

## ПЕРЦЕПТУАЛЬНЫЙ КОДЕР ЗВУКА НА БАЗЕ БЫСТРОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ С ДИНАМИЧЕСКОЙ ТРАНСФОРМАЦИЕЙ ЧАСТОТНО-ВРЕМЕННОГО ПЛАНА

*Петровский А. А., к.н.т., доцент Белорусского государственного университета информатики и радиоэлектроники, email: petrovsky@bsuir.by*

**Ключевые слова:** кодер звука, перцептуальный, вейвлет-преобразование, фрейм, частотно-временной план, психоакустический анализ, порог маскирования, алгоритм, структура, вычислительная платформа.

### Введение

Структуры большинства кодеров звука на основе психоакустики сходны и могут быть представлены обобщенной схемой (рис. 1) [2-4].

Входной звуковой сигнал  $x(n)$  разбивается на квазистационарные фреймы длительностью от 2 до 50 миллисекунд в зависимости от алгоритмов обработки и методов кодирования. Блок частотно-временного анализа аппроксимирует временные и спектральные параметры звука 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 невосприимчивости акустической информации слушателем, который затем передается в блок квантования и кодирования. 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 порога маскирования, известный как порог едва различимых искажений  $JND$  (just noticeable distortion) [5].

Значения психоакустических порогов маскирования неявно показывают результат кодирования звука, а точнее вероятный шаг квантования. Минимальное число

*Предлагается перцептуальный кодер звука на основе адаптируемого под сигнал и заданный вычислительный ресурс быстрого вейвлет-преобразования с динамической трансформацией частотно-временного плана [1].*

бит, необходимое для кодирования воспринимаемого слушателем сигнала, оценивается перцептуальной энтропией  $PE$  (perceptual entropy) [6]:

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

где  $N$  – число частотных компонент в частотном диапазоне  $f_l$  и  $f_h$ ;  $f_l$  – нижняя частота диапазона;  $f_h$  – верхняя частота диапазона, например  $f_h = 22050$  Гц;  $|signal(f)|$  – амплитуда частотной компоненты  $f$ ;  $threshold(f)$  – оценка порога маскирования на частоте  $f$ . На практике,  $PE$  в форме (1) часто называют функцией Джонсона и вычисляют на основе субполосного анализа звука [6]:

$$PE = \sum_{i=1}^{25} \sum_{\omega=bl_i}^{bh_i} \log_2 \left( 2 \left| \text{nint} \left( \frac{Re(\omega)}{\sqrt{6 \cdot T_i / k_i}} \right) \right| + 1 \right) + \log_2 \left( 2 \left| \text{nint} \left( \frac{Im(\omega)}{\sqrt{6 \cdot T_i / k_i}} \right) \right| + 1 \right), \quad (2)$$

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

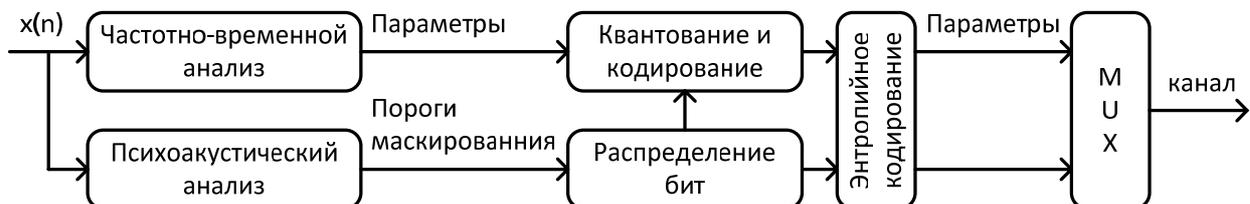


Рис. 1. Общая структура перцептуального кодера звука

Выбор метода частотно-временного анализа неразрывно взаимосвязан с точностью вычисления психоакустической модели и общей концепцией построения системы кодирования сигналов звука. При этом выбор методологии частотно-временного анализа всегда обуславливает решение фундаментального вопроса: обеспечения требуемого разрешения как в частотной так и временной областях. Развитие этих работ в настоящее время идет по пути построения перцептуальных кодеров звука на основе подходящего частотно-временного преобразования, в частности, банков цифровых фильтров и наиболее гибкого, в смысле управления частотно-временным разрешением, быстрого вейвлет преобразования (пакет дискретного вейвлет преобразования (ПДВП)) с динамической трансформацией дерева алгоритма.

ПДВП есть обобщение диадического вейвлет преобразования, которое позволяет получить множество структур путем его соответствующих декомпозиций [7]. Структура ПДВП больше согласуется с сигналом, чем вейвлет преобразование [8], и характеризуется следующими свойствами: малая вычислительная сложность процедуры декомпозиции звукового сигнала в выбранном базисе (процедура анализа); малая вычислительная сложность процедуры реконструкции сигнала (синтеза) в выбранном базисе; конвейерность вычислительного процесса процедур анализа и синтеза, что способствует организации поточных и параллельно-поточных структур процессоров реального времени; гибкое изменение частотно-временного разрешения, обеспечивающее локализацию нестационарностей в сигнале; единственность преобразования, т.е. в ограниченном числе структур ПДВП имеется одна структура, идентифицирующая соответствующие компоненты сигнала [9].

**Психоакустический анализ**  
**Аппроксимация критического частотно-временного плана**

Пусть  $\{\varphi_n(t): n \in \mathbb{Z}\}$  определяет множество структур деревьев ПДВП, и пусть  $E \subset \{(l, n): 0 \leq l \leq L, 0 \leq n \leq 2^l\}$  представляет собой узлы дерева ПДВП [9], тогда отрезок  $[0,1)$  разделяется на диадические интервалы

$$I_{l,n} = [n2^{-l}, (n+1)2^{-l}), \tag{3}$$

которые соответствуют специфическому множеству узлов  $E$ . В частности,

$$\{\varphi_{l,n,k}(t): (l, n) \in E, k \in \mathbb{Z}\}, \tag{4}$$

где  $\varphi_{l,n,k}(t) \triangleq 2^{-\frac{l}{2}}\varphi_{l,n}(2^{-l}t - k)$  является базовой формой в пространстве сигнала  $\overline{\text{span}}\{\varphi_{0,0}(t - k): k \in \mathbb{Z}\}$ . Узел  $(l, n) \in E$  дерева ПДВП ассоциируется с частотной полосой, у которой центральная частота и полоса пропускания приблизительно задаются следующими соотношениями:

$$f_{l,n} = 2^{-l}(GC^{-1}(n) + 0.5) \cdot \frac{f_s}{2}, \tag{5}$$

$$\Delta f_{l,n} = 2^{-l} \cdot \frac{f_s}{2}, \tag{6}$$

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. Соотношение между шкалой частот в герцах и барках приведено в [10].

На рис. 2 показано дерево ПДВП (Critical Band Wavelet Packet Decomposition (CB – WPD)), полученное эмпирически, которое осуществляет разделение частотного интервала звукового сигнала на полосы, согласно критической шкале частот [11]:

$$CB - WPD: (l, n) \in E_{CB}, l = \overline{0, 8}, \tag{7}$$

где  $E_{CB}$  – обозначает множество узлов дерева ПДВП, соответствующего CB – WPD.

