_b)\left[\frac{P_{y,m}(k_b)}{P_{s,m}(k_b)} - 1\right]^{\frac{1}{\nu_m}}, \quad \text{åñëè} \quad P_{y,m}(k_b) \geq T_m(k_b) \quad \text{è} \quad P_{s,m}(k_b) \geq T_m(k_b) \quad (I), \\ &a_m(k) \geq P_{y,m}(k_b)\left[\frac{P_{y,m}(k_b)}{T_m(k_b)} - 1\right]^{\frac{1}{\nu_m}}, \quad \text{åñëè} \quad P_{y,m}(k_b) \geq T_m(k_b) \quad \text{è} \quad P_{s,m}(k_b) \leq T_m(k_b) \quad (II). \end{aligned} \qquad (П2.4))$$

Если принять во внимание (П2.1) и положить в (П.2.4 *I*) равенство, а также сделать замену $P_{s,m}(k_b)$ на $T_m(k_b)$, причем $P_{s,m}(k_b) \geq T_m(k_b)$ для $0 < \nu_m \leq 1$, получается, что

$$\left[T_m(k_b)+P_{n,m}(k_b)\right]\left[\frac{P_{n,m}(k_b)}{P_{s,m}(k_b)}-1\right]^{\frac{1}{\nu_m}} \geq \left[P_{s,m}(k_b)+P_{n,m}(k_b)\right]\left[\frac{P_{n,m}(k_b)}{P_{s,m}(k_b)}-1\right]^{\frac{1}{\nu_m}}. \qquad (П.2.5)$$

Следовательно, переход *I* (П.2.4) выполняется при переходе от оценки чистой речи к порогу восприятия речевого сигнала. Из условия *II* (П.2.4) следует, что $a_m(k_b)$ пропорциональна $P_{s,m}(k_b)$. Следовательно, замена $T_m(k_b)$ на $P_{s,m}(k_b)$ также будет справедлива.

PERCEPTUAL CODING OF AUDIO AND SPEECH SIGNALS

A.A. PETROVSKY, K. BIELAWSKI, AL.A. PETROVSKY

Abstract

This paper introduce the new approach to design of perceptual audio and speech coders based on the psychoacoustically wavelet packet decomposition. The combined noise reduction and speech coding system is proposed also. The system based on the critical band wavelet packet decomposition (*CB-WPD*) and psychoacoustic weighting rule of input signal.

Литература

1. Application of digital signal processing to audio and acoustics / edited by Mark Kahrs, Karl-heinz Brandenurg. // Kluwer Academic Publishers, Boston, 1998. 545 p.
2. Multimedia System, Standards, and Networks / edited by Atul Puri, Tsuhan Chen // Marcel Dekker, Inc., New York, 2000. 636 p.
3. *Kondoz A.M.* Digital Speech: coding for low bit rate communication systems // John Wiley & Sons, New York, 1994. 442 p.
4. *Jayant N.S., Chen E.Y.* Audio compression: technology and applications // AT&T technical journal. 1995. Vol. 74, №2. P. 23-34.
5. *Bosi M.* Filter banks in perceptual audio coding // The Proc. of the AES 17th International Conference "High-Quality Audio Coding". Florence, Italy. 1999. P. 125-136.
6. *Рабинер Л., Гоулд Б.* Теория и применение цифровой обработки сигналов. М.: Мир, 1978. 848 с.
7. *Johnston J.* "Audio coding with filter banks," in Subband and Wavelet Transforms / A. Akansu and M. J. T. Smith // Eds: Kluwer Academic, 1996. P. 287-307.
8. *Painter T., Spanias A.* Perceptual Coding of Digital Audio // Proceedings of the IEEE. April 2000. Vol. 88, № 4. P. 451-513.
9. *Петровский Ал.А.* Компрессия аудиосигналов на базе психоакустики: подходы и структуры // Республ. межвед. сб. научн. тр. "Радиотехника и электроника". Мн.: БГУИР, 1999. Вып. 24. С. 140-149.
10. *Brandenburg K.* Perceptual coding of high quality digital audio // Applications of Digital Signal Processing to Audio and Acoustics / M. Kahrs, K. Brandenburg, Eds. – Boston. MA: Kluwer Academic, 1998. P. 39-83.
11. *Wickerhauser M.V.* Adaptive Wavelet Analysis from Theory to Software. A.K. Peters Ltd., Massachusetts, 1994. 486 p.
12. *Coifnam R., Meyer Y., Quake S., Wickerhauser V.* Signal Processing and compression with wavelet packet // Numerical Algorithms Research Group. New Haven, CT: Yale University, 1990. 196 p.
13. *Sinha D., Tewfik A.* Low bit rate transparent audio compression using adapted wavelets // IEEE Trans. Signal Processing. Dec. 1993. Vol. 41. P. 3463-3479.
14. *Philippe P., Saint-Martin F. M., Lever M.* Wavelet packet filterbanks for low time delay audio coding // IEEE Trans. Speech Audio Processing. 1999. Vol. 7, № 3. P. 310-322.
15. *Zwicker E., Fastl H.* Psychoacoustics: Facts and Models. Berlin, Germany: Springer-Verlag, 1990. 380 p.
16. *Петровский Ал.А.* Расчет маскирующих порогов для аудио кодеров на базе пакетного дискретного вэйвлетного преобразования // Республ. межвед. сб. научн. тр. "Радиотехника и электроника". Мн.: БГУИР, 2000. Вып. 25. С. 44-57.
17. *Cohen I.* Enhancement of speech using bark-scaled wavelet packet decomposition // The Proc of EUROSPEECH. Aalborg, Denmark, 3-7 Sep. 2001. P. 1933-1936.

18. *Петровский Ал. А.* Динамическая реконфигурация пакетного вэйвлетного преобразования на основе вычисления перцептуальной энтропии // Идентификация образов. Мн.: ИТК НАН Беларуси, 2001. С. 45-52.
19. *Petrovsky Al.* Perceptually optimized time-varying wavelet packet decomposition and its ap-plications in acoustic signal processing // 17th International Congress of Acoustics (ICA'2001). Rome, Italy, 2-7 Sept. 2001.
20. *Petrovsky Al. A., Petrovsky A.* Dynamic algorithm transforms for reconfigurable real-time audio coding processor // Proc. "Parallel computing in electrical engineering". IEEE Computer Soc. Press, NJ, 2002. P. 231-234.
21. Audio coding with a masking threshold adapted wavelet packet based on run-time reconfigur-able processor architecture / Al. Petrovsky, A.A. Petrovsky. Amsterdam, Netherlands, May 2001. 8 p.
22. *Petrovsky Al., Krahe D., Petrovsky A.A.* Real-Time Wavelet Packet-based Low Bit Rate Audio Coding on a Dynamic Reconfigurable System // Proc. of the 114th AES Convention. Amsterdam, Netherlands, 22-25 May, 2003. 22 p.
23. *Bielawski K., Petrovsky A.A.* Proposition of minimum bands multirate noise reduction system which exploits properties of human auditory system and all-pass transformed filter bank // IEEE Workshop SIGNAL PROCESSING' 2001. Poznan, 2001. P. 65-70.
24. *Tsoukalas D.E., Mourjopolous J.N., Kokkinakis G.* Improving the intelligibility of noise speech using an audible noise suppression technique // Proc. of 5th European Conference on Speech Communication and Technology. 1997. P. 1415-1418.