

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
“БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ”

УДК 004.032.6 + 004.383.3

ПАВЛОВЕЦ
Александр Николаевич

**МЕТОДЫ И АЛГОРИТМЫ
ПАРАМЕТРИЧЕСКОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ
С ПСИХОАКУСТИЧЕСКОЙ ОПТИМИЗАЦИЕЙ
В СИСТЕМАХ МУЛЬТИМЕДИА РЕАЛЬНОГО ВРЕМЕНИ**

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

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

Минск 2010

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 учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Официальные оппоненты: **Лобанов Борис Мефодьевич**, доктор технических наук, профессор, заведующий лабораторией распознавания и синтеза речи государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларуси»

Бранцевич Пётр Юльянович, кандидат технических наук, доцент, доцент кафедры программного обеспечения информационных технологий учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Оппонирующая организация – Белорусский государственный университет

Защита состоится 18 ноября 2010 г. в 16.00 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, ауд. 232-1, тел. 293-89-89, e-mail: dissovet@bsuir.by.

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

КРАТКОЕ ВВЕДЕНИЕ

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.

Традиционно высокое внимание к повышению качества звучания реконструированного речевого сигнала в низкоскоростных системах компрессии речи обуславливает актуальность исследования в рамках совершенствования методов и аппаратно-программных средств кодирования речи в мобильных мультимедийных системах реального времени. Развитие этих работ: в настоящее время идет по пути построения параметрических кодеров речевых сигналов на основе подходящей модели речевого сигнала с последующей психоакустической оптимизацией.

Вышеизложенное подчеркивает актуальность исследований, разработки методов и алгоритмов анализа и кодирования системы параметрической обработки речевых сигналов на базе их декомпозиции на периодическую и аperiodическую компоненты. Важность проблемы, совершенствование методов и алгоритмов параметрической обработки речевых сигналов в реальном масштабе времени и их отображение на архитектуру вычислительной платформы определили цель работы, решаемые задачи и методы исследований.

В настоящей работе предлагается метод построения устройства компрессии речевого сигнала, основанный на его декомпозиции для анализа на периодическую и аperiodическую компоненты с последующей психоакустической оптимизацией при кодировании. Основу аппаратной платформы данной системы составляет цифровой процессор обработки сигналов (ЦПОС) фирмы Texas Instruments TMS320C6713 и периферийный процессор на базе программируемой логической интегральной схемы (ПЛИС) Xilinx Spartan-3. Такое двухпроцессорное ядро обеспечивает возможность гибкой реконфигурации структуры кодера, перераспределения алгоритмов и распараллеливания вычислений, а также встраиваемость в аппаратуру заказчика.

Полученные научные результаты имеют практическое применение в приложениях мультимедийных систем.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Связь работы с крупными научными программами и темами

Исследования, положенные автором в основу диссертационной работы, а также выполняемые им в ходе работы, нашли отражение в следующих НИР:

1. ГНТП «Развитие методов и средств системы комплексной защиты информации» по теме «Разработать процедуры сжатия речевой информации, обеспечивающие коммерческое качество восстановленной речи» (ГБ № 01-1089, № ГР 2002518, научный руководитель НИР – А.А. Петровский).

2. «Разработка низкоскоростного речепреобразующего устройства с препроцессором редактирования шума и компенсатором эхо-сигнала для аппаратуры М-461» (х/д № 05-1160, № ГР 2006589, научный руководитель НИР – А.А. Петровский).

3. «Методики и алгоритмы обработки зашумленных речевых сигналов» (№ ГР 20066753, научный руководитель НИР – Г.В. Давыдов).

Цель и задачи исследования

Целью работы является разработка методов и алгоритмов анализа и синтеза для устройств параметрической обработки речи, использующих декомпозицию речевого сигнала на периодическую и аperiodическую составляющие и психоакустическую оптимизацию.

Поставленная цель работы определяет следующие задачи исследования:

1. Анализ подходов построения параметрических кодеров речи с целью выбора методов анализа речевого сигнала, обеспечивающих высокую эффективность системы компрессии.

2. Разработка методов и алгоритмов параметрической обработки речевых сигналов для достижения высокой степени компрессии и качества синтезируемого сигнала в мультимедийных устройствах реального времени.

3. Разработка метода квантования параметров модели речевого сигнала для представления их минимальным количеством бит.

4. Аппаратно-программная реализация устройства компрессии речевых сигналов на основе параметрической модели, использующей декомпозицию речи на периодическую и аperiodическую компоненты, предназначенного для работы в реальном масштабе времени.

Объектом исследования является устройство параметрического кодера речи на основе декомпозиции речевого сигнала на периодическую и аperiodическую компоненты и психоакустической оптимизации. Предмет исследования – методы компрессии речевых сигналов, основанные на раздельной обработке периодической и аperiodической компонент речевого

сигнала с психоакустической оптимизацией для более качественного квантования.

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. Метод параметрической обработки речевого сигнала на основе его декомпозиции на периодическую и аperiodическую компоненты.

2. Метод и алгоритм одновременного определения частоты основного тона и параметров периодической компоненты.

3. Метод квантования амплитуд гармоник периодической компоненты на основе квантования векторов переменной длины, учитывающий особенности слуховой системы человека.