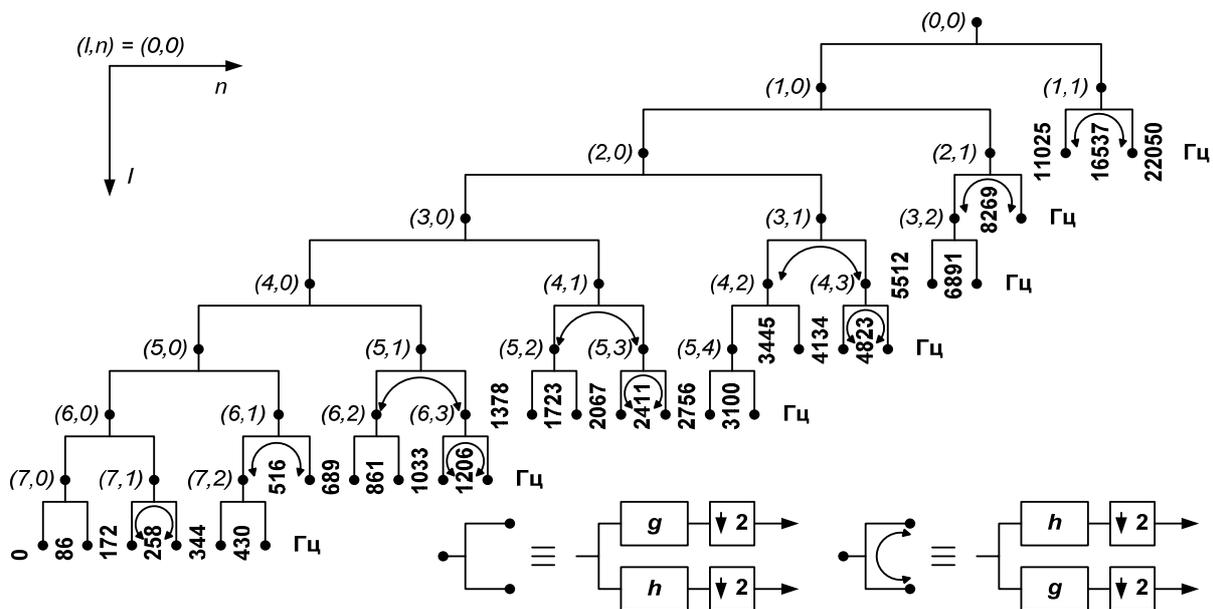


Рис.2. Дерево ПДВП  $(l, n) \in E_{CB}$

На рис. 3 и рис. 4 показаны аппроксимации центральной частоты и ширины каждой частотной полосы критической шкалы частот соответственно деревом ПДВП, структура которого приведена на рис. 2.

Дерево  $CB - WPD$  делит частотный диапазон  $[0 - 22,05 \text{ кГц}]$  на 25 неравных полос  $CBW(f)$  согласно [10], 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 (вейвлет коэффициенты) которых децимируются в соотношении 2:1:

$$X_{l,n,k}(t) = \langle x(t), \varphi_{l,n,k}(t) \rangle, (l, n) \in E_{CB}, k \in \mathbb{Z}. \quad (8)$$

где  $l$  - номер масштабного уровня преобразования ( $0 \leq l \leq 8$ ),  $n$  - номер узла масштабного уровня преобразования,  $k$  - число вейвлет коэффициентов в полосе (узле  $(l, n)$  дерева  $E$ ). Банк вейвлет фильтров ( $CB - WPD: (l, n) \in E_{CB}$ ), согласованный с критической шкалой частот восприятия акустической информации человеком, является предельной структурой для метода перцептуального кодирования сигнала звука [11]. Ширина каждой клеточки частотно-временного плана есть длина фрейма, которая определяется как  $F_l = 2^l (F_{min} = 2$  отсчетов и  $F_{max} = 256$  отсчетов). Следовательно, длина анализируемого окна равна  $W = (P - 1)(F_l - 1) + 1$  отсчетов. Для первого уровня  $l = 1$  преобразования определяющей является область верхних частот, и длина окна  $W = 40$  отсчетов при длине фильтра прототипа  $P = 40$ . Для уровня  $l = 8$  преобразования наибольшая частотная разрешающая способность в области нижних частот, а окно  $W = 9946$  отсчетов.

#### Расчет порогов маскирования в области вейвлет-коэффициентов

Порог маскирования  $T_{l,n}$  в каждой критической полосе  $(l, n) \in E_{CB}$  вычисляется на основе спектральных компонент сигнала, попавших в полосу. Значение маскирующего порога непосредственно зависит от уровня

давления звука, частоты маскира, тональности маскира и маскируемых маскиров. Для точного определения маскирующих порогов нужен учет совместного влияния как частотного, так и временного маскирования. Причем, пост-маскирование, т.е. маскирование вперед, имеет более важное значение перед маскированием назад в силу своей продолжительности по времени.

Процедура расчета порогов маскирования в вейвлет области для кодера звука на базе ПДВП, согласованного с критической шкалой частот, показана в [12].

#### Динамическая трансформация частотно-временного плана на основе перцептуального критерия

##### Критерий оптимизации структуры ПДВП

Поиск структур деревьев ПДВП базируется на известном утверждении [9]: любая комбинация целых индексов  $(l, n, k) \in \mathbb{Z}$ , для которых вейвлеты 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. Причем, ПДВП ассоциируется с алгоритмом выбора лучшей структуры  $E$  преобразования из множества структур, путем изменения и минимизации определенной меры качества (стоимостной функции)  $J$ , зависящей от обрабатываемых данных  $\min J(Ax)$ , где  $A$  - структура ПДВП,  $x$  - входной сигнал. В перцептуальном кодировании звукового сигнала, в первую очередь, интересна такая декомпозиция ПДВП, при которой минимизируется скорость передачи с сохранением соответствующего высокого качества восприятия человеком декодированного сигнала. Таким образом, оптимизация - это итеративный процесс, и «лучшее» дерево преобразования выбирается как можно ближе к шкале барков [13]. Конструкция алгоритма выбора лучшей декомпозиции  $E$  ПДВП следующая [9]. Для каждого узла  $(l, n) \in E$  дерева ПДВП вычисляется стоимостная функция  $J$ . Далее рост структуры дерева осуществляется в соответствии со следующей процедурой:

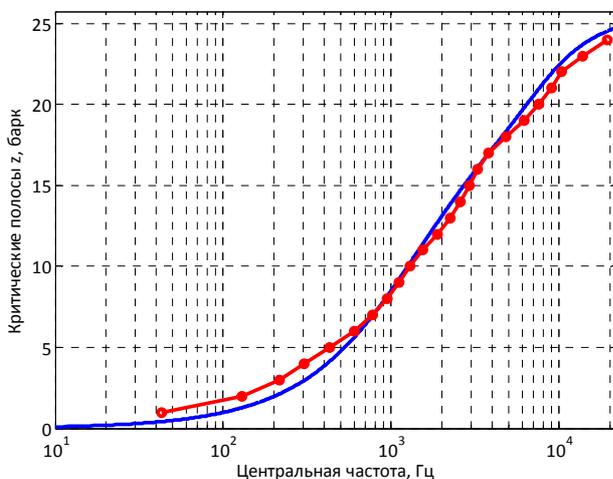


Рис. 3 Аппроксимация центральных частот  $CB - WPD: (l, n) \in E_{CB}, l = 0,8$

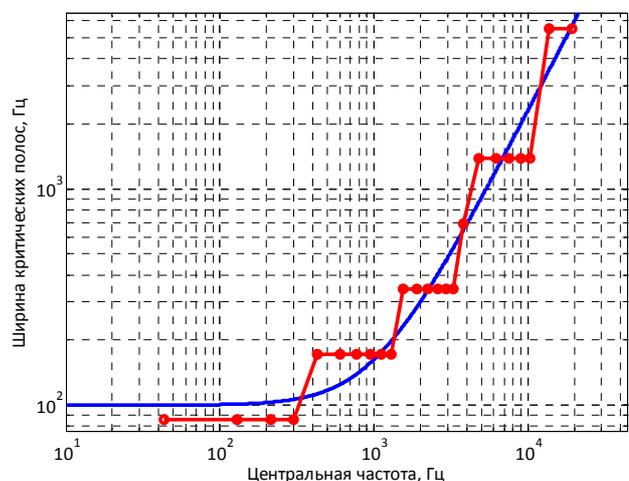


Рис. 4 Аппроксимация ширины критических частотных полос  $CB - WPD: (l, n) \in E_{CB}, l = 0,8$

