

Учреждение образования  
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ»

УДК 004.032.6 + 004.383.3

ЛИВШИЦ  
Михаил Зенадьевич

**КОМБИНИРОВАННАЯ СИСТЕМА ПЕРЦЕПТУАЛЬНОГО  
ШИРОКОПОЛОСНОГО КОДЕРА РЕЧИ С ПРЕПРОЦЕССОРОМ  
РЕДАКТИРОВАНИЯ ШУМА**

АВТОРЕФЕРАТ  
диссертации на соискание ученой степени  
кандидата технических наук

по специальности 05.13.05  
«Элементы и устройства вычислительной техники и систем управления»

Минск 2007

Работа выполнена в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель – **ПЕТРОВСКИЙ Александр Александрович**, доктор технических наук, профессор, заведующий кафедрой электронных вычислительных средств, учреждение образования «Белорусский государственный университет информатики и радиоэлектроники»

Официальные оппоненты: **АПАНАСОВИЧ Владимир Владимирович**, доктор физико-математических наук, профессор, заведующий кафедрой системного анализа, учреждение образования «Белорусский государственный университет», директор Института бизнеса и менеджмента технологий

**БОРИСКЕВИЧ Анатолий Антонович**, кандидат технических наук, доцент кафедры сетей и устройств телекоммуникаций, учреждение образования «Белорусский государственный университет информатики и радиоэлектроники»

Оппонирующая организация – Научно-исследовательское республиканское унитарное предприятие «Научно-исследовательский институт средств автоматизации»

Защита состоится 6 декабря 2007г. в 14 часов на заседании совета по защите диссертаций Д 02.15.01 при учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, г. Минск, ул. П. Бровки, 6, БГУИР, 1 учебный корпус, ауд. 232, email: dissovet@bsuir.by, тел. 293-89-89.

## **ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ**

Современная вычислительная техника характеризуется развитием мобильных проблемно-ориентированных вычислительных платформ, функционирующих в реальном масштабе времени и базирующихся на специализированных процессорах и системах, оптимизированных для решения тех или иных задач в специальной области применения. По мере совершенствования элементной базы вычислительной техники спектр этих процессоров и систем расширяется при одновременном сужении их специализации и увеличении масштабов применения.

Значительно возросшее внимание к повышению качества звучания реконструированного речевого сигнала обуславливает актуальность исследования в рамках совершенствования методов и аппаратно-программных средств кодирования аудиосигналов в мобильных мультимедийных системах реального времени. Развитие этих работ в настоящее время идет по пути построения перцептуальных широкополосных кодеров речевых сигналов на основе подходящего частотно-временного преобразования и модели речеобразования. Один из перспективных подходов в организации вычислительных платформ для решения задач компрессии речевых и аудиосигналов в реальном масштабе времени базируется на адаптации алгоритма обработки и архитектуры процессора под кодируемый сигнал и пропускную способность коммуникационного канала. При этом обеспечивается минимальная избыточность в представлении сигнала и повышение показателя быстродействия/качество звучания реконструированного речевого сигнала.

Вышеизложенное подчеркивает актуальность исследований, разработки алгоритмов кодирования и структуры перцептуального компрессора широкополосных речевых сигналов с динамически реконфигурируемой структурой на основе работы психоакустической модели на базе дискретного преобразования Фурье с неравномерным разложением коэффициентов по частотной оси (WDFT). Важность проблемы, совершенствование методов и алгоритмов обработки широкополосных речевых сигналов в реальном масштабе времени и их отображение на архитектуру вычислительной платформы определили цель работы, решаемые задачи и методы исследований.

### **Связь работы с крупными научными программами и темами**

Диссертационная работа выполнена в соответствии с научно-техническими заданиями и планами работ кафедры «Электронные вычислительные средства» и НИЛ 3.1 «Мультипроцессорные системы реального времени» Белорусского государственного университета информатики и радиоэлектроники в ходе составной части научно-исследовательской работы (НИР) «Разработать процедуры сжатия речевой информации, обеспечивающие коммерческое качество восстановленной речи» (ГБ 01-1089), выполненной в рамках Государственной научно-технической

программы «Развитие методов и средств системы комплексной защиты информации»; НИР «Разработка широкополосного компрессора речевых сигналов с психоакустической мотивацией на базе мультиполосного CELP-алгоритма для низкоскоростных каналов связи» (ГБЦ 05-3040) и НИР «Аппаратно-программная реализация широкополосного компрессора речевых сигналов с психоакустической мотивацией для применения в системах мультимедиа реального времени на базе ЦПОС TMS320C6x» (ГБЦ 06-3097), выполненных в рамках грантов Министерства образования РБ, а также НИР «Разработка низкоскоростного речепреобразующего устройства с препроцессором редактирования шума и компенсатором эхо-сигнала» (х/д 05-1160).

### **Цель и задачи исследования**

Целью работы является разработка методов и устройств компрессии широкополосных речевых и аудиосигналов, а также снижение уровня шумов окружающей среды в мультимедийных системах реального времени, базирующихся на психоакустических особенностях восприятия человеком акустической информации.

Поставленная цель работы определяет следующие задачи исследования:

1. Анализ подходов построения перцептуальных широкополосных кодеров речевой информации с целью выбора методов анализа речевого сигнала и его компрессии.
2. Разработка метода и алгоритма перцептуальной обработки речевых сигналов для применения в мультимедийных устройствах реального времени.
3. Разработка метода квантования параметров модели речеобразования для их представления минимальным количеством бит.
4. Аппаратно-программная реализация устройства компрессии речевых сигналов с препроцессором редактирования шума: комбинированная система, мультиполосная CELP-модель с перцептуальной обработкой субполос, реальный масштаб времени.

Объектом исследования является устройство широкополосного CELP-кодера с многополосным возбуждением и перцептуальной обработкой сигнала в субполосах.

Предмет исследования – методы компрессии широкополосных речевых и аудиосигналов на основе CELP-модели с мультиполосным возбуждением, использующей психоакустическую модель для более детального анализа и кодирования субполос.

Совокупность использованных в диссертационной работе средств исследований базируется на аппарате линейной алгебры, математического анализа, численных математических методов, теории вероятностей и математической статистики, математического моделирования, психоакустики, цифровой обработки сигналов, методах проектирования проблемно-ориентированных вычислительных средств.

## **Положения, выносимые на защиту**

1. Модифицированный алгоритм формирования (обучения) кодовых книг для различных параметров модели речеобразования, обеспечивающий более быструю сходимость.

2. Метод субполосной декомпозиции частотной шкалы, учитывающий частотную избирательность слуховой системы человека и формантную структуру речи.

3. Алгоритм динамической трансформации банка субполосных кодовых книг под частотно-временной план фрейма кодируемого сигнала на базе субполосной перцептуальной энтропии.

4. Метод квантования структуры банка кодовых книг, обеспечивающий «прозрачное» кодирование при минимальном количестве бит.

5. Структура комбинированной системы кодирования широкополосного речевого сигнала на основе совмещения препроцессора редактирования шума и широкополосного CELP-кодера с реконфигурируемым банком кодовых книг, построенная на общей психоакустической модели на базе ДПФ с неравномерным разложением коэффициентов по частотной оси (WDFT).

6. Особенности аппаратно-программной реализации устройства комбинированной системы кодирования широкополосного речевого сигнала на основе параллельно-поточной обработки на однокристалльном ЦПОС Texas Instruments TMS320C6713 с VLIW-архитектурой и поддержкой SIMD-команд, функционирующего в реальном масштабе времени в составе комплекса мультимедийных устройств для сжатия аудиосообщений.