4. Способ кодирования речевого сигнала на основе гибридной модели, использующей декомпозицию речевого сигнала на периодическую и аperiodическую компоненты.

5. Аппаратно-программная реализация устройства системы кодирования 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 устройств мультимедиа для сжатия речевых сообщений.

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 синтезированной речи.

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, 2004); Российско-белорусских научно-

технических конференциях “Технические средства защиты информации” (Нарочь, Беларусь, 2003); 8-й Международной конференции “Цифровая обработка сигналов и ее применение” (Москва, Россия, 2005); 7-м Международном симпозиуме по Спектральным методам и многоскоростной обработке сигналов (Москва, Россия, 2007); international conference “Computer Information Systems and Industrial Management Applications” (Elk, Poland, 2004); XI Symposium AES «New Trends in Audio and Video» (Bialystok, Poland, 2006); the 9-th international conference “Pattern Recognition and Information Processing” (Minsk, Belarus, 2007); the 13-th international conference “Speech and Computer” (Saint-Petersburg, Russia, 2009); научно-технических конференциях БГУИР (Минск, Беларусь, 2002–2006).

Результаты диссертационной работы внедрены и используются в:

- 1) Государственном центре безопасности информации при Президенте Республики Беларусь;
- 2) Управлении правительственной связи КГБ Республики Беларусь;
- 3) инженерно-техническом ЗАО “Ай-Ти Мобайл”, г. Москва, Российская Федерация;
- 4) научно-исследовательской лаборатории НИЛ 5.3 “Материалы и элементы электронной и сверхпроводниковой техники” БГУИР;
- 5) учебном процессе специальности “Электронные вычислительные средства”.

Опубликованность результатов диссертации

Основные результаты диссертации опубликованы в 14 научных работах, в том числе 6 статей – в научно-технических журналах, рекомендованных ВАК для опубликования результатов диссертационных исследований, 4 статьи в научных сборниках, 3 статьи в сборниках материалов научно-технических конференций и симпозиумов, 1 тезис докладов на научных конференциях. Суммарный объем публикаций составляет около 5,8 авторских листов. Результаты работы включены в 4 отчёта о НИР.

Структура и объём диссертации

Диссертационная работа состоит из введения, пяти глав, заключения и приложений. **Во введении** показана актуальность темы диссертационной работы, формулируются цель и задачи исследования. В **главе 1** проведён анализ существующих методов параметрического кодирования речи и выполняется предварительный выбор аппаратной платформы для реализации устройства компрессии. В **главе 2** осуществляется разработка метода параметрической обработки речевого сигнала на основе его декомпозиции на периодическую и аperiodическую компоненты. **Глава 3** посвящена разработке

метода квантования параметров периодической компоненты речевого сигнала. В **4-й главе** дано описание алгоритмического обеспечения системы кодирования речевого сигнала, обсуждаются особенности аппаратной реализации данной системы в реальном масштабе времени на базе ЦПОС TMS320C6713. Представлены характеристики отладочного модуля и разработанной платформы. В **5-й главе** приведены результаты субъективных и объективных оценок восстановленной в декодере речи, а также исследование метода декомпозиции речевого сигнала. В **приложениях** диссертации даны методики определения объективных оценок качества звучания реконструированной речи, описание алгоритма выбора контура частоты основного тона методом динамического программирования и акты внедрения результатов диссертационной работы.

Работа представлена на 194 страницах машинописного текста, в том числе основная часть – на 142 страницах (84 страницы текста, 91 рисунок на 53 страницах, 20 таблиц на 5 страницах), 3 приложения на 22 страницах, библиографический список из 187 наименований на 16 страницах, включая 14 публикаций автора на 2 страницах.

СОДЕРЖАНИЕ РАБОТЫ

В **главе 1** дан краткий обзор развития методов низкоскоростного кодирования речи, рассмотрены наиболее эффективные для этой цели параметрические и гибридные модели.

Главным достоинством применяемых моделей и методов является высокая разборчивость реконструированной речи, достигаемая при низкой скорости.

Основным недостатком рассмотренных систем обработки речи на основе параметрических и гибридных моделей является использование недостаточно точных методов разделения вокализованной и шумовой компонент, что приводит к снижению качества восстановленной речи.

Другая причина погрешностей в синтезированной речи – неточное определение амплитуд гармоник основного тона при значительном изменении фундаментальной частоты на интервале анализа. При определении амплитуд гармоник спектр полагается стационарным, а при их синтезе осуществляется линейная интерполяция как частоты основного тона (a , следовательно, и частот гармоник), так и амплитуд гармоник между соседними окнами. Наиболее сильно это сказывается на качестве богатой интонациями речи.

На основе анализа существующих подходов параметрической и гибридной обработки речевых сигналов делается вывод, что для повышения качества при высокой степени компрессии требуется повысить степень соответствия используемой модели природе обрабатываемого речевого отрезка.

Для этого необходимо дополнить её методами обработки переходных сегментов и повысить точность выделения периодической и шумовой компонент.

Существующие в настоящее время параметрические модели анализа-синтеза речевого сигнала рассматривают речь как квазистационарный процесс, в то время как частота основного тона может изменяться даже в течение одного сегмента, отчего особенно страдает обработка богатой эмоциями речи. Следовательно, анализ речи должен быть согласован с изменением частоты основного тона.