Процедура роста структуры дерева ПДВП:

**Пусть**  $PE_{\text{предок}}$  - мера качества узла предка дерева ПДВП,

$PE_{\text{потомки}}$  - сумма мер качества соответствующих узлов потомков,

**Тогда,**

**если**  $PE_{\text{предок}} > PE_{\text{потомки}}$ ,

**то**  $PE_{\text{предок}} = PE_{\text{потомки}}$ ,

**иначе** отсечь узлы потомков.

Структура банка фильтров (дерева преобразования ПДВП) определяет вычислительную сложность алгоритма процедур анализа и синтеза на каждом фрейме анализа. Стратегия выбора структуры дерева ПДВП должна базироваться под заданный вычислительный ресурс (вычислительный ресурс – реконфигурируемая вычислительная среда: поточные, параллельно-поточные структуры процессоров). При этом декомпозиция дерева преобразования осуществляется в процессе работы ПДВП кодера, т.е. «на ходу». Преимущество данного подхода в том, что кодер учитывает локальные изменения спектра звукового сигнала. Любая структура дерева  $E$  может быть выражена конечными его узлами  $(l, n)$ :

$$E_j = \{(l, n) : l = \overline{1, L}; n = \overline{0, 2^l - 1}\}, \quad (9)$$

где  $j$  – номер структуры дерева из множества  $E$ ;  $L$  – максимальное значение масштабного уровня  $j$ -ой структуры дерева ПДВП.

Множество структур  $E_j$  образуют пространство структур деревьев ПДВП. В силу того, что в структуре дерева ПДВП каждый узел может является предком двух потомков (двух новых узлов) при переходе с уровня  $l$  на новый уровень  $l + 1$ , то структуры ПДВП зависимы друг от друга:

$$E_j \in E_{j+1}. \quad (10)$$

По мере обработки входного фрейма сигнала решение задачи предлагается разбить на два этапа:

1. 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 вейвлет коэффициентов  $X_{l,n,k}$ , а шумы квантования не воспринимаются слухом человека. Таким образом, осуществляется рост дерева преобразования.

2. Оценивается информативность новой структуры дерева преобразования, т.е. как точно новое частотно-временное разрешение банка фильтров анализа локализовало нестационарности сигнала, а также определяется требуемый вычислительный ресурс.

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

$$PE_{l,n} = \sum_{\forall X_{l,n,k}} \log_2(2[\text{rint}(SMR_{l,n,k})] + 1), \quad (11)$$

$(l, n) \in E, k \in \mathbb{Z}$ ,

где  $SMR_{l,n,k}$  – отношение сигнала к порогу маскирования в полосе узла  $(l, n)$  дерева  $E$ . Соотношение (11) вычисляется для каждой частотной полосы (узла дерева ПДВП) и представляет собой отношение среднеквадратического значения вейвлет коэффициентов 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  $PE_{l,n}$  является индикатором необходимого числа битов для кодирования звука.

В качестве меры информативности дерева ПДВП может быть выбрана энтропия. Исходя из свойств энтропии [9], в частности, характеризующих 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:

$$J_{l,n} = \sum_{\forall (l,n) \in E_j} |X_{l,n,k}|^2 \log \frac{1}{|X_{l,n,k}|^2}, \quad (l, n) \in E, k \in \mathbb{Z}. \quad (12)$$

С учетом определения энтропии несложно показать, что

$$H(X_{l,n,k}) = \frac{J_{l,n}}{\|X_{l,n,k}\|^2} + \log \|X_{l,n,k}\|^2, \quad (l, n) \in E_i, k \in \mathbb{Z}, \quad (13)$$

т.е. минимизация стоимостной функции  $J_{l,n}$  ведет к минимизации энтропии  $H(X_{l,n,k})$ ,  $(l, n) \in E_j$ .

**Динамическая трансформация алгоритма ПДВП**

Как отмечалось выше, декомпозиция дерева ПДВП должна осуществляться в процессе работы ПДВП кодера, т.е. «на ходу», а структура дерева  $CB - WPD$  является предельной структурой преобразования. Таким образом, стоит задача оценки  $SMR$  не в критических частотных полосах восприятия акустической информации человеком, а в частотных полосах соответствующих узлов  $(l, n)$  текущего дерева  $E_j$  преобразования (см. [11]). Отношение среднеквадратического значения вейвлет 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 [11]:

$$SMR_{l,n,k} = \frac{|X_{l,n,k}|}{\sqrt{12 \cdot T_{l,n} / K_{l,n}}}, \quad (14)$$

где знаменатель  $\sqrt{12 \cdot T_{l,n} / K_{l,n}}$  представляет собой максимальный шаг квантователя  $\Delta_{l,n}$  вейвлет-коэффициентов в узле  $(l, n) \in E_j$ , а величина  $SMR_{l,n,k}$  задает минимальное количество уровней квантования.

Следовательно, стоимостная функция  $PE_{l,n}$  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 звукового сигнала в частотной полосе, определяемой узлом  $(l, n)$ :

$$PE_{l,n} = \sum_{k=0}^{K_{l,n}-1} \log_2(2[\text{rint}(SMR_{l,n,k})] + 1), \quad \left[ \frac{\text{бит}}{(l,n)} \right], \quad (15)$$

$(l, n) \in E_j, k \in \mathbb{Z}$ .

Функция  $PE_{l,n}(l, n) \in E_j$  представляет собой функцию перцептуальной энтропии Джонстона (1), однако вычисляемую для действительных коэффициентов и в вейвлет области для текущего дерева  $E_j$  ПДВП. Декомпози-

ция ПДВП, т.е. «рост» дерева преобразования или другими словами динамическая трансформация частотно-временного плана, может осуществляться на основании следующего алгоритма:

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  $l = 0, 1, 2, 3, \dots, n = 0, 1, 2, 3, \dots$

Шаг 1. Пусть  $l = 0$ ,  $split(l, n) = YES$ , т.е. задан корневой узел  $(0, 0)$  дерева преобразования  $E_0 = E_j$ , где  $j = 0$  – входной фрейм звукового сигнала, перцептуальная энтропия которого равна  $PE_{0,0}$  и информативность дерева равна  $H_{E_0}$

Шаг 2. Осуществляется декомпозиция узлов предков  $(l, n)$  входного сигнала на основе ячейки – банка из двух ортонормальных вейвлет фильтров.

Шаг 3.  $l = l + 1, j = j + 1$ .

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

ТОГДА STOP – конец роста дерева ПДВП.

Шаг 4. Вычисляется перцептуальная энтропия  $PE_{l,2n}$  и  $PE_{l,2n+1}$  в узлах декомпозиции  $(l, 2n)$  и  $(l, 2n + 1)$  соответственно для всех  $n$ .

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

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. Оценивается информативность  $H_{E_j}$  дерева  $E_j$ :

ЕСЛИ  $H_{E_j} > H_{E_{j-1}}$ ,

ТОГДА STOP – конец роста дерева, результирующее дерево ПДВП  $E_{j-1}$ .

Шаг 7. Переход к шагу 2.