## **Личный вклад соискателя**

Результаты, приведённые в диссертации, получены либо соискателем лично, либо при его непосредственном участии. Вклад научного руководителя Петровского А.А. связан с постановкой цели и задач исследования. В публикациях с соавторами вклад соискателя определяется рамками излагаемых в диссертации результатов. Соавторы работ принимали участие в получении объективных и субъективных оценок качества синтезированного речевого сигнала.

## **Апробация результатов диссертации**

Результаты, полученные в ходе выполнения исследований, докладывались и обсуждались на VI и VII Международной летней школе-семинаре студентов и аспирантов «Современные информационные технологии» (Браслав, Беларусь, 2003, 2005); 8-й международной конференции «Современные средства связи» (Нарочь, Беларусь, 2003); VI и VII Международных конференциях «Цифровая обработка сигналов и ее применение» (Москва, Россия, 2004, 2005); the 8th International Conference «Pattern Recognition and Information Processing» (Minsk, Belarus, 2005);

International Conference «Computer as a tool» EUROCON 2005 (Serbia & Montenegro, Belgrade, 2005); XI International Conference «SPEECH AND COMPUTER» (St.Petersburg, Russia, 2006); XI Symposium AES «New Trends in Audio and Video» (Bialystok, Poland, 2006); the 9th International Conference «Pattern Recognition and Information Processing» (Minsk, Belarus, 2007); научно-технических конференциях БГУИР (Минск, Беларусь, 2004–2007).

Результаты диссертационной работы внедрены и используются в следующих учреждениях:

1. Государственном центре безопасности информации при Президенте Республики Беларусь;
2. Управлении правительственной связи КГБ Республики Беларусь;
3. В учебном процессе специальности «Электронные вычислительные средства» (БГУИР).

### **Опубликованность результатов диссертации**

По тематике представленной диссертационной работы опубликовано 14 печатных работ (6 – в научных журналах, 7 – в статьях и сборниках трудов научно-технических конференций, 1 глава в книге). Суммарный объем публикаций составляет 7 авторских листов. Результаты работы включены в 5 отчетов о НИР.

### **Структура и объём диссертации**

Диссертационная работа состоит из введения, пяти глав, заключения и четырех приложений. **Во введении** показана актуальность темы диссертационной работы, формулируются цель и задачи исследования. В **главе 1** представлен анализ архитектур, структур и концепций построения современных перцептуальных широкополосных кодеров речевых сигналов. Предлагается модель и структура широкополосного перцептуального CELP-кодера с мультиполосным возбуждением и обработкой субполос в порядке убывания их перцептуальной важности. Осуществляется предварительный выбор аппаратной платформы для реализации системы в реальном масштабе времени. В **главе 2** осуществляется формирование структуры перцептуального широкополосного CELP-кодера согласно концепции, изложенной в главе 1. Предлагается психоакустическая модель на базе WDFТ, осуществляется синтез алгоритма управления структурой банка субполосных кодовых книг, реализуется интеграция препроцессора редактирования шума в структуру компрессора. **Глава 3** посвящена формированию, оптимизации и квантованию структуры банка кодовых книг векторов возбуждения. В **4-й главе** дано описание алгоритмического обеспечения комбинированной системы кодирования широкополосного речевого сигнала, обсуждаются особенности аппаратной реализации данной системы в реальном масштабе времени на базе ЦПОС TMS320C6713. Представлены характеристики отладочного модуля и разработанной платформы с реконфигурируемой логикой. В **5-й главе**

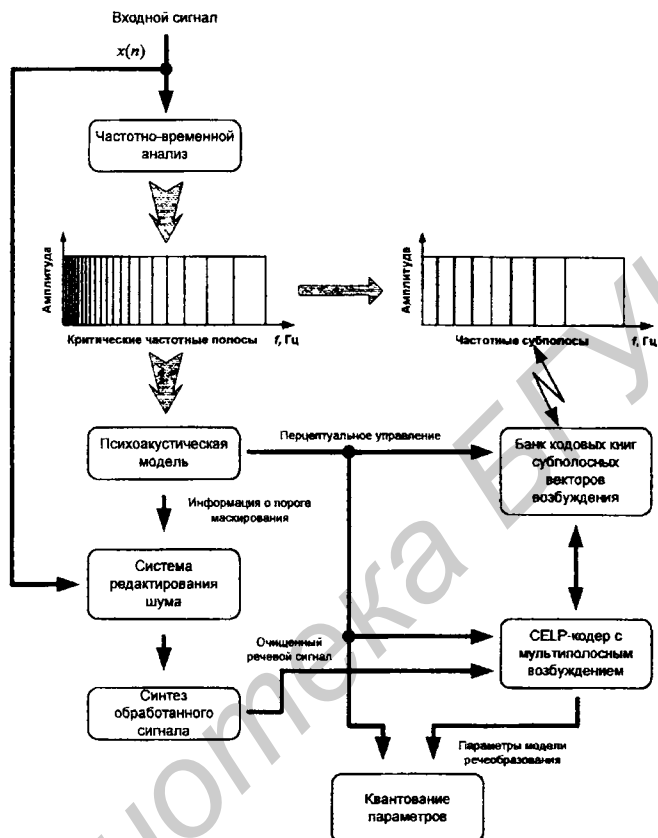
приведены результаты субъективных и объективных оценок качества восстановленного в декодере речевого сигнала, а также оценки качества и эффективности системы редактирования шума. В приложениях представлены примеры отображения С-кода на архитектуру ЦПОС TMS320C67х, методики определения объективных оценок качества звучания реконструированной речи, вычислительные аспекты WDFT, а также акты внедрения результатов диссертационной работы.

Работа представлена на 226 страницах машинописного текста, в том числе основная часть – на 181 странице (100 страниц текста, 97 рисунков на 69 страницах, 18 таблиц на 12 страницах), 4 приложения на 30 страницах, библиографический список из 180 наименований на 15 страницах, включая 14 публикаций автора на 2 страницах.

## СОДЕРЖАНИЕ РАБОТЫ

В главе 1 представлен анализ концепций построения современных перцептуальных широкополосных кодеров речи. Проводится обзор основных методов кодирования речевой информации: кодеры на основе частотно-временных преобразований, субполосное и параметрическое кодирование, а также гибридные схемы перцептуальных компрессоров. Предлагается модель и структура широкополосного перцептуального CELP-кодера с мультиполосным возбуждением и обработкой субполос в порядке убывания их перцептуальной важности. Ядром данного кодера является банк субполосных кодовых книг с многоуровневой организацией, позволяющей моделировать частотно-временной план фрейма кодируемого сигнала [1,2,8]. Для расширения диапазона мобильных мультимедийных приложений, в которых возможно применение разрабатываемой системы, в ее состав вводится препроцессор редактирования шума [7,14]. Представляется обобщенная модель комбинированной системы кодирования речи (рисунок 1). Модель формализуется следующим образом. Пусть имеется множество обрабатываемых (кодируемых) субполос  $B$  с мощностью равной  $C(B)$ , кодер имеет  $C(B)$ -полосную кодовую книгу с длиной субполосных векторов, равной  $L$  отсчетов. Длина обрабатываемого фрейма сигнала равна  $N$  отсчетов, длина векторов книги соответствует длине субфреймов LTP-анализа. Количество субфреймов равно  $\lambda_c$ . Номер субфрейма  $\lambda$ . Индексы коэффициентов усиления субполосных кодовых векторов  $b$ -й субполосы для всех субфреймов определяются как вектор  $g_b = \{g_b^{(1)}, g_b^{(2)}, \dots, g_b^{(\lambda_c)}\}$ , а индексы самих векторов возбуждения как  $i_b = \{i_b^{(1)}, i_b^{(2)}, \dots, i_b^{(\lambda_c)}\}$ . Набор индексов оптимальных коэффициентов усиления и субполосных векторов возбуждения обозначим как  $g_{b,opt} = \{g_{b,opt}^{(1)}, g_{b,opt}^{(2)}, \dots, g_{b,opt}^{(\lambda_c)}\}$  и  $i_{b,opt} = \{i_{b,opt}^{(1)}, i_{b,opt}^{(2)}, \dots, i_{b,opt}^{(\lambda_c)}\}$ , соответственно. Тогда синтезированный субполосный сигнал для субфрейма  $\lambda$  и его коэффициент

усиления определяются как функции от соответствующих индексов:  $\hat{s}_w^{(\lambda)}$  и  $G_b(g_b^{(\lambda)})$ .



**Рисунок 1 – Модель комбинированной системы кодирования широкополосных речевых сигналов**

Если обозначить фрейм входного сигнала длиной  $N$  отсчетов как  $s(n)$ , а субфрейм  $\lambda$  длиной  $L$  отсчетов –  $s^{(\lambda)}(l)$ , то  $s(n) = \{s^{(1)}(l), s^{(2)}(l), \dots, s^{(\lambda)}(l)\}$ . Фрейм взвешенного оригинального сигнала можно представить как набор субфреймов  $\tilde{s}_w(n) = \{\tilde{s}_w^{(1)}(l), \tilde{s}_w^{(2)}(l), \dots, \tilde{s}_w^{(\lambda)}(l)\}$ , операцию взвешивания через оператор  $LP[\cdot]$ , вектор ошибки квантования предыдущей субполосы как  $e_{w_b^{(\lambda-1)}}(n) = \{e_{w_b^{(\lambda-1)}}^{(1)}(l), e_{w_b^{(\lambda-1)}}^{(2)}(l), \dots, e_{w_b^{(\lambda-1)}}^{(\lambda)}(l)\}$ , остаточный сигнал после кодирования текущей субполосы представить  $e_{w_b^{(\lambda)}}(n) = \{e_{w_b^{(\lambda)}}^{(1)}(l), e_{w_b^{(\lambda)}}^{(2)}(l), \dots, e_{w_b^{(\lambda)}}^{(\lambda)}(l)\}$ , то процесс определения оптимального субполосного сигнала возбуждения может быть записан в следующем виде [9,11]:



$$\begin{aligned} \bar{s}_w(n) &= LP[s(n)], \quad n=1,2,\dots,N, \quad 1 \leq \lambda \leq \lambda_c, \\ e_w^{(\lambda)}(l) &= \bar{s}_w(k), \quad l=1,2,\dots,L, \quad k=(\lambda-1) \cdot L+1, (\lambda-1) \cdot L+2, \dots, \lambda \cdot L, \\ \{g_{b,opt}^{(\lambda)}, i_{b,opt}^{(\lambda)}\} &= \arg \min_{g_b^{(\lambda)}, i_b^{(\lambda)}} e_w^{(\lambda)}(l) = e_{w_{(b-1)}}^{(\lambda)}(l) - G_b(g_b^{(\lambda)}) \cdot \hat{s}_{w_b}(i_b^{(\lambda)}), \\ & 1 \leq b \leq C(B), \\ e_w^{(\lambda)}(l) &= e_{w_0}^{(\lambda)}(l) - \sum_{b=1}^{C(B)} G_b(g_{b,opt}^{(\lambda)}) \cdot \hat{s}_{w_b}(i_{b,opt}^{(\lambda)}), \end{aligned}$$

$$e_w(n) = \{e_w^{(1)}(l), e_w^{(2)}(l), \dots, e_w^{(\lambda_c)}(l)\},$$

где  $e_w(n)$  – остаточный сигнал (ошибка) после кодирования текущего фрейма взвешенного оригинального сигнала  $\bar{s}_w(n)$ .

В качестве основы для реализации аппаратной платформы устройства компрессии речи предлагается использовать ЦПОС Texas Instruments TMS320C6713 с VLIW-архитектурой и поддержкой SIMD-инструкций, который позволит реализовать в реальном масштабе времени алгоритмы компрессии речи, а также процедуры редактирования шума и шифрования [7].

В **главе 2** описывается поэтапное проектирование структуры перцептуального широкополосного CELP-кодера согласно концепции, изложенной в главе 1. Рассматриваются базовые принципы психоакустики. Осуществляется формирование субполосных групп на основе анализа их перцептуальной значимости и формантной структуры речи. Предлагается схема декомпозиции (группирования) всей частотной шкалы (21 барка) в 8 субполос, обрабатываемых кодером. Для устранения перцептуальной избыточности и статистически равнозначной обработки субполос в качестве частотно-временного преобразования применяется WDFT, которое сочетает в себе достоинства методов, использующих банки цифровых фильтров с субполосами неравной ширины и ДПФ [10,12,13]. WDFT позволяет получить неравномерное  $z$ -преобразование конечной последовательности, используя фазовый фильтр  $A(z)$  и для последовательности  $x[n]$  из  $N$  точек определяется следующей формулой:

$$\hat{X}(z_k) = X(\hat{z}_k) = \sum_{n=0}^{N-1} x[n] \hat{z}_k^{-n}, \quad k=0, \dots, N-1, \quad (2)$$

где  $\hat{z}_k$  – изображения равноотстоящих точек на единичной окружности в  $z$ -плоскости, получаемые из преобразования

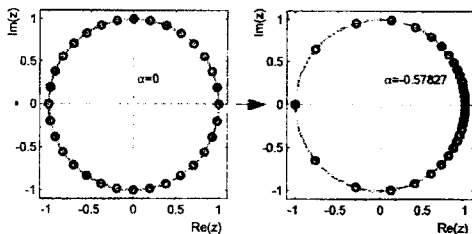
$$z_k^{-1} = e^{-j \frac{2\pi k}{N}} \rightarrow \hat{z}_k^{-1} = A(z_k) \quad k=0, \dots, N-1 \quad (3)$$

с фазовым звеном  $A(z)$  произвольного порядка.

Простейший вариант WDFT основан на фазовом звене первого порядка с действительным коэффициентом  $a$ , отвечающим за неравномерность шага по частоте:

$$\hat{\omega} = \omega + 2 \arctan \left( \frac{a \sin \omega}{1 - a \cos \omega} \right) \quad \text{для} \quad \begin{cases} z = e^{j\omega} \\ \hat{z} = e^{j\hat{\omega}} \end{cases} \quad (4)$$

то есть осуществляется билинейное преобразование  $z$ -плоскости в новую искривленную  $\hat{z}$ -плоскость (рисунок 2).



**Рисунок 2 – Расположение частотных отсчетов ДПФ (слева) и WDFТ (справа)**

Значение коэффициента фазового фильтра для заданной частоты дискретизации определяется по следующему выражению:

$$a_{Bark} = 0.1957 - 1.048 \cdot \left[ \frac{2}{\pi} \arctan \left( 0.07212 \frac{f_s}{1000} \right) \right]^2. \quad (5)$$

На базе данного преобразования строится психоакустическая модель для оценки субполосных порогов маскирования  $P_{TT,b}$  и перцептуальной значимости обрабатываемых субполос  $SPE_b$ . Далее предлагается алгоритм управления структурой банка субполосных кодовых книг для возможности его адаптации к частотно-временному плану фрейма обрабатываемого сигнала [5,10]:

$$SPE_{\max} = \max(SPE),$$

$$DPE_b = \begin{cases} \frac{SPE_{\max}}{\ell_b}, & \text{если } \ell_b \geq SPE_{\max} \\ \text{иначе } \frac{\ell_b}{SPE_{\max}} \end{cases}, \quad b \in Bands,$$

$$d_b = \min(\ell_b, \text{nint}(SPE_b \cdot DPE_b)),$$

$$\text{depth}(i) = \begin{cases} d_i, & \text{если } i \in Bands \\ \text{иначе } 0 \end{cases}, \quad i = 1, 2, \dots, C(B),$$

где  $\ell_b$  – максимально возможная глубина поиска в кодовой книге для субполосы  $b$ ;

$DPE_b$  – вклад каждого уровня книги в суммарное значение субполосной перцептуальной энтропии;

$Bands \subset B_{SPE_b > 0}$  – подмножество обрабатываемых субполос;

$d_b$  – номер уровня детализации в соответствующей субполосной кодовой книге;

$\text{rint}$  – операция округления;

$\text{depth}$  – вектор, содержащий номера требуемых уровней детализации для каждой субполосной кодовой книги;

$SPE_b$  определяется следующим образом:

$$SPE_b = \log_2 \left( \frac{E_b}{\sqrt{\frac{6 \cdot P_{TT,b}}{\text{bandwidth}_b}}} + 1 \right), \text{ [бит/полосу]}, \quad (7)$$

где  $E_b$  – энергия сигнала в субполосе  $b$ ;

$\text{bandwidth}_b$  – ширина субполосы  $b$  в спектральных отсчетах.

С учетом алгоритма (6), синтезированный субполосный сигнал будет являться функцией от индекса вектора возбуждения на заданном уровне детализации  $\hat{s}_{w_b}(i_b^{(\lambda)}, \text{depth}(b))$  и уточненная математическая модель кодирования широкополосного речевого сигнала может быть выражена следующим образом:

$$\{B, SPE, P_{TT}\} = \{D, MT[\cdot], PE[\cdot]\},$$

$$\text{Bands} \subset \text{descend} \left[ \frac{B}{SPE > 0} \right],$$

$$\tilde{s}_w(n) = LP[s(n)], \quad n = 1, 2, \dots, N, \quad 1 \leq \lambda \leq \lambda_c,$$

$$e_{w_0}^{(\lambda)}(l) = \tilde{s}_w(k), \quad l = 1, 2, \dots, L, \quad k = (\lambda - 1) \cdot L + 1, (\lambda - 1) \cdot L + 2, \dots, \lambda \cdot L,$$

$$b = \overline{1, C(\text{Bands})}, \quad (8)$$

$$\{g_{b,opt}^{(\lambda)}, i_{b,opt}^{(\lambda)}\} = \arg \min_{g_b^{(\lambda)}, i_b^{(\lambda)}} e_{w_b}^{(\lambda)}(l) = \frac{e_{w_{(b-1)}}^{(\lambda)}(l) - G_b(g_b^{(\lambda)}) \cdot \hat{s}_{w_b}(i_b^{(\lambda)}, \text{depth}(b))}{P_{TT,b}},$$

$$e_w^{(\lambda)}(l) = e_{w_0}^{(\lambda)}(l) - \sum_{b=1}^{C(B)} G_b(g_b^{(\lambda)}) \cdot \hat{s}_{w_b}(i_{b,opt}^{(\lambda)}),$$

$$e_w(n) = \{e_w^{(1)}(l), e_w^{(2)}(l), \dots, e_w^{(\lambda_c)}(l)\},$$

где  $D$  – прямое частотно-временное преобразование WDFG;

$MT[\cdot]$  – оператор расчета порогов маскирования, включающий оценку субполосной перцептуальной энтропии;

$PE[\cdot]$  – оператор, представляющий алгоритм (6) определения требуемого уровня детализации в субполосных кодовых книгах;

$\text{descend}[\cdot]$  – операция сортировки субполос, участвующих в процессе кодирования, в порядке убывания  $SPE$ .

Схема кодера и декодера с динамически трансформируемым банком кодовых книг представлена на рисунках 3 и 4.

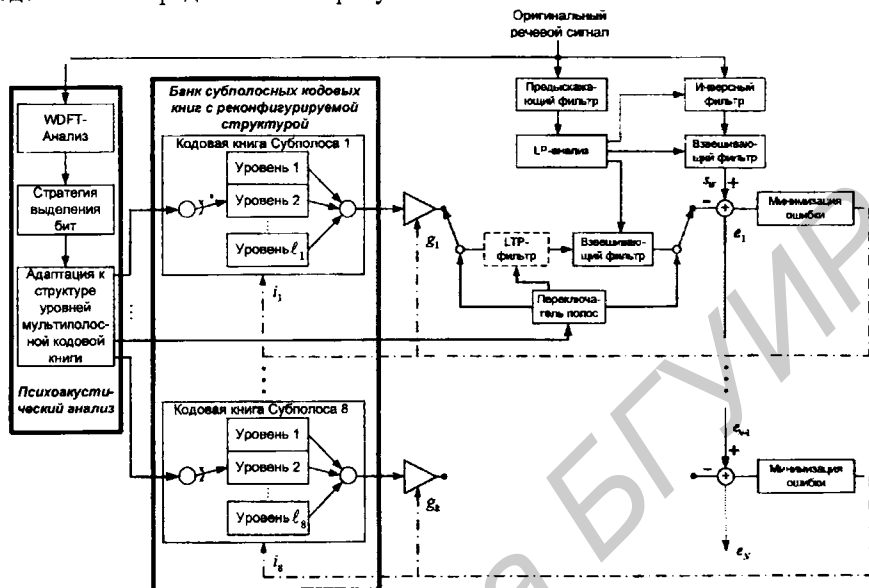


Рисунок 3 – Схема перцентуального широкополосного кодера речевых сигналов с динамически реконфигурируемой структурой банка кодовых книг на базе WDFT-психоакустической модели [12]

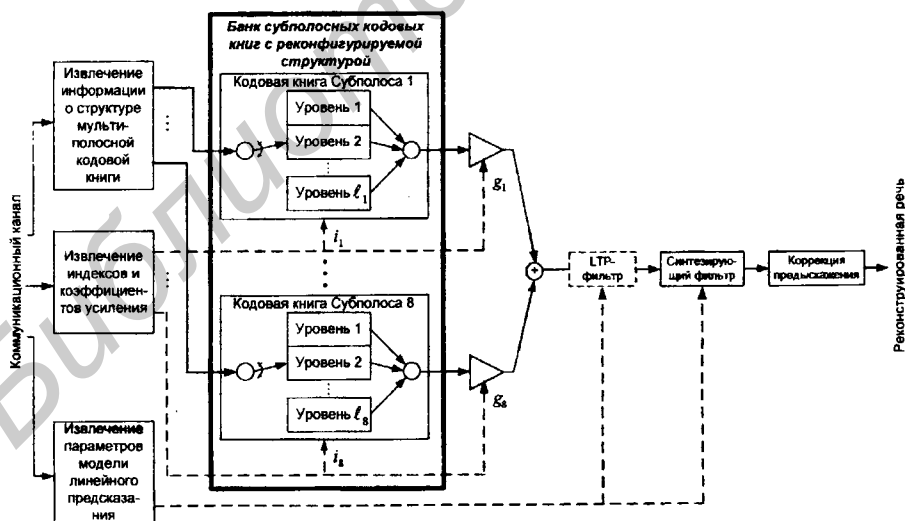


Рисунок 4 – Схема декодера [12]

Далее в главе рассматривается вопрос повышения качества реконструированной речи в условиях окружающих акустических шумов. С учетом специфики частотно-временного преобразования и адаптированной для него психоакустической модели, применяемой в компрессоре, принимается решение использовать тот же математический аппарат и в системе редактирования шума. Данный подход нацелен на обеспечение работы обеих систем в единой перцептуальной шкале, что не требует различных преобразований и согласований, а также минимизацию вычислений [7,13,14]. При этом результаты работы психоакустической модели в препроцессоре редактирования шума используются для реконфигурации банка кодовых книг кодера. Таким образом, математическая модель, отражающая работу комбинированной системы кодирования широкополосных речевых сигналов с препроцессором повышения качества речи, может быть записана следующим образом:

$$\begin{aligned}
 & [\tilde{s}(n), B, SPE, P_{TT}] = NRS[y(n)], \\
 & Bands \subset \text{descend} \left[ \begin{array}{c} B \\ SPE > 0 \end{array} \right], \\
 & \tilde{s}_w(n) = LP[\tilde{s}(n)], \quad n=1,2,\dots,N, \quad 1 \leq \lambda \leq \lambda_c, \\
 & e_w^{(\lambda)}(l) = \tilde{s}_w(k), \quad l=1,2,\dots,L, \quad k = (\lambda-1) \cdot L + 1, (\lambda-1) \cdot L + 2, \dots, \lambda \cdot L, \\
 & \quad \quad \quad b = \overline{1, C(Bands)}, \\
 & \{g_{b,opt}^{(\lambda)}, i_{b,opt}^{(\lambda)}\} = \arg \min_{g_b^{(\lambda)}, i_b^{(\lambda)}} e_w^{(\lambda)}(l) = \frac{e_{w(b-1)}^{(\lambda)}(l) - G_b(g_b^{(\lambda)}) \cdot \hat{s}_{w_b}(i_b^{(\lambda)}, \text{depth}(b))}{P_{TT,b}}, \\
 & e_w^{(\lambda)}(l) = e_w^{(1)}(l) - \sum_{b=1}^{C(B)} G_b(g_{b,opt}^{(\lambda)}) \cdot \hat{s}_{w_b}(i_{b,opt}^{(\lambda)}), \\
 & e_w(n) = \{e_w^{(1)}(l), e_w^{(2)}(l), \dots, e_w^{(\lambda_c)}(l)\},
 \end{aligned} \tag{9}$$

где  $y(n)$  – фрейм зашумленного входного сигнала;

$\tilde{s}(n)$  – фрейм очищенного сигнала;

$NRS[\cdot] = \{D, LSA, MT, H^{JND}, PE, D^{-1}\}$  – обобщенный оператор системы

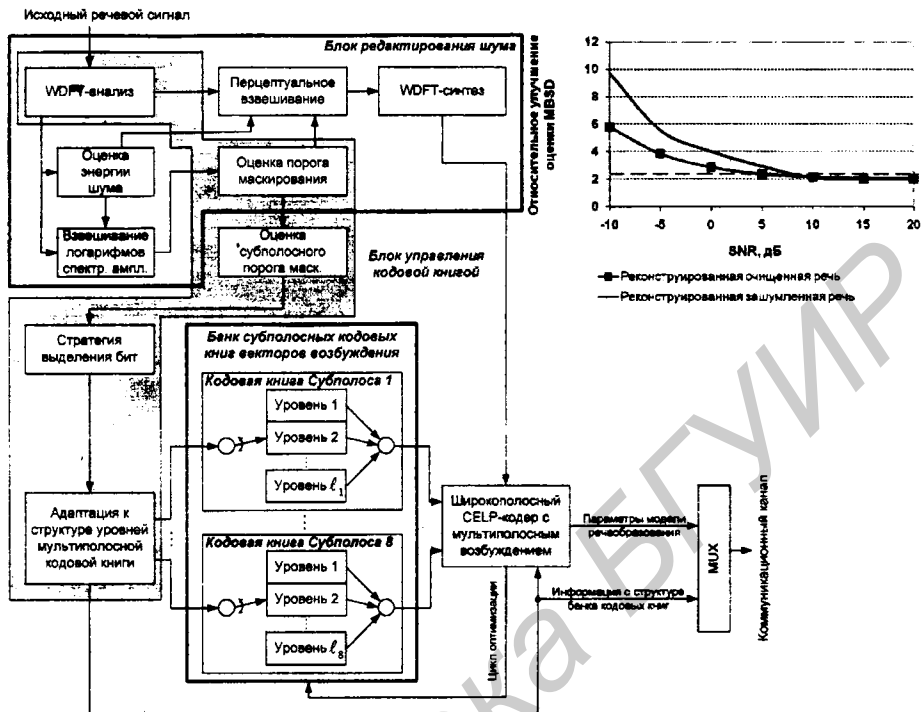
редактирования шума;

$LSA[\cdot]$  – правило оценки логарифмов спектральных амплитуд;

$H^{JND}[\cdot]$  – функция перцептуального взвешивания;

$D^{-1}$  – обратное частотно-временное преобразование WDFТ для синтеза очищенного речевого сигнала.

На основании модели (9) представляется структурная схема комбинированной системы кодирования широкополосного речевого сигнала (рисунок 5) [7,13,14].

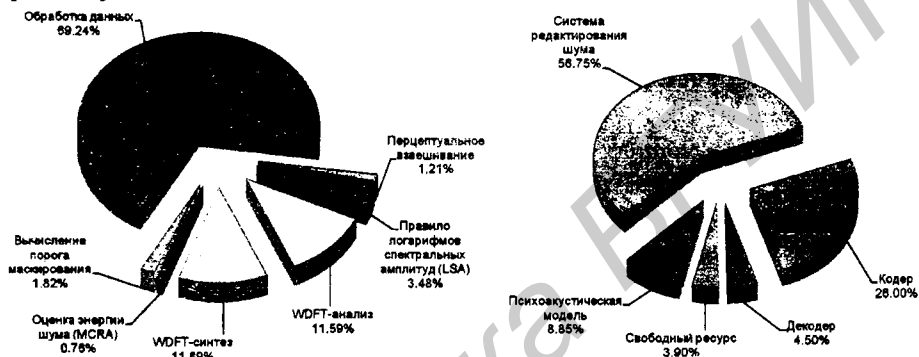


**Рисунок 5 – Комбинированная система кодирования широкополосных речевых сигналов и область эффективной работы препроцессора повышения качества речи [7]**

**Глава 3** посвящена синтезу банка кодовых книг векторов возбуждения, формированию обучающих множеств, алгоритму обучения кодовых книг [3], определению перцептуальной значимости уровней детализации в субполосных кодовых книгах, на основании которой осуществляется оптимизация избыточной структуры банка кодовых книг. Предлагается древовидная организация уровней в субполосных кодовых книгах, которая обеспечивает семикратное увеличение скорости поиска оптимального вектора возбуждения по сравнению с кодовой книгой с независимой организацией уровней и позволяет масштабировать поток данных в зависимости от перцептуальных свойств кодируемого сигнала и ограничений трафика. Представлен метод квантования информации о структуре реконфигурируемого банка кодовых книг на базе девятибитного векторного квантователя [6]. По результатам оптимизации осуществляется перераспределение высвобожденного бюджета бит между наиболее чувствительными к квантованию параметрами модели речеобразования.

В **4-й главе** представлено алгоритмическое обеспечение комбинированной системы кодирования широкополосного речевого сигнала с

масштабируемой скоростью потока данных (до 24.2 кбит/с), рассматриваются вопросы аппаратной реализации данной системы на базе VLIW ЦПОС TMS320C6713 [7,13], отображение алгоритмов на архитектуру аппаратной платформы и организация вычислительного процесса (диспетчеризация задач) с учетом работы в реальном масштабе времени с оценкой вычислительной сложности алгоритмов комбинированной системы кодирования широкополосных речевых сигналов (рисунок 6). Приводится описание отладочного модуля и разработанной платформы с реконфигурируемой логикой на базе FPGA Xilinx Spartan-3, выполняющей функции периферийного процессора.



а) алгоритмы системы редактирования шума

б) комбинированная система кодирования широкополосных речевых сигналов

**Рисунок 6 – Распределение вычислительной сложности**

Предлагаемое архитектурное решение позволяет получить следующие особенности аппаратной платформы: мощное вычислительное ядро с плавающей запятой для обработки данных с производительностью до 1350/1800 MFLOPS/MIPS; возможность распараллеливания задач обработки в двухпроцессорной системе; реконфигурируемость платформы путем изменения конфигурационной последовательности FPGA и программ ЦПОС; возможность встраивания платформы в разнообразное оборудование за счет гибкой настройки внешнего интерфейса с помощью периферийного процессора; использование стандартных средств конфигурирования и программирования. Основные электрические параметры платформы: напряжение питания модуля +5В постоянного тока; ток потребления – не более 300 мА; потребляемая мощность зависит от загрузки ЦПОС и ПЛИС (от реализуемых алгоритмов) и составляет не более 1.5 Вт.

**Глава 5** является экспериментальной частью диссертационной работы. В ней осуществлена оценка качества и эффективности системы редактирования шума (рисунок 7) [7,13], представлен анализ существующих методик субъективной и объективной оценки качества звучания реконструированного речевого сигнала, проведены соответствующие исследования (таблица 1 и 2).

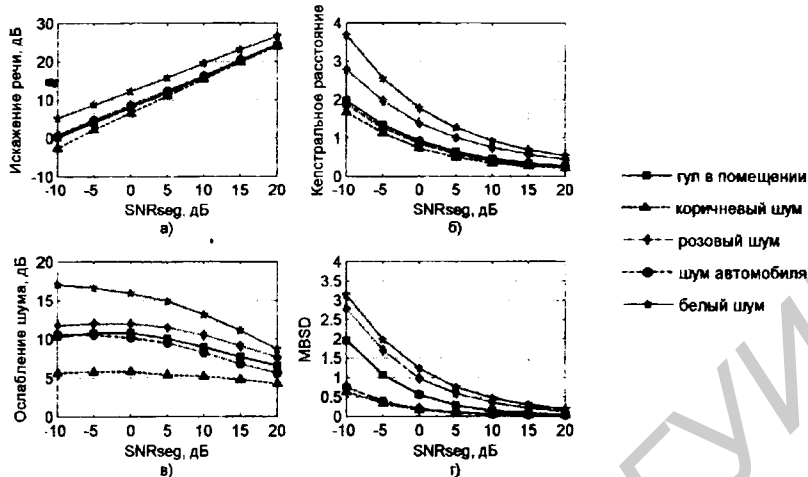


Рисунок 7 – Объективные оценки качества системы редактирования шума

Таблица 1 – Объективные и субъективные оценки качества восстановленного в декодере речевого сигнала

Отношение сигнал/шум (SNR), дБ	Отношение шум/порог маскирования (NMR), дБ	Оценка искажений спектра барков (BSD)	Модифицированная оценка искажений спектра барков (MBSD)
11.59	-6.76	0.05	0.02

Таблица 2 – Субъективные оценки качества звучания реконструированной речи

Слоговая разборчивость речи, %	Субъективное качество речи, баллы	Особенности восстановленной в декодере речи	Узнаваемость голоса диктора, %
98	4.2	Хорошая естественность и узнаваемость	95

Значения полученных объективных оценок являются достаточно высокими для компрессоров подобного класса, восстановленный речевой сигнал отличается довольно высокой степенью разборчивости и узнаваемости диктора [14]. Анализ качества препроцессора редактирования шума показывает, что предлагаемая система достаточно хорошо справляется с шумами различной природы при различных уровнях SNR. Препроцессор обеспечивает улучшение перцептуального качества реконструированной очищенной речи в условиях окружающих шумов с диапазоном уровней SNR от -10 дБ до +6 дБ (рисунок 5). Для SNR более 6 дБ система редактирования шума не влияет на качество реконструированной речи и, следовательно, может быть постоянно включена. Сравнительный анализ качества реконструированной речи с стандартизованными широкополосными кодеками речевых и



аудиосигналов [5,12] (MPEG1 Layer III и ITU-T G.722), показывает более высокое перцептуальное качество предлагаемой системы компрессии при сравнимой или меньшей скорости потока данных.

В приложениях представлены примеры отображения С-кода на архитектуру ЦПОС TMS320C67x, методики определения объективных оценок качества реконструированного речевого сигнала, вычислительные аспекты WDFT, а также акты внедрения результатов диссертационной работы.

## ЗАКЛЮЧЕНИЕ

### Основные научные результаты диссертации

В ходе представленной диссертационной работы были получены следующие научные результаты.

1. На основании анализа подходов построения перцептуально мотивированных методов и устройств компрессии широкополосных речевых сигналов показано, что наилучшего перцептуального качества реконструированного речевого сигнала и более глубокой степени компрессии позволяют добиться гибридные структуры кодеров. Предложена новая структура и модель CELP-кодера с мультиполосным возбуждением и обработкой субполос в порядке убывания их перцептуальной значимости [5,9 10]. В качестве вычислительной платформы предлагается использовать однокристалльный VLIW-процессор с поддержкой SIMD-инструкций, позволяющий осуществлять параллельно-поточную обработку данных в реальном масштабе времени [6,13].

2. Разработана структура компрессора широкополосных речевых и аудиосигналов на базе мультиполосного CELP-алгоритма с реконфигурируемой структурой банка кодовых книг [5,10], обеспечивающая масштабируемость потока данных (максимальная скорость 24.2 кбит/с) и более высокое перцептуальное качество по сравнению с устройствами подобного класса (оценка по процедуре MOS составляет 4.2 балла).

3. В структуру широкополосного кодера речевых сигналов интегрирована система редактирования шума, при этом препроцессор повышения качества речи и кодер работают на базе единой психоакустической модели и в единой перцептуальной шкале. Для обеспечения статистически равнозначного анализа энергии сигнала и оценки порога маскирования в критических частотных полосах используется психоакустическая модель на базе WDFT, которая, в отличие от существующих моделей, обеспечивает более высокое перцептуальное качество [1,8,10,12]. Интегрированный препроцессор повышения качества речи обеспечивает высокую степень подавления акустических шумов различной природы (до 26 дБ) и расширяет диапазон работы системы в шумах вплоть до шумов с SNR=-10 дБ без осязаемой на слух деградации качества реконструированного речевого сигнала [13,14].

4. Разработан метод декомпозиции частотной шкалы на субполосы (группирования критических частотных полос), с учетом формантной структуры речи и особенностей слухового восприятия человека, который обеспечивает равную концентрацию энергии в субполосах (их равнозначность) [2,4,5,9,10,11].

5. Проведен анализ влияния каждой субполосы на качество кодирования исходного сигнала. Получена новая оптимизированная структура банка кодовых книг, обеспечивающая лучшее качество кодирования по сравнению с избыточной структурой. При этом достигнуто снижение скорости потока данных с 12800 бит/с до 11200 бит/с и обеспечено «прозрачное» кодирование структуры банка при помощи 9-битного векторного квантователя [6].

6. Предложена древовидная организация субполосных кодовых книг. Данная структура позволила увеличить скорость поиска оптимального вектора возбуждения почти в семь раз, по сравнению с кодовой книгой с независимой организацией уровней, и обеспечить плавное снижение перцептуального качества реконструированной речи при уменьшении скорости потока данных [6].

7. Синтезирован алгоритм управления структурой банка кодовых книг на базе перцептуальной энтропии каждой субполосы с целью моделирования частотно-временного плана фрейма кодируемого сигнала [5].

8. Осуществлена аппаратная реализация комбинированной системы компрессии широкополосных речевых сигналов с препроцессором редактирования шума на базе платформы VLIW ЦПОС Texas Instruments TMS320C6713B с реконфигурируемой логикой на основе FPGA Xilinx Spartan-3 [7,13,14]. Данная платформа является универсальной и подходит для решения широкого круга задач, касающихся разработки мобильных мультимедийных устройств реального времени.

9. Создано ПО экспериментальной исследовательской системы, позволяющей проводить моделирование и исследования в рамках предложенной модели и методов кодирования широкополосной речевой информации с оценкой перцептуального качества [2,3,8].

#### **Рекомендации по практическому использованию результатов**

Разработанная система компрессии может использоваться в составе персонального компьютера как дополнительное мультимедийное устройство для сжатия речевых сообщений с целью их хранения в памяти и передачи по компьютерным сетям или как вспомогательное устройство для связи посредством персонального компьютера (интернет-телефония). Кроме того, комбинированная система компрессии широкополосных речевых сигналов может применяться в мобильных средствах связи, работающих в акустических шумах.

# СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ ПО ТЕМЕ ДИССЕРТАЦИИ

## Глава в книге

1. Livshitz, M. «Szerokopasmowy koder mowy typu CELP o wielopasmowym wzbudzeniu» / M. Livshitz // Cyfrowe banki filtrów: analiza, synteza i implementacja dla systemów multimedialnych / Aleksander Piotrowski, Marek Parfieniuk. – Politechnika Białostocka, Białystok, 2006. – Section 10.4. – P. 310–325.

## Статьи в научных журналах

2. Лившиц, М.З. Многополосный CELP-вокодер / М.З. Лившиц // Известия Белорусской инженерной академии. – 2003. – №1(15)/1. – С. 259–262.

3. Лившиц, М.З. Векторное квантование в широкополосных кодерах речи / М.З. Лившиц // Известия Белорусской инженерной академии. – 2003. – №1(15)/2. – С. 281–284.

4. Лившиц, М.З. Линейное предсказание при обработке сигналов в неравномерных частотных шкалах / М.З. Лившиц // Известия Белорусской инженерной академии. – 2005. – №1(19)/1. – С. 234–238.

5. Лившиц, М.З. Широкополосный CELP-кодер с мультиполосным возбуждением и многоуровневым векторным квантованием по кодовой книге с реконфигурируемой архитектурой / М.З. Лившиц, М. Парфенюк, А.А. Петровский // Цифровая обработка сигналов, Москва. – 2005. – №2. – С. 20–35.

6. Livshitz, M. Synthesis of Codebooks with Perceptually Monitored Structure for Multiband CELP-Coders» / M. Livshitz, A. Petrovsky // Biometrics, Computer Security Systems and Artificial Intelligence Applications; edited by K. Saeed, J. Pejas, R. Mosdorf. – Springer Science + Business Media, New York, USA, 2006. – P. 57–67.

7. Livshitz, M. Real-Time Implementation of Combined Noise Reduction and Wideband Speech Coding Systems based on the DSP TMS320C6713 / M. Livshitz, A. Borowicz, A. Petrovsky // New Trends in Audio and Video; edited by Andrzej Dobrucki, Alexander Petrovsky, Wladislaw Skarbek. – Politechnika Białostocka, Białystok, 2006. – vol. I. – P. 473–478.

## Материалы научно-технических конференций

8. Лившиц, М.З. Широкополосный компрессор речевых сигналов с психоакустической мотивацией на базе алгоритма CELP // М.З. Лившиц, А.А. Петровский // Материалы 8-й Международной научно-технической конференции «Современные средства связи» / Известия Белорусской инженерной академии. – 2003. – №1(15)/3. – С. 117–119.

9. Лившиц, М.З. Многоуровневое векторное квантование речевого сигнала по мультиполосной кодовой книге в широкополосном CELP-кодере с психоакустической мотивацией / М.З. Лившиц, А.А. Петровский // VI

Международная конференция «Цифровая обработка сигналов и ее применение» / Труды РНТОРЭС им. А.С.Попова, выпуск VI-1, Москва. – 2004. – С. 119–123.

10. Лившиц, М.З. Многоуровневое векторное квантование с переменной глубиной поиска в перцептуальных кодерах речи с психоакустической моделью на основе сжатого ДЦФ / М.З. Лившиц, М. Парфенюк, А. Петровский // VII Международная конференция «Цифровая обработка сигналов и ее применение» / Труды РНТОРЭС им. А.С.Попова, выпуск VII-1, Москва. – 2005. – С. 187–191.

11. Livshitz, M.Z. Perceptual Weighting and LP-Analysis Using Time-Varying Windows in CELP-Encoders / M.Z. Livshitz // Proceedings of the 8th International Conference on Pattern Recognition and Information Processing (PRIP'2005), May 18-20, Minsk, Belarus. – Minsk, Belarus, 2005. – P. 45–48.

12. Livshitz, M. Perceptually Constrained Variable Bitrate Wideband Speech Coder / M. Livshitz, A. Petrovsky // Proceedings of EUROCON, Serbia & Montenegro, Belgrade, November 22-24, 2005. – Belgrade, 2005. – P. 1296–1299.

13. Livshitz, M. An Overcomplete WDFT-based Perceptually Constrained Variable Bit Rate Wideband Speech Coder with Embedded Noise Reduction System / M. Livshitz, A. Petrovsky // Proceedings of XI International Conference «SPEECH AND COMPUTER», St. Petersburg, Russia, June 25-29, 2006. – St. Petersburg, Russia, 2006. – P. 343–348.

14. Perceptually-Based Front-End Real-Time Processor for Very Low Bit-Rate Speech Coding / M. Livshitz, A. Borowicz, A. Petrovsky // Proceedings of the 9th International Conference on Pattern Recognition and Information Processing (PRIP'2007), May 22-25, Minsk, Belarus. – Minsk, Belarus, 2007. – vol. I. – P. 45–48.

*M. Livshitz*

# РЭЗІЮМЭ

Ліўшыц Міхаіл Зенадз'евіч

## Камбінаваная сістэма перцептуальнага шырокапалоснага кодэра мовы з прапрацэсарам рэдагавання шуму

**Ключавыя словы:** банк субпалосных кодавых кніг з дынамічнай трансфармацыяй структуры, ДПФ (Дыскрэтнае Пераўтварэнне Фур'е) з нероўнамерным раскладаннем каэфіцыентаў па частотнай восі, кампрэсія моўнага сігналу, перцептуальная энтрапія, псіхаакустыка, сістэма рэдагавання шуму.

Мэтай работы з'яўляецца распрацоўка метадаў і прыстасаванняў кампрэсіі шырокапалосных моўных і аудыёсігналаў, а таксама зніжэнне шумоў навакольнага асяроддзя ў мультымедыя-сістэмах рэальнага часу, якія грунтуюцца на псіхаакустычных асаблівасцях успрымання чалавекам акустычнай інфармацыі.

Аб'ектам даследавання з'яўляецца прыстасаванне шырокапалоснага CELP-кодэра з многापалосным узбуджэннем і перцептуальнай апрацоўкай сігналу ў субпалосах.

Прадмет даследавання – метады кампрэсіі шырокапалосных моўных і аудыёсігналаў на падставе CELP-мадэлі з мультыпалосным узбуджэннем, якая выкарыстоўвае псіхаакустычную мадэль для больш дэталёвага аналізу і кадыравання субпалос.

Навуковая навізна атрыманых аўтарам вынікаў ў наступным: прапанавана новая структура і мадэль CELP-кодэра з мультыпалосным узбуджэннем і апрацоўкай субпалос у парадку змяншэння іх перцептуальнай значнасці; сінтэзаваны банк субпалосных кодавых кніг з рэканфігуруемай структурай, якая адлюстроўвае частотна-часовы план фрэйма кадыруемага сігналу; распрацаваны алгарытм дынамічнай трансфармацыі банка на базе субпалоснай перцептуальнай энтрапіі; прапанавана выкарыстоўваць псіхаакустычную мадэль на базе ДПФ з нероўнамерным раскладаннем каэфіцыентаў па частотнай восі, якая забяспечвае лепшую перцептуальную якасць у параўнанні з вядомымі мадэлямі; у склад кампрэсара інтэгравана сістэма рэдагавання шуму на базе таго ж частотна-часовага пераўтварэння; створана праграма-апаратнае забеспячэнне прыстасавання кампрэсіі, прызначанага для выкарыстання ў складзе персанальнага камп'ютара; распрацавана апаратная платформа на базе VLIW DSP Texas Instruments TMS320C713B з рэканфігуруемай логікай на аснове FPGA Xilinx Spartan-3. Асноўныя вынікі дысертацыйнай працы ўкаранёныя і выкарыстоўваюцца ў Дзяржаўным цэнтры бяспекі інфармацыі пры Прэзідэнце РБ, Упраўленні ўрадавай сувязі КДБ РБ.

# РЕЗЮМЕ

Лившиц Михаил Зенадьевич

## Комбинированная система перцептуального широкополосного кодера речи с препроцессором редактирования шума

**Ключевые слова:** банк субполосных кодовых книг с динамической трансформацией структуры, ДПФ с неравномерным разложением коэффициентов по частотной оси, компрессия речевого сигнала, перцептуальная энтропия, психоакустика, система редактирования шума.

Целью работы является разработка методов и устройств компрессии широкополосных речевых и аудиосигналов, а также снижение шумов окружающей среды в мультимедиа системах реального времени, базирующихся на психоакустических особенностях восприятия человеком акустической информации.

Объектом исследования является устройство широкополосного CELP-кодера с многополосным возбуждением и перцептуальной обработкой сигнала в субполосах.

Предмет исследования – методы компрессии широкополосных речевых и аудиосигналов на основе CELP-модели с мультиполосным возбуждением, использующей психоакустическую модель для более детального анализа и кодирования субполос.

Научная новизна полученных автором результатов заключается в следующем: предложена новая структура и модель CELP-кодера с мультиполосным возбуждением и обработкой субполос в порядке убывания их перцептуальной значимости; синтезирован банк субполосных кодовых книг с реконфигурируемой структурой, отражающей частотно-временной план фрейма кодируемого сигнала; разработан алгоритм динамической трансформации банка на базе субполосной перцептуальной энтропии; предложено использовать психоакустическую модель на базе ДПФ с неравномерным разложением коэффициентов по частотной оси, обеспечивающую лучшее перцептуальное качество по сравнению с известными моделями; в состав компрессора интегрирована система редактирования шума на базе того же частотно-временного преобразования; создано программно-аппаратное обеспечение устройства компрессии, предназначенного для использования в составе персонального компьютера; разработана аппаратная платформа на базе VLIW ЦПОС Texas Instruments TMS320C6713B с реконфигурируемой логикой на основе FPGA Xilinx Spartan-3.

Основные результаты диссертационной работы внедрены и используются в Государственном центре безопасности информации при Президенте РФ, Управлении правительственной связи КГБ РФ.

## SUMMARY

of the thesis by Michael Livshitz

### **Combined system of perceptual wideband speech coder with noise reduction preprocessor**

**Keywords:** bank of subband codebooks with dynamic reconfigurable structure, Warped Discrete-Fourier Transform, wideband speech compression, perceptual entropy, psychoacoustics, noise reduction system.

The purpose of the work is a development of methods and devices for wideband speech and audio compression and environmental noise suppression in real-time multimedia systems based on human perception features.

The object of the research is a wideband CELP coder with multiband excitation and perceptual signal processing in subbands.

The subjects of the research are wideband speech and audio compression methods based on CELP model with multiband excitation using psychoacoustic model for in-depth analysis and subband coding.

The scientific novelty of the results received by the author consists of the following issues: new CELP coder structure and model with multiband excitation and processing of subbands in descending order of their perceptual relevance are proposed; the bank of subband codebooks with reconfigurable structure mapped time-frequency plane of coded frame of signal is realized; the dynamic transformation algorithm of bank of codebooks based on subband perceptual entropy are synthesized; it is proposed to use a new psychoacoustic model based on Warped Discrete-Fourier Transform that provides better perceptual quality in comparison with known models; a noise reduction system based on the same time-to-frequency transformation is integrated into the structure of speech compressor; hardware-software implementation of the combined system is designed for use in personal computer as an additional multimedia device; hardware platform based on Texas Instruments DSP TMS320C6713B with VLIW architecture and reconfigurable logic on FPGA Xilinx Spartan-3 (peripheral processor) are designed and developed.

Results of the dissertational work are implemented into practice and used in the State Center of Information Security at the President of the Republic of Belarus and Department of Government Telecommunication of State Security Committee of the Republic of Belarus.

ЛИВШИЦ  
Михаил Зенадьевич

**КОМБИНИРОВАННАЯ СИСТЕМА ПЕРЦЕПТУАЛЬНОГО  
ШИРОКОПОЛОСНОГО КОДЕРА РЕЧИ С ПРЕПРОЦЕССОРОМ  
РЕДАКТИРОВАНИЯ ШУМА**

**АВТОРЕФЕРАТ**  
диссертации на соискание ученой степени  
кандидата технических наук

по специальности 05.13.05  
«Элементы и устройства вычислительной техники и систем управления»

---

Подписано в печать	19.10.2007.	Формат 60x84 <sup>1</sup> / <sub>16</sub> .	Бумага офсетная.
Гарнитура «Таймс».	Печать ризографическая.		Усл. печ. л. 1,63.
Уч.-изд. л. 1,4.	Тираж 60 экз.		Заказ 617.

---

Издатель и полиграфическое исполнение: Учреждение образования  
«Белорусский государственный университет информатики и радиоэлектроники»  
ЛИ №02330/0056964 от 01.04.2004. ЛП №02330/0131666 от 30.04.2004.  
220013, Минск, П. Бровки, 6.