Ни в одном из рассмотренных примеров реализации систем компрессии речи не используются в полной мере закономерности восприятия речи человеком. Для достижения высокой степени компрессии следует использовать психоакустическую оптимизацию.

В главе выполняется анализ аппаратных платформ для реализации устройства компрессии. На основании проведённого анализа в качестве основы для аппаратной реализации предлагается использовать ЦПОС с VLIW-архитектурой и с поддержкой SIMD-инструкций. Ожидается, что такой подход позволит не только достичь хороших экономических показателей, но и построить вычислительную систему, достаточно производительную для реализации в реальном масштабе времени, как алгоритмов компрессии речи, так и процедур шумоподавления, шифрования, конверсии голоса и других.

В главе 2 выполняется разработка метода анализа речевого сигнала на основе его декомпозиции на периодическую и аperiodическую компоненты.

На основе анализа существующих подходов делается вывод, что для повышения точности представления речевого сигнала целесообразно использовать модель “гармоники+шум / переходный сегмент”. Классификация речевого сегмента осуществляется с точки зрения возможности декомпозиции его на гармоническую и шумовую компоненты. Признаком такой возможности является вокализованность сегмента (рисунок 1).

Показано, что для успешного решения задачи декомпозиции речевого сигнала необходимо учитывать его нестационарность. При достаточно малой величине окна анализа (до 25 мс) можно предположить, что изменение частоты основного тона носит линейный характер:

$$h(i) = \sum_{k=1}^M A_k e^{-\beta i} \cos \left(\frac{2\pi k i}{F_s} \left(F_0 + \frac{\Delta F_0 i}{2N} \right) + \theta_k \right), \quad (1)$$

где F_0 – частота основного тона (ЧОТ); ΔF_0 – изменение частоты основного тона за N отсчётов; A_k – амплитуда k -ой гармоники основного тона; θ_k – фаза k -ой гармоники основного тона; M – количество гармоник; F_s – частота дискретизации; N – длина сегмента; β – фактор изменения; предполагается, что

значения гармонических амплитуд эволюционируют по экспоненциальному закону:

$$\beta = \ln(x)/T_0, \quad x = \frac{S_0^T S_{T_0}}{S_{T_0}^T S_{T_0}}, \quad (2)$$

где $S_0 = [|s_0|, \dots, |s_{N-T_0-1}|]^T$; $S_{T_0} = [|s_{T_0}|, \dots, |s_{N-1}|]^T$; T_0 – значение периода основного тона в отсчётах; s_0, s_1, \dots, s_{N-1} – отсчёты речевого сигнала.

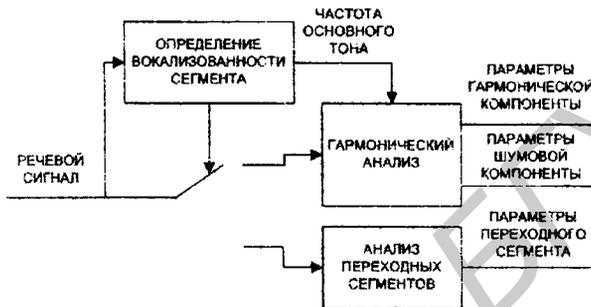


Рисунок 1 – Схема обработки речевого сигнала на основе его декомпозиции на периодическую и аperiodическую компоненты

С целью более точного представления периодической компоненты речи предлагается заменить классический спектральный анализ на основе кратковременного преобразования Фурье спектральным анализом, согласованным с изменением контура частоты основного тона (следящим гармоническим анализом):

$$X(k) = \sum_{n=0}^{N-1} x(n)w(n)e^{-j\varphi(n,k)}, \quad k = 1 \dots K, \quad (3)$$

где $X(k)$ – спектральный отсчет k -ой гармоники (коэффициент Фурье); $w(n)$ – временное окно; K – число всех гармоник основного тона во входном речевом сигнале.

Если частота основного тона меняется линейно на анализируемом фрейме на величину ΔF_0 , тогда аргумент под экспонентой будет вычисляться по формуле:

$$\varphi(n,k) = \frac{2\pi nkF_0}{F_s} + \frac{2\pi k\Delta F_0 n^2}{2NF_s} = \frac{2\pi nk}{F_s} \left(F_0 + \frac{\Delta F_0 n}{2N} \right). \quad (4)$$

Амплитуды и начальные фазы гармоник вычисляются в соответствии со следующими выражениями:

$$A_k = \sqrt{\operatorname{Re}^2 X(k) + \operatorname{Im}^2 X(k)},$$

$$\Phi_k = -\arctan\left(\frac{\operatorname{Im} X(k)}{\operatorname{Re} X(k)}\right). \quad (5)$$

Поскольку частоты синусоидальных компонент в (3) изменяются, такая система базовых функций не является ортогональной, что приводит к снижению точности анализа. С целью уменьшения погрешности определения параметров сигнала предлагается форму окна согласовывать с изменением частоты основного тона. Например, при использовании в качестве прототипа окна Кайзера (рисунок 2), согласованное окно описывается следующим образом:

$$w_n(i) = \frac{I_0\left(\beta\sqrt{1 - [(2x - L_n + 1)/(L_n - 1)]^2}\right)}{I_0(\beta)}, \quad (6)$$

где $I_0(x)$ – функция Бесселя нулевого порядка; β – параметр окна;

$$x = \frac{a_{2,n}(L_n - 1 - i)^2 + a_{1,n}(L_n - 1 - i)}{a_{2,n}(L_n - 1) + a_{1,n}}, \quad (7)$$

где $a_{1,n}$, $a_{2,n}$ – параметры, обеспечивающие линейное изменение частоты базовых функций анализа в соответствии с рассчитанной траекторией частоты основного тона:

$$a_{2,n} = \frac{2\pi(F_{0,n} - F_{0,n}^{(1)})}{2F_s N}, \quad a_{1,n} = \frac{2\pi F_{0,n}}{F_s}, \quad (8)$$

где $F_{0,n}$, $F_{0,n}^{(1)}$ частота основного тона в конце и в начале n -го окна анализа.

Алгоритм формирования матрицы компенсации неортогональности преобразования и коррекции результатов имеет следующий вид:

- а) вычислить частотные отсчеты $X(m)$;
- б) осуществить синтез косинусоидального сигнала соответствующей гармоники m :

$$H_m(n) = \cos((a_1 n + a_2 n^2)m); \quad (9)$$

- в) вычислить влияние m частотного отсчета на другие

$$e_{lm} = \sum_{n=0}^{Nl-1} H_m(n)(\cos(l(a_1 n + a_2 n^2))) - j \sin(l(a_1 n + a_2 n^2))); \quad (10)$$

- г) повторить пункты б) и в) для всех частотных отсчетов;

- д) построить матрицу компенсации неортогональности преобразования

$$E_1 = \begin{bmatrix} e_{11} & e_{12} & e_{13} & \dots & e_{1m} \\ e_{21} & e_{22} & e_{23} & & \\ e_{31} & e_{32} & e_{33} & & \\ \vdots & & & \ddots & \\ e_{l1} & & \dots & & e_{lm} \end{bmatrix}; \quad (11)$$

е) повторить пункты б) – д) алгоритма лишь с той разницей, что в пункте б) генерируется синусоидальный сигнал, а в пункте д) формируется матрица компенсации E_2 ;

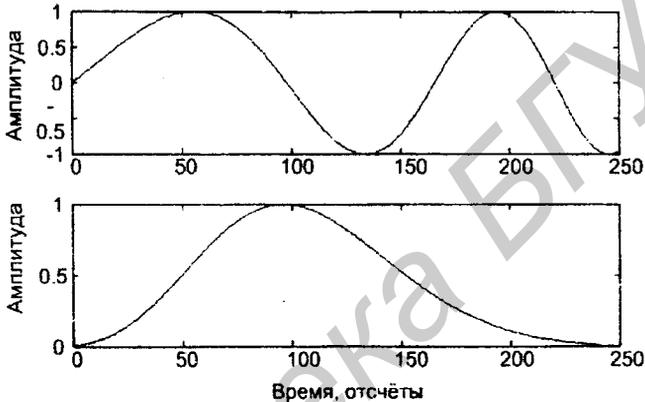


Рисунок 2 – Синусоида с линейно изменяющейся частотой (вверху) и соответствующая ей форма окна Кайзера (внизу)

ж) рассчитать новые значения частотных отсчетов в соответствии с формулой

$$\tilde{X} = E_1^{-1} \operatorname{Re} X - E_2^{-1} \operatorname{Im} X. \quad (12)$$

В главе показано, что качество декомпозиции критически зависит от точности определения частоты основного тона. Показателем качества декомпозиции предлагается использовать отношение “гармоники / шум”:

$$HNR = 10 \lg \frac{E_h}{E_r}, \quad (13)$$

где E_h и E_r – энергии гармонической и шумовой компоненты соответственно.

Показано, что зависимость $HNR(F_0)$ имеет локальный максимум в точке, равной частоте основного тона данного сегмента речи и носит унимодальный характер в её окрестности. На основании этого предлагается следующая схема определения частоты основного тона с одновременной декомпозицией речевого сигнала (рисунок 3).

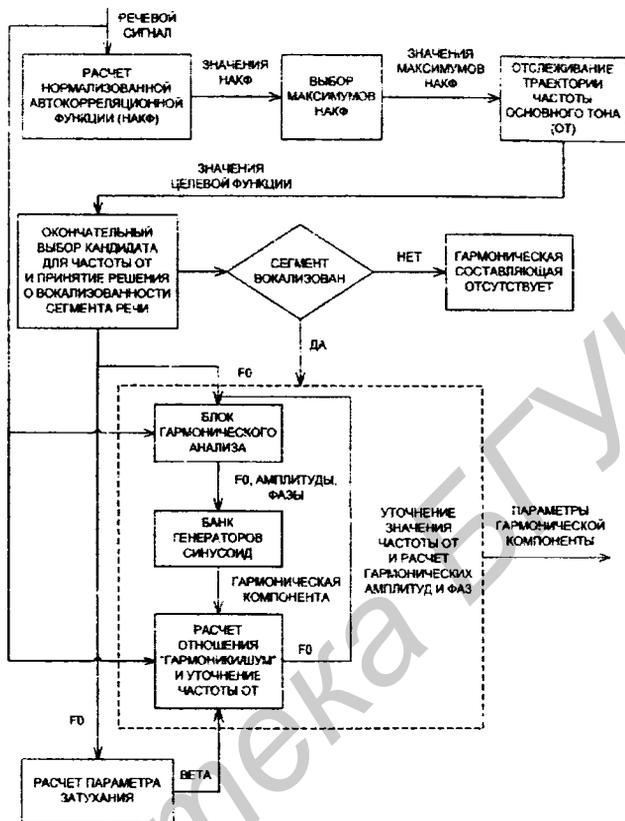


Рисунок 3 – Схема определения частоты основного тона

Алгоритм определения параметров гармонической компоненты выглядит следующим образом.

Шаг 1. Рассчитать значения нормализованной автокорреляционной функции (НАКФ) для всех допустимых значений периода основного тона:

$$\psi(p) = \frac{\sum_{j=1}^N s_j s_{j+p}}{\sqrt{\sum_{j=1}^N s_j^2 \sum_{j=1}^N s_{j+p}^2}}, \quad (14)$$

где p – период тона.

Шаг 2. Найти максимумы НАКФ и отобрать из соответствующих им значений периода тона кандидатов.

$$p_{ci} = \arg(\max(\psi(p))), \psi(p_{ci}) \geq 0.3 * \psi_{\max}(p), \quad (15)$$

где $\psi_{\max}(p)$ – максимальное значение НАКФ на данном сегменте.

Шаг 3. Для каждого кандидата рассчитать функцию стоимости с учетом прошлой информации о контуре частоты основного тона.

$$D_{i,j} = d_{i,j} + \min_{k \in I_{i-1}} \{D_{i-1,k} + \delta_{i,j,k}\}, \quad (16)$$

где: $d_{i,j}$ – локальная стоимость j -го кандидата в момент времени i , $\delta_{i,j,k}$ – стоимость перехода от k -го кандидата в момент времени $i-1$ к j -му кандидату в момент времени i , $1 \leq j \leq I$; I – количество кандидатов.

Шаг 4. Выбрать в качестве предварительной оценки периода основного тона на анализируемом сегменте кандидат j с минимальной стоимостью $D_{i,j}$.

Шаг 5. Если сегмент вокализован, провести процедуру расчёта параметров гармонической компоненты и одновременного уточнения значения частоты основного тона по формулам. Процедура расчёта параметров гармонической компоненты и одновременного уточнения значения частоты основного тона включает в себя следующие действия:

1) рассчитывается коэффициент затухания по формуле (2);

2) осуществляется поиск максимума отношения “гармоники / шум” в окрестности значения F_0^{est} . На каждой итерации рассчитываются векторы амплитуд A и фаз θ гармоник по формулам (3), (5) с последующим синтезом гармонической компоненты по формуле (1) и определением нового значения HNR с использованием формулы (13).

Шаг 6. Если сегмент вокализован, провести процедуру расчёта параметров гармонической компоненты и одновременного уточнения значения отклонения частоты основного тона. Здесь осуществляется поиск максимума отношения “гармоники / шум” в окрестности нулевого значения данного отклонения. На каждой итерации рассчитываются векторы амплитуд A и фаз θ гармоник по формулам (3), (5) с последующим синтезом гармонической компоненты по формуле (1) и определением нового значения HNR с использованием формулы (13).

Условием оптимальности будет следующее выражение

$$\Delta F_0^{opt} = \arg \max(HNR(F_0^{opt} + \Delta F_0)), \Delta F_0 \leq \Delta F_0 \leq \Delta F_0, \quad (17)$$

где диапазон $[\Delta F_0; \Delta F_0]$ представляет возможную область значений изменения частоты основного тона.

Глава 3 посвящена разработке метода квантования параметров используемой модели речевого сигнала. *Ставится задача* – оптимальным образом квантовать параметры гармоник периодической компоненты, представляя такие параметры минимальным количеством бит, необходимым для сохранения хорошего качества синтезированного сигнала.

В контексте гармонической модели проблема квантования в большей степени связана с передачей вектора амплитуд гармоник. На основании анализа

гармонической структуры спектров речевого сигнала различных дикторов (рисунок 4), можно сделать вывод, что векторы амплитуд гармоник, даже определяющие голос одного и того же диктора, имеют различную размерность в разные моменты времени.

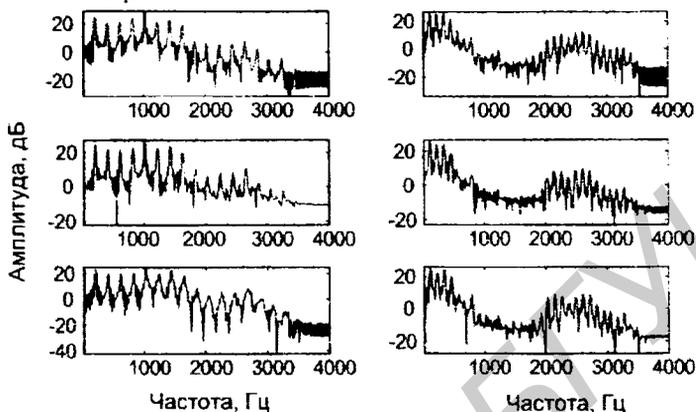


Рисунок 4 – Изменение спектров речи различных дикторов во времени

Анализ существующих подходов показал недостаточность применяемых методов квантования для использования в системах компрессии речи, использующих гармоническую модель речевого сигнала. Предложен метод квантования амплитуд гармоник с переменной размерностью вектора, в котором отсутствует необходимость дополнительных преобразований и нерациональные требования к объему памяти. Метод тренировки кодовых книг с использованием алгоритма *K*-средних адаптирован для квантования вектора амплитуд гармоник с переменной размерностью.

Поскольку кодовые книги имеют большую длину кодовых слов, имеет смысл ограничить размерность квантуемого вектора таким образом, чтобы не учитывать достаточно малые амплитуды.

Применяемая модель анализа речевого сигнала предполагает разделение речи на гармоническую и шумовую компоненту по всему спектру. Используя закономерности психоакустики, можно определить, в какой степени шумовая компонента влияет на восприятие человеком гармонической компоненты, т.е. определить гармоники, не влияющие на восприятие речи в целом. Для решения данной проблемы использовалась психоакустическая модель Джонстона. Максимальной частотой вокализованности считается последняя гармоника частоты основного тона, превышающая порог маскирования.

Экспериментальные результаты свидетельствуют, что величина ошибки квантования зависит главным образом от длины кодовой книги. Метод определения граничной частоты вокализованности, учитывающий особенности

восприятия речи человеком, позволяет повысить качество реконструированной речи.

В **4-й главе** дано описание структуры и функционирования устройства компрессии. Обсуждаются проблемы и особенности аппаратной реализации данной системы. Представлены характеристики отладочного модуля и разработанной платформы. Даны и детально описаны алгоритмы компрессии и декомпрессии речевого сигнала, даётся оценка вычислительной сложности применяемых алгоритмов.

На основе анализа вычислительной сложности алгоритмов в качестве основы для аппаратной реализации устройства компрессии предлагается использовать ЦПОС с VLIW-архитектурой с поддержкой SIMD-инструкций из серии TMS320C6x. Процессоры данной серии имеют достаточную для используемого алгоритма компрессии вычислительную мощность, содержат оперативную память на кристалле и широкий набор аппаратных интерфейсов.

Разработано алгоритмическое обеспечение устройства компрессии-декомпрессии речи, основанного на гибридной модели и методе декомпозиции речевого сигнала на периодическую и аperiodическую компоненты. Устройство обеспечивает скорость передачи 2,4 и 4,8 кбит/с.

Разработана высокопроизводительная вычислительная платформа, построенная на базе двухпроцессорного ядра, в состав которого входят ЦПОС TMS320C6713B и ПЛИС Xilinx семейства Spartan-3 XC3S200 со структурой *FPGA*. Данная платформа является универсальной и подходит для решения широкого круга задач, касающихся разработки мобильных мультимедийных устройств реального времени.

Предъявленные требования к объему памяти и вычислительной мощности позволяют говорить о возможности эффективной аппаратной реализации рассматриваемого устройства на базе выбранной платформы и принятого архитектурного решения.

Разработана текстозависимая система конверсии голоса на базе гибридной модели и декомпозиции речевого сигнала на периодическую и аperiodическую компоненты, что демонстрирует универсальность модели анализа-синтеза. Система конверсии голоса может использоваться как дополнительная опция в системе кодирования речи, так и автономно.

Глава 5 является экспериментальной частью диссертационной работы. В главе проведены исследования метода декомпозиции речевого сигнала, а также оценки качества восстановленной в декодере речи.

Показано, что метод декомпозиции речевого сигнала на периодическую и аperiodическую компоненты, основанный на дискретном преобразовании Фурье, согласованном с изменением контура частоты основного тона, превосходит свои аналоги PAPD и PSHF.

При тестировании устройств компрессии с помощью субъективных методик предполагалось, что качество восстановленной речи соответствует общему впечатлению слушателя от принимаемой речи. Качество речи определяется как параметр, представляющий субъективную оценку звучания речи в испытываемой системе передачи речи, выраженную в баллах по пятибалльной шкале или в процентах предпочтения при сравнении с эталонным трактом.

В рамках НИР «Разработать процедуры сжатия речевой информации, обеспечивающие коммерческое качество восстановленной речи» в соответствии с методикой, приведённой в СТБ ГОСТ 50840-2000, для предлагаемого устройства были получены следующие результаты (скорость потока данных 2,4 и 4,8 кбит/с):

- 1) слоговая разборчивость речи – 96 %;
- 2) субъективное качество речи – 3,5 и 3,6 баллов соответственно;
- 3) особенности восстановленной в декодере речи – некоторое нарушение естественности и узнаваемости, иногда присутствуют всхлипывания и искажения тембра голоса;
- 4) узнаваемость голоса диктора – 94 и 95 % соответственно.

Результаты субъективной оценки качества синтезированной речи подтверждают, что речь отличается довольно высокой степенью разборчивости и хорошей узнаваемостью диктора.

В рамках НИР «Разработать процедуры сжатия речевой информации, обеспечивающие коммерческое качество восстановленной речи» были получены объективные оценки качества реконструированной речи (по параметру искажений спектра барков *MBSD*), которые показывают, что разработанный кодер речи, основанный на декомпозиции речевого сигнала на периодическую и аperiodическую компоненты, инвариантен к языку и полу диктора.

Сравнительный анализ показывает, что оба разработанных устройства компрессии речи 2400 бит/с и 4800 бит/с по показателю *MBSD* превосходят кодек ITU-T G.726 16000 бит/с и кодер MELP 2400 бит/с; кодер 4800 бит/с превосходит кодек ITU-T G.723.1 5300 бит/с и G.729 8000 бит/с; кодер 4800 бит/с превосходит кодеки AMR 4750 бит/с и 5150 бит/с и сопоставим по качеству с кодеками AMR 5900 бит/с, 6500 бит/с и 7400 бит/с.

В приложениях диссертации даны методики определения объективных оценок качества звучания реконструированной речи, описание алгоритма выбора контура частоты основного тона методом динамического программирования и акты внедрения результатов диссертационной работы.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

В ходе выполнения представленной диссертационной работы были получены следующие научные результаты.

1. На основании анализа существующих подходов параметрического кодирования речевых сигналов применительно к системам компрессии показано, что для достижения высокой степени компрессии и качества требуется точная декомпозиция речевого сигнала на периодическую и аperiodическую компоненты для их последующего анализа [2-А, 4-А, 8-А, 9-А, 10-А]. Для повышения степени адекватности модели речевого сигнала необходимо выделение переходных сегментов для анализа в отдельный класс [1-А, 2-А, 13-А]. В качестве вычислительной платформы предлагается использовать *VLIW*-процессор с поддержкой *SIMD*-инструкций, позволяющий осуществлять параллельно-поточную обработку данных в реальном масштабе времени.

2. Предложено построение системы анализа-синтеза речи на базе гибридной схемы, предполагающей амплитудную и частотную модулированность речевого сигнала на участке анализа [2-А].

3. Разработан метод декомпозиции речевого сигнала на основе гармонического следящего анализа, позволяющий значительно более качественно выделять периодическую компоненту по сравнению с подобными алгоритмами. Применение следящего анализа позволяет согласовать определение параметров периодической компоненты с изменением контура частоты основного тона [2-А, 7-А, 12-А] и тем самым добиться повышения точности декомпозиции. Метод декомпозиции превосходит аналогичные методы *PAPD* и *PSHF*.

4. Разработан метод определения частоты основного тона с одновременной декомпозицией речевого сигнала в цикле с обратной связью [2-А, 5-А, 9-А]. Данный подход позволяет увязать точность определения частоты основного тона с качеством декомпозиции. Оценка качества метода [6-А] показала, что количество грубых ошибок определения частоты основного тона снизилось в пределах от 1 до 30 % по сравнению с аналогами, количество ошибок классификации сегмента снизилось в пределах от 7 до 40 % по сравнению с аналогами. Метод сохраняет превосходство над аналогами при определении частоты основного тона в речи, зашумленной аддитивным белым шумом с *SNR* до 5 дБ.

5. Разработан метод квантования гармонических параметров модели, учитывающий особенности психоакустики [2-А, 3-А]. Метод обучения кодовых книг с использованием алгоритма *K*-средних адаптирован для квантования

векторов различной размерности. Использование психоакустической оптимизации позволило повысить perceptуальное качество реконструированной речи на 0,45 по шкале MBSD.

6. Создано алгоритмическое обеспечение гибридного вокодера с психоакустической оптимизацией, особенностью которого является декомпозиция речевого сигнала на периодическую и аperiodическую компоненты с последующим отдельным анализом [2-А, 3-А].

7. Осуществлена аппаратная реализация устройства компрессионно-декомпрессионного речевого сигнала со скоростями передачи 2,4 и 4,8 кбит/с [2-А] на базе платформы *VLIW* ЦПОС Texas Instruments TMS320C6713B с реконфигурируемой логикой на основе ПЛИС Xilinx Spartan-3. Данная платформа является универсальной и подходит для решения широкого круга задач, касающихся разработки мобильных мультимедийных устройств реального времени. Предлагаемый кодек 2400 бит/с сопоставим по показателю *PESQ* с кодером MELPе, по показателю *MBSD* превосходит кодек ITU-T G.726 16000 бит/с и кодек MELP 2400 бит/с. Предлагаемый кодек 4800 бит/с превосходит кодек ITU-T G.723.1 5300 бит/с и G.729 8000 бит/с, кодеки AMR 4750 бит/с и 5150 бит/с, сопоставим по качеству с кодеками AMR 5900 бит/с, 6500 бит/с и 7400 бит/с.

8. Показана возможность построения различных систем обработки речи на базе разработанной модели на примере системы конверсии голоса [1-А, 11-А, 12-А, 13-А]. Экспериментальные результаты показывают, что данная система конверсии превосходит по качеству системы конверсии, построенные на базе моделей анализа ACELP и HNM.

Рекомендации по практическому использованию результатов

Концепция гибридного представления речевого сигнала с использованием гармонического следа для его декомпозиции позволяет получить восстановленный речевой сигнал с сохранением помимо разборчивости и узнаваемости его интонационных признаков.

Система компрессионного речевого сигнала, разработанная на базе данной концепции, 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 (IP-телефония). Кроме того данная система может быть расширена системой конверсии голоса. В свою очередь, система конверсии голоса может применяться в областях синтеза речи по тексту (устранение “компьютерного акцента”), а также в судебной практике (защита свидетелей).

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в научных журналах

1-А. Конверсия голоса с использованием модели сепарации речевого сигнала на компоненты “гармоники+шум” и переходные фреймы / А.Н. Павловец, М.З. Лившиц, Д.С. Лихачев, А.А. Петровский // Речевые технологии. – 2008. – №4. – С. 37–49.

2-А. Павловец, А.Н. Гармоническая модель речевого сигнала: определение параметров и их квантование / А.Н. Павловец, П.Р. Зубрыcki, А.А. Петровский // Доклады Белорусского государственного университета информатики и радиоэлектроники. – 2007. – № 4 (20). – С. 19–34.

3-А. Павловец, А.Н. Использование закономерностей психоакустики в процедуре квантования параметров гармонической модели речевого сигнала / А.Н. Павловец, А.А. Петровский // Речевые технологии. – 2008. – №4. – С. 50–60.

4-А. Павловец, А.Н. Квантование огибающей спектра в вокоде, основанное на декомпозиции речевого сигнала на периодическую и аperiodическую составляющие / А.Н. Павловец, А.А. Петровский // Цифровая обработка сигналов. – 2005. – № 3. – С. 13–21.

5-А. Павловец, А.Н. Метод определения частоты основного тона на основе субгармонического анализа / А.Н. Павловец // Известия Белорусской инженерной академии. – 2003. – № 1(15)/2. – С. 244–247.

6-А. Павловец, А.Н. Методика объективной оценки качества алгоритмов для определения частоты основного тона речевого сигнала / А.Н. Павловец, А.А. Петровский // Известия Белорусской инженерной академии. – 2004. – № 1(17)/1. – С. 159–162.

Статьи в научных сборниках

7-А. Petrovsky, A Split vector quantization of psychoacoustical modified LSF coefficients in speech coder based on pitch-tracking periodic-aperiodic decomposition / A. Petrovsky, A. Sawicki, A. Pavlovec // Information processing and security systems. – Eds. K. Saeed, J. Pejas – Springer, 2005. – P. 67–76.

8-А. Sawicki, A. Incorporating psychoacoustical principles in LSF quantization for pitch synchronous speech coder / A. Sawicki, A. Pavlovec, A. Petrovsky // Computer information systems and applications. – Eds. K. Saeed [et al.]. – Bialystok: University of finance and management, 2004. – Vol. 1. – P. 88–97.

9-А. Zubrycki, P. Analysis-by-synthesis parameters estimation in the harmonic coding framework by pitch tracking modified DFT / P. Zubrycki, A. Pavlovec, A. Petrovsky // New trends in audio and video / Eds. A. Dobrucki, A. Petrovsky and W. Skarbek. – Bialystok: Politechnica Bialostocka, 2006. – P. 233–246.

10-А. Павловец, А.Н. Исследование векторного квантования расщепленного множества LSF-коэффициентов в приложении к кодированию периодической компоненты речевого сигнала в вокодерных системах / А.Н. Павловец, А.А. Петровский, А. Савицкий // Труды РНТОРЭС им. А. С. Попова, серия "Цифровая обработка сигналов и её применение" – 2005. – Вып. VII-1, том 1. – С. 179–181.

Материалы конференций

11-А. Pavlovets, A. Voice conversion as a part of the voice analysis/synthesis system based on the periodic-aperiodic decomposition of speech / A. Pavlovets, A. Petrovsky // Pattern recognition and information processing: proceedings of the 9th int. conf. (PRIP'2007), Minsk, Belarus, May 22–24, 2007. – Minsk, 2007. – Vol. II. – P. 71–76.

12-А. Speech analysis-synthesis based on the PTDF for voice conversion / A. Pavlovets, T. Kien, P. Zubrycki, A. Petrovsky // Spectral methods and multirate signal processing: proceedings of the 2007 Int. TICSP Workshop (SMMSP'2007), Moscow, Russia, September 1–2, 2007. – Tampere, 2007. – P. 203–210.

13-А. Voice conversion based on the HNT model of speech and separate VQ learning / A. Pavlovets, M. Livshitz, D. Lichachov, A. Petrovsky // Speech and computer: proceedings of the 13th int. conf. (SPECOM'2009), St. Petersburg, Russia, June 21–25, 2009 – St. Petersburg, 2009. – P. 283–288.

Тезисы докладов

14-А. Павловец, А.Н. Субгармонический анализ в системах кодирования речевых сигналов / А.Н. Павловец // Доклады БГУИР. Электроника, материалы, технологии, информатика. – 2003. – том 1, № 2/1. – С. 44.



РЭЗІЮМЭ

Паўлавец Аляксандр Мікалаевіч

Метады і алгарытмы параметрычнай апрацоўкі маўленчага сігналу з псіхаакустычнай аптымізацыяй у сістэмах мультымедыа рэальнага часу

Ключавыя словы: кампрэсія мовы, параметрычная апрацоўка, гібрыдная апрацоўка, дэкампазіцыя маўленчага сігналу, гарманічны аналіз з сачэннем.

Мэтай працы з'