Достоинством данного алгоритма является то, что рост дерева осуществляется сверху вниз, без возвратов на меньшие масштабные уровни преобразования и необходимости построения полного дерева ПДВП [1]. Данный алгоритм роста дерева ПДВП позволяет определить субоптимальную структуру декомпозиции ПДВП при минимальном числе бит на отсчет звукового сигнала без воспринимаемых на слух искажений, вносимых в процессе кодирования входного сигнала.

Эффективность данного алгоритма динамической трансформации частотно-временного плана иллюстрируется на 5. На 5б для входного сигнала 5а показаны этапы работы алгоритма роста дерева ПДВП и формирование частотно-временного плана.

Как видно, по мере увеличения декомпозиции дерева ПДВП порог маскирования – порог едва различимых искажений приближается к такому же порогу, который был построен для предельного дерева преобразования  $CB - WPD$  (рис. 6).

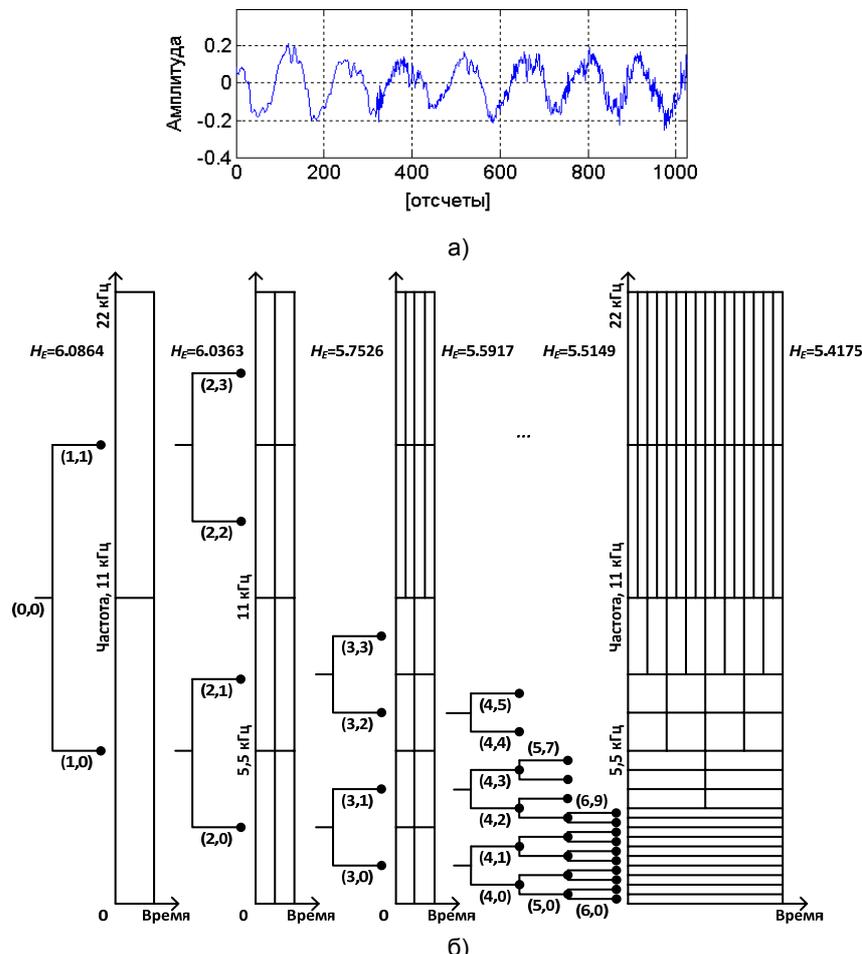


Рис.5. Формирование дерева ПДВП на основании алгоритма динамической трансформации частотно-временного плана

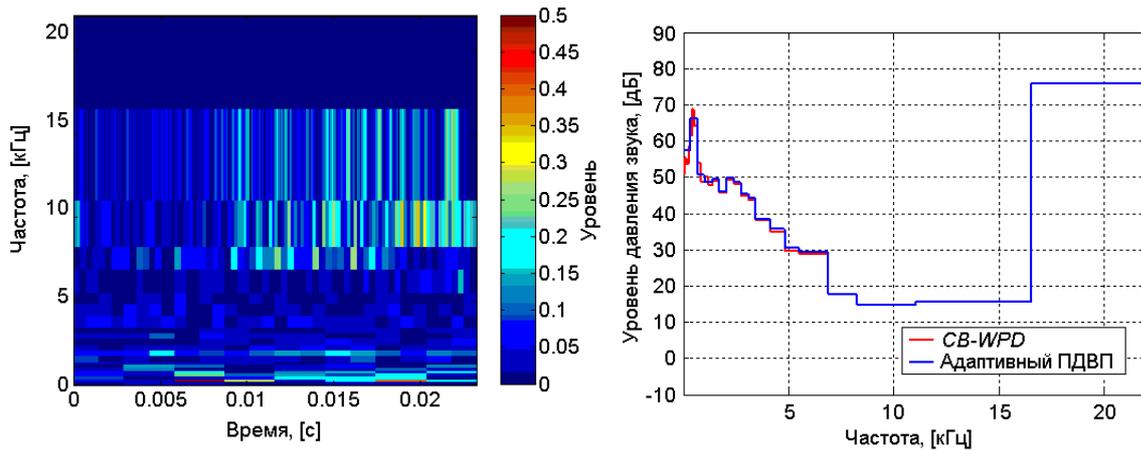


Рис.6. Спектрограмма и порог маскирования звукового сигнала

**Структура кодера звука на базе динамической трансформации алгоритма ПДВП**

На рис. 7 представлена структура перцептуального кодера звукового сигнала, ядром которой является блок «WPD», выполняющий реализацию адаптивного ПДВП. В темпе обработки сигналов для каждого текущего дерева  $E_i$  вычисляются пороги маскирования  $T_{l,n}$ , в соответствии с [12], значения перцептуальной энтропии  $PE_{l,n}$ , а также энтропии  $H_{E_i}$  структуры дерева ПДВП  $E_i$ . На основании данной информации в блоке «Split» рассчитываются параметры реконфигурации дерева ПДВП  $r_{l,n}$  [1]. Данный процесс осуществляется поступательно, без возвратов на меньшие масштабные уровни преобразования. Следовательно, весь вычислительный процесс идеально ложится на архитектуру параллельно-поточных процессоров [14]. Обработка звукового сигнала

ла в ПДВП кодере выполняется кадрами, и ввиду того, что преобразование осуществляется над каждым последующим кадром в новой структуре дерева преобразования, то для устранения фазовых разрывов, соседние кадры сигнала перекрываются на 1/8 своей длины. Следующим этапом работы кодера является квантование и кодирование вейвлет коэффициентов  $X_{l,n,k}: (l,n) \in E, k \in \mathbb{Z}$ , результирующего субоптимального дерева ПДВП  $E$ . Управление данным процессом осуществляется в блоке «Алгоритм размещения бит». Кодирование структуры дерева ПДВП  $E$  выполняется в блоке «Кодирования структуры дерева ПДВП». Принимая во внимание тот факт, что рост дерева осуществляется поступательно и изменчивость сигнала во времени более инерционна, чем время обработки, кодируются только изменения структуры дерева от фрейма к фрейму.

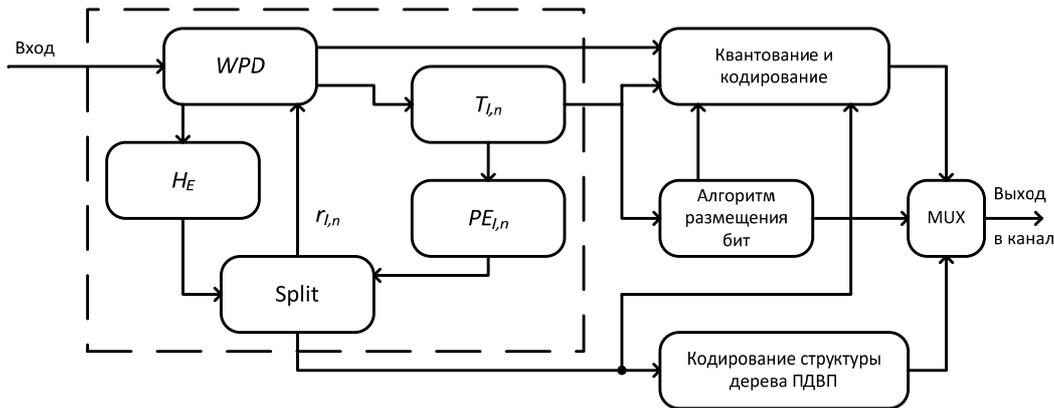


Рис. 7. Структура ПДВП кодера звука

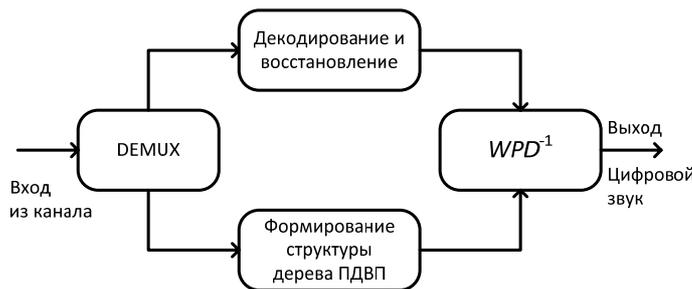


Рис. 8. Структура ПДВП декодера звука

Структура ПДВП декодера звука показана на рис. 8. Работа декодера выполняется в следующем порядке: входная информация разделяется на два потока данных в блоке «DEMUX»: один содержит кодированные вейвлет коэффициенты, а второй – код структуры дерева ПДВП; формирование структуры дерева ПДВП происходит в блоке «Формирование структуры дерева ПДВП» и осуществляется путём суммирования кодовой последовательности, описывающей изменение структуры дерева ПДВП, с текущей кодовой последовательностью структуры дерева ПДВП. В случае передачи первого фрейма, структура дерева ПДВП кодируется полностью в кодере. Реконструкция звукового сигнала выполняется в блоке « $WPD^{-1}$ » на основе обратного адаптивного ПДВП над декодированными и восстановленными вейвлет коэффициентами (блок «Декодирования и восстановления») в соответствии со структурой дерева ПДВП. Синтезированные кадры реконструированного сигнала умножаются на модифицированное треугольное окно и складываются для формирования выходной последовательности  $\hat{x}(n)$ .

### Вычислительная платформа системы кодирования

#### Структура вычислительной платформы

Структура вычислительной платформы кодера звука на основе ДТА-подхода включает управляющий процессор цифровой обработки сигналов (ЦПОС) и ПДВП процессор с реконфигурируемой архитектурой. На ЦПОС возложены следующие задачи: обработка вейвлет коэффициентов  $X_{l,n,k}$  в частотных полосах  $(l, n)$ , соответствующих текущей структуре дерева  $E_i$ ; вычисление психоакустической модели дерева  $(l, n) \in E_i: T_{l,n}$  и  $PE_{l,n}$ ; получение вектора реконфигурации ПДВП процессора  $r_{l,n}, (l, n) \in E_i$ ; формирование и разборка данных передачи  $[code(X_{l,n,k}, (l, n) \in E_i, k \in \mathbb{Z})]$  и приема  $\hat{X}_{l,n,k}(code), (l, n) \in E_i, k \in \mathbb{Z}$ , если используется дуплексный режим работы кодера звука. ПДВП процессор реализуется на параллельно-поточной архитектуре с динамической реконфигурацией. Выбор данной архитектуры обусловлен алгоритмом роста дерева ПДВП. Процессор работает в дуплексном режиме: конвейер кодера  $X_{l,n,k} = WPD_{l,n}(x_j(t)), (l, n) \in E_i, x_j(t) - j$ -й фрейм звукового сигнала и конвейер декодера  $\hat{x}_j(t) = WPD_{l,n}^{-1}(\hat{X}_{l,n,k}), (l, n) \in E_i, \hat{x}_j(t) - j$ -й фрейм реконструированного звукового сигнала. Длина конвейера параллельно-поточного

процессора определяется предельной структурой дерева ПДВП ( $CB - WPD$ ) (рис. 2). Анализ сложности реализации базовой операции ПДВП (пара КИХ фильтров нижних и верхних частот с децимацией), модулей коммутации и буферизации и требуемого аппаратного ресурса ПДВП процессора дает следующий результат: объём КЛБ для построения 8-ми модулей базовой операции ПДВП на MAC архитектуре – 2048, и для схем буферизации и переключения – 504. Расчетное число КЛБ с учётом 20% надбавки, обусловленной накладными расходами при реализации на конкретной ПЛИС, равно 3062 КЛБ.

Структурная схема вычислительной платформы представлена на рис. 9. Управляющий процессор реализован на ЦПОС TMS320C6713, а ПДВП процессор синтезирован на ПЛИС Xilinx семейства Spartan-3 XC3S2000 со структурой FPGA. На ПЛИС реализован также внешний цифровой интерфейс аппаратной платформы. Внешняя память SDRAM общим объёмом 8 Мбайт и ПЛИС подключаются к TMS320C6713 через 32-разрядный интерфейс внешней памяти EMIF. Интерфейс внешней памяти EMIF имеет четыре отдельные адресные области CE0-CE3. Flash память общим объёмом 2 Мбайта предназначена для хранения программ. Процессор использует Flash память в загрузочном режиме. После начальной инициализации модуля программа из Flash памяти автоматически, с помощью начального загрузчика, загружается во внутреннюю память управляющего процессора. Для работы со стерео звуком используется кодек TLC320AIC23. Кодек TLC320AIC23 подключается к ЦПОС с помощью двух последовательных каналов: McBSP0 – для доступа к конфигурационным регистрам, McBSP1 – для приема и передачи данных. Период частоты синхронизации может настраиваться за счет использования блоков DMC (Digital Clock Manager) ПЛИС. Для реализации внешнего интерфейса системы используются порты ввода/вывода ПЛИС.

Разработка проекта на ПЛИС реализована с помощью САПР Xilinx ISE. В данном случае конфигурационная последовательность для ПЛИС генерируется с помощью утилиты iMPACT пакета и загружается через интерфейс JTAG в конфигурационную Flash память платформы. Каждый раз после включения питания ПЛИС считывает эту последовательность из Flash памяти.

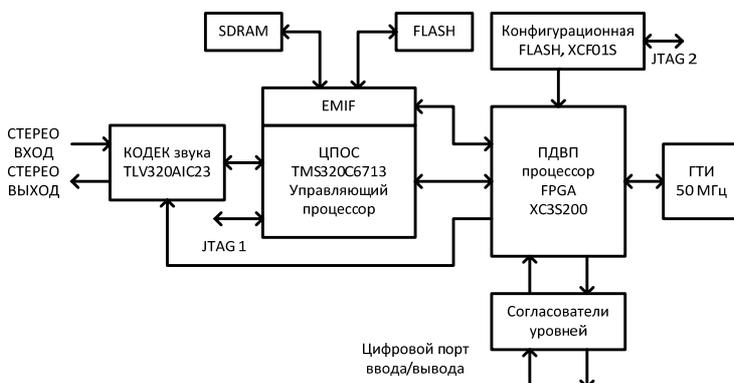


Рис. 9. Структурная вычислительной платформы



Рис. 10. Модуль цифровой обработки сигналов с реконфигурируемой логикой

Архитектурное решение вычислительной платформы позволяет получить следующие особенности: мощное вычислительное ядро с плавающей точкой для обработки данных с производительностью для 2400/1800 MIPS/MFLOPS; возможность распараллеливания задач обработки в двух-процессорной системе; реконфигурируемость платформы путем изменения конфигурационной последовательности ПЛИС и программ ЦПОС; возможность встраивания платформы в разнообразное оборудование за счет гибкой настройки внешнего интерфейса с помощью ПЛИС; использование стандартных средств конфигурирования и программирования. На рис. 10 представлен внешний вид модуля ЦОС с реконфигурируемой логикой.

**Организации вычислительного процесса**

Процесс кодирования звукового сигнала можно представить в виде последовательности следующих действий: прием очередного кадра сигнала длительностью 23 мс (1024 отсчета при частоте дискретизации 44.1 кГц); передача в ПЛИС вектора реконфигурации ; анализ входного сигнала; расчет параметров психоакустической модели на ЦПОС; формирование вектора реконфигурации ПДВП процессора задачи анализа; кодирование вейвлет коэффициентов на управляющем процессоре; формирование пакетов передачи коди-

ванной информации через цифровой порт.

Следует отметить, что работа кодера звука в реальном масштабе времени обуславливает достаточно жесткую синхронизацию вычислительных процессов управляющего процессора и ПДВП процессора. Так, перестройка ПДВП процессора осуществляется на основе формирования вектора реконфигурации согласно алгоритму динамической трансформации дерева ПДВП. При этом для фреймов, обрабатываемых согласно текущей структуре дерева на ступенях ПДВП процессора, управляющий процессор осуществляет выполнение процедур расчета порогов маскирования, оценки перцептуальной энтропии и информативности структуры дерева ПДВП. Распределение вычислительного ресурса между процедурами в управляющем процессоре показано на рис. 11. Время выполнения процедуры 1 (алгоритм вычисления порогов маскирования (рис. 12)) и процедуры 2 (оценка перцептуальной энтропии (рис. 12)) не постоянно, оно зависит от числа ступеней конвейера ПДВП процессора (кодер), задействованных в обработке текущего фрейма входного сигнала (см рис. 13). Этой же причиной обусловлено неперiodическое выполнение процедур 4 и 5 (квантование и кодирование вейвлет коэффициентов).

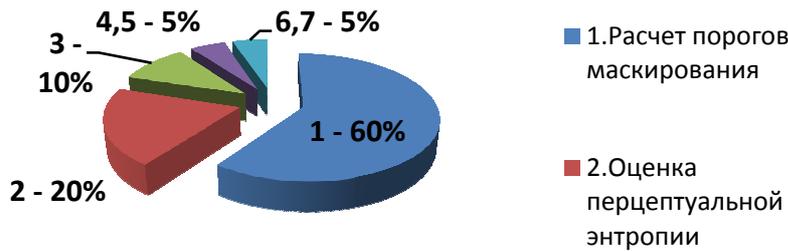


Рис. 11. Распределение вычислительного ресурса между процедурами в управляющем процессоре

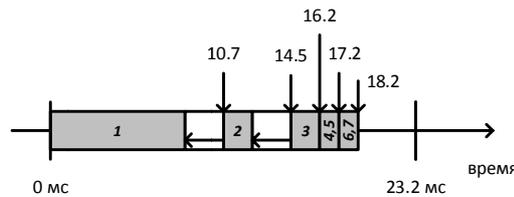


Рис. 12. Время выполнения процедур в управляющем процессоре

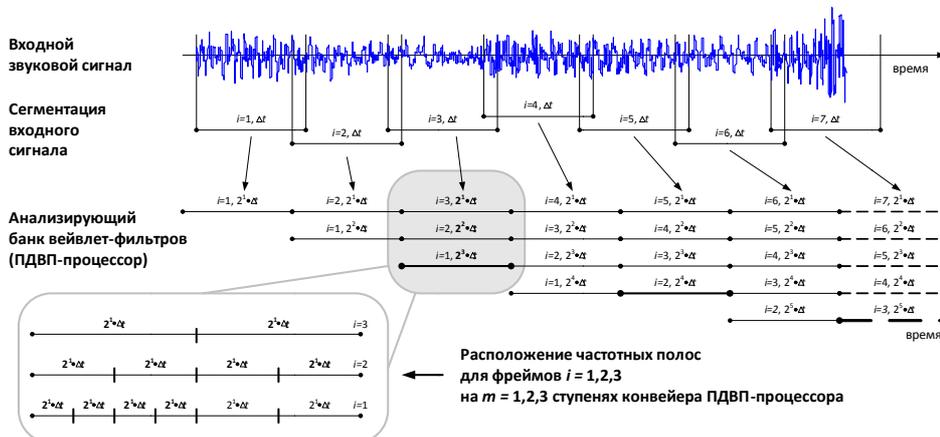


Рис. 13. Анализа входного сигнала в ПДВП процессоре

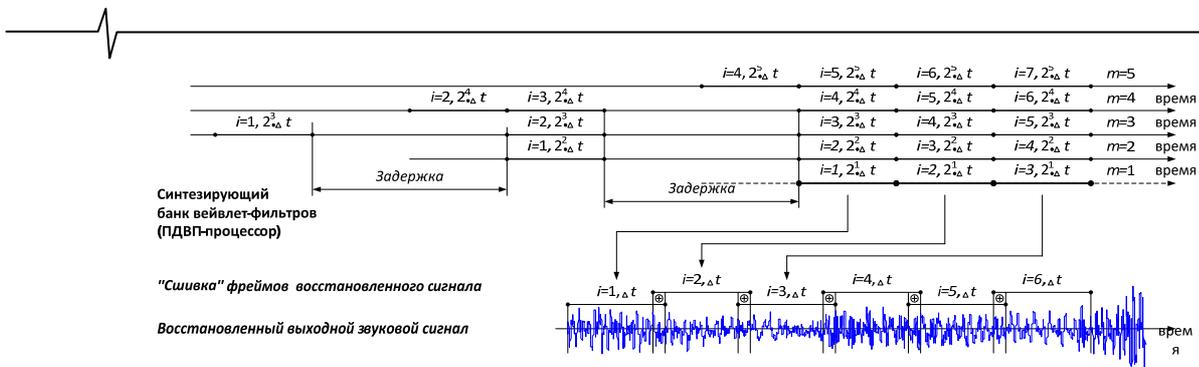


Рис. 14. Синтез выходного сигнала в ПДВП процессоре

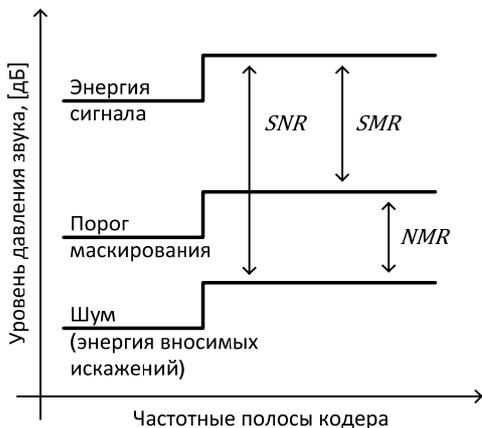


Рис. 15. Объективные оценки качества сигнала

В режиме декодирования управляющий процессор декодирует и восстанавливает фреймы из принимаемой кодовой последовательности, выполняя процедуры 6,7 (декодирование и восстановление), загружает восстановленный фрейм  $i$  на соответствующую ступень  $m$  конвейера ПДВП процессора (декодера). Переход фрейма  $i$  на следующую ступень процессора выполняется тогда, когда управляющий процессор примет следующий фрейм  $i + 1$ , который необходимо будет загрузить уже на задействованную ступень конвейера ПДВП процессора (декодера) (см. рис.14). Для согласования работ, выполняемых на каждой ступени конвейера процессора, необходимо ввести задержку, кратную времени обработки одного фрейма на одной ступени, тем самым будет обеспечиваться ритмичная работа параллельно-поточной структуры ПДВП процессора (декодер).

### Результаты тестирования

Ниже представлены результаты исследований данного субполосного перцептуального ПДВП кодера звука. Для разных сигналов (шумовых, тональных, ритмичных) оценивалась возможная минимальная скорость передачи (кбит/с), максимальная степень компрессии, среднее количество бит на отсчет входного звукового сигнала, а также определялись объективные оценки качества:  $NMR_{total}$ ,  $NMR_{SEG}$  – общее и сегментное соотношение шума к порогу маскирования, показывающее границу порога восприятия артефакторов на слух,  $SB$  – ширина полосы частот реконструированного сигнала,  $MF$  – процент искаженных фреймов реконструированного сигнала (см. рис.15) [15].

В таблице 1 приведены результаты экспериментов в сравнении с популярным кодером звука MPEG-1 уровень III [16], минимальная скорость передачи которого 64 кбит/с. Анализ полученных данных показывает, что минимальная скорость передачи для ПДВП кодера звука меняется от 36 кбит/с для тональных, ритмичных звуковых сигналов до 45 кбит/с для шумовых сигналов. При этом, оценка  $NMR_{total}$  меняется от -4,037 дБ до -8,857 дБ, т.е. граница порога восприятия минимум в два раза ниже порога маскирования. Для сравнения у MPEG-1 данная оценка позитивная, т.е. энергия воспринимаемых шумов не менее чем в два раза больше порога маскирования и, соответственно, практически каждый фрейм реконструированного сигнала содержит искажения и,  $MF_{max} \cong 99,78\%$ . Это обусловлено скорее тем, что ширина частотной полосы  $SB$  реконструированного сигнала достаточно мала:  $SB_{min} = 11,89$  кГц, против 22,05 кГц в оригинальном сигнале (см. рис.16 – рис.18).

Таблица 1.

Оценки перцептуального ПДВП кодера звука и MPEG-1 уровень III

Тестовый материал $f_s = 44.1$ кГц, 16 бит ИКМ	Тип кодера	Мин. скорость передачи, [кбит/с]	Макс. степень компрессии	Среднее кол., [бит/отсчет]	Ширина частотной полосы SB, [кГц]	$NMR_{total}$ , [дБ]	$NMR_{SEG}$ , [дБ]	МГ, % искаженных фреймов
Boney-M 15с.	ПДВП-кодер	45	15.68	1.02	18.49	-8.297	-0.8699	27.04
	MPEG-1 ур. III	64	11.02	1.45	15.13	4.430	0.4076	99.85
Classic 10с.	ПДВП-кодер	42	16.8	0.95	17.62	-6.495	-0.7987	43.45
	MPEG-1 ур. III	64	11.02	1.45	12.58	3.284	0.3145	99.78
ABBA "Take a chance on me" 2с.	ПДВП-кодер	36	19.7	0.81	18.58	-4.037	-0.5844	66.25
	MPEG-1 ур. III	64	11.02	1.45	15.39	3.720	0.3312	98.70

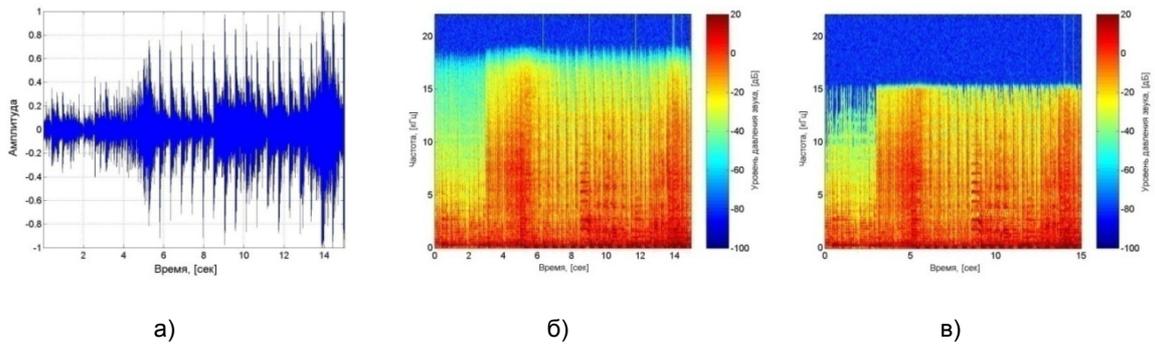


Рис. 16. Звуковой сигнал «VoiceM-2000»: а) оригинальный сигнал; б) спектрограмма реконструированного сигнала (ПДВП-кодер); в) спектрограмма реконструированного сигнала (MPEG-1 уровень III)

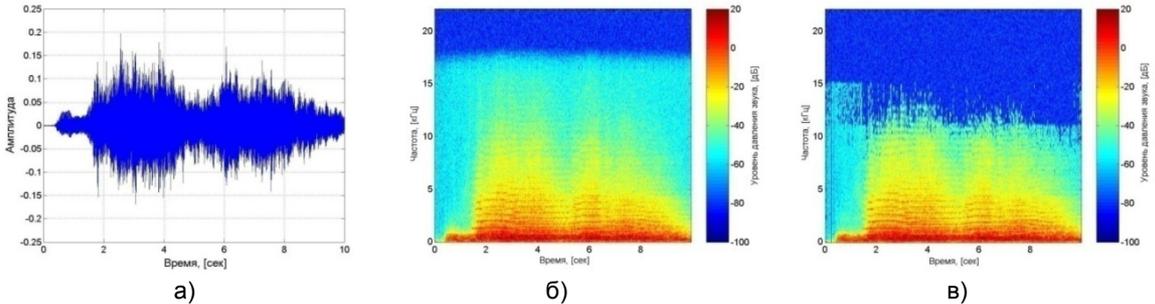


Рис. 17. Звуковой сигнал «Classic»: а) оригинальный сигнал; б) спектрограмма реконструированного сигнала (ПДВП кодер); в) спектрограмма реконструированного сигнала (MPEG-1 уровень III)

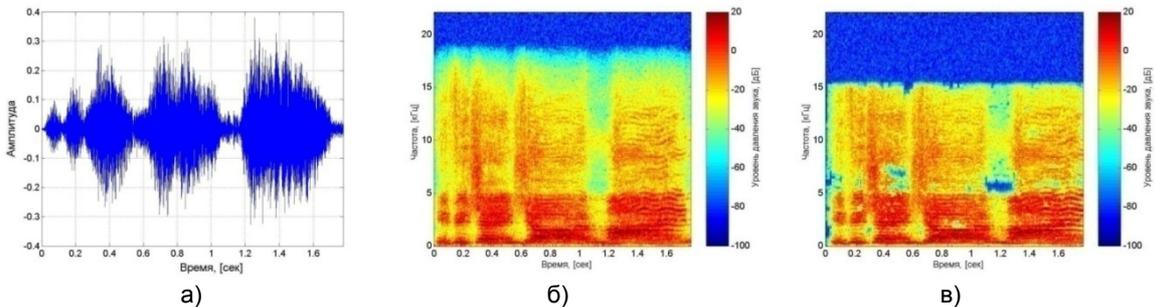


Рис.18. Звуковой сигнал «ABBA "Take a chance on me"»: а) оригинальный сигнал; б) спектрограмма реконструированного сигнала (ПДВП кодер); в) спектрограмма реконструированного сигнала (MPEG-1 уровень III)

## Заключение

Таким образом, объективные оценки, полученные при экспериментальных исследованиях перцептуального ПДВП кодера звуковых сигналов, показывают, что реконструированный сигнал имеет достаточно хорошее качество, соответствующее требованиям стандарта ITU-R PEAQ [17] при высокой степени компрессии в 15 и более раз. Причём, для  $NMR_{total}$  меньше -10 дБ считается, что восстановленный сигнал не содержит слышимых артефактов. Для данного субполосного перцептуального ПДВП кодера звука данный показатель  $\approx -9$  дБ, но для схемы измерения  $NMR_{total}$ , главным достоинством которой, а также оценок  $MF$  и  $SB$  является то, что в отличие от [15], они вычисляются в вейвлет области и при тестировании кодеров звука на базе адаптивных ПДВП дают минимальную погрешность вследствие отсутствия взаимных частотно-вейвлетных преобразований [18,19].

## Список литературы

1. Al.Petrovsky, A.Petrovsky, Dynamic algorithm transforms for reconfigurable real-time audio coding processor // The proc. of the Intern. Conference on Parallel Computing in Electrical Engineering, PARELEC'02, Warsaw, Poland, Sep. 22-25, 2002, - IEEE Press, Los Alamitos, California, 2002. – pp.422-424.
2. A. Spanias, T. Painter, V. Atti, Audio signal processing and coding. - Hoboken, NJ: John Wiley & Sons, Inc., 2007. – 464 p.
3. M. Bosi, R. Goldberg, Introduction to digital audio coding and standards. – Springer, Inc., 2003. – 434 p.
4. J.A. Kovalgin, D.A. Hitrov, M.V. Zyrianov. Advanced perceptual digital audio coding algorithm // In the proc. of the AES 17th intern. Conference "High-quality audio coding", Florence, Italy. – 1999. – pp.233-243.
5. N. Jayant, P. Noll. Digital coding of waveforms: principles and applications to speech and video. – Englewood Cliffs, NJ: Prentice-Hall, 1984. – 964 p.
6. J.D. Johnston. Transform coding of audio signals using perceptual noise criteria. IEEE Trans. on Select. Areas Commun. – Feb. 1988, – vol. 6. – pp. 314-323.
7. D. Sinha, A. Tewfik. Low bit rate transparent audio compression using adapted wavelets. IEEE Trans. on Signal Processing.- Dec. 1993, vol. 41 - pp. 3463-3479.
8. M. Vetterli, J. Kovacevic. Wavelets and subband coding. - Englewood Cliffs, NJ, Prentice-Hall, 1995. – pp. 488.

9. M.V. Wickerhauser. Adaptive Wavelet Analysis from Theory to Software. – A.K. Peters Ltd., Massachusetts, 1994. – 486 p.
10. E. Zwicker, H. Fastl. Psychoacoustics: facts and models. – Berlin, Springer-Verlag, 1990. – 380 p.
11. Al.Petrovsky, D.Krahe, A.A.Petrovsky. Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfiguration system // AES, Convention paper 5778, 114th Convention, 2003 March 22-25, Amsterdam, The Netherlands. – 22 p.
12. Ал.А. Петровский. Построение психоакустической модели в области вейвлет-коэффициентов для перцептуальной обработки звуковых и речевых сигналов. – Речевые технологии. - № 4, 2008.- Москва. – С.61-71.
13. P.Srinivasan, L.H.Jamieson. High-quality audio compression using an adaptive wavelet packet decomposition and psychoacoustic modeling. IEEE Trans. on Signal Processing, vol. 46, No 4, April 1998. - pp.1085-1093.
14. Al.A.Petrovsky, A.A.Petrovsky. Reconfigurable processing: the solution to audio coding in real-time // The proc. of the 9th International conference "Mixed design of integrated circuits and systems", MIXDES'2002, 20-22 June 2002, Wroclaw, Poland. – pp.629-634.
15. K. Brandenburg, T. Sporer. "NMR" and "Masking Flag": evaluation of quality using perceptual criteria // In the proc. of the 11th intern. AES conference "Test and measurement conference", Portland, Oregon. – 1992. – pp.169-179.
16. Ю.А. Ковалгин, Э.И. Вологдин. Цифровое кодирование звуковых сигналов. Санкт-Петербург, Корона принт, 2004. - 240 с.
17. C. Colomes, C. Schmidmer, T. Thiede, W. Treurniet. Perceptual quality assessment for digital audio: PEAQ – the new ITU standard for objective measurement of the perceived audio quality // In the proc. of the AES 17<sup>th</sup> Intern. Conference "High-quality audio coding", Florence, Italy. – 1999. – pp.337 – 351.
18. Ал.А. Петровский. Объективная оценка качества восстановленного аудиосигнала перцептуальным ПДВП-кодером на базе периферийной модели уха человека // Сборник трудов 5-й Международной научной конференции "Цифровая обработка сигналов и её применение" (DSPA'2003), т.2, Москва, Россия, 2003. – С. 123-126.
19. Ал.А. Петровский. Субполосное перцептуальное кодирование звуковых и речевых сигналов.- Минск, «Беспринт», 2008. – 218 с.

### ***Уважаемые коллеги!***

***Предлагаем вам принять участие в формировании тематических выпусков журнала «Цифровая обработка сигналов» и размещению рекламы продукции (услуг) Вашей фирмы на его страницах. В случае положительного решения просим представить в редакцию журнала Ваши предложения по плановому размещению информационных материалов и макет рекламы продукции (услуг) Вашей фирмы с указанием желаемого её месторасположения: обложка (2-я, 3-я или 4-я стр.), цветная внутренняя полоса (объем полосы).***

В 2010 году планируется выпуск 4-х номеров журнала (тираж до 700 экз.). Журнал распространяется по подписке через агентство «Роспечать» в России, СНГ и странах Балтии (индекс 82185), а также на отраслевых всероссийских и международных Выставках.

Размещение рекламы Вашей фирмы на страницах журнала «Цифровая обработка сигналов» на плановой основе (не менее 2-х полных или 4-х половинчатых рекламных полос в течение года) предоставит Вам следующие возможности и права:

1. Первоочередное право расположения рекламных материалов на всех обложках (кроме 1-й) и страницах журнала.
2. Публикация представленных Вами рабочих (рекламных) материалов (статей) объемом до 6 полос в каждом очередном номере (в счет оплаченной рекламы).
3. Установка баннера Вашего сайта (или логотипа вашей организации) на 1-й странице сайта журнала «Цифровая обработка сигналов» ([www.dspra.ru](http://www.dspra.ru)) в течение всего года, что привлечет внимание к продукции (услугам) Вашей фирмы новых участников на рынке DSP-технологий (ежедневно фиксируется до 100 и более посещений сайта [www.dspra.ru](http://www.dspra.ru)).
4. Предоставление до 10 экз. очередного выпуска журнала.

Ориентировочная стоимость рекламных услуг:

4-я (внешняя) страница цветной обложки - 20 тысяч рублей.

2-я и 3-я (внутренние) страницы цветной обложки - 15 тысяч рублей.

1/2 цветной внутренней полосы - 7 тысяч рублей.

1/2 черно-белой внутренней полосы – 1 тысяча рублей.

*Ждем Ваших предложений.*

С наилучшими пожеланиями, зам. главного редактора  
д.т.н., профессор Витязев Владимир Викторович

**Предложения прошу направлять по адресу: E-mail: [tor@rgta.ryazan.ru](mailto:tor@rgta.ryazan.ru) или [info@dspra.ru](mailto:info@dspra.ru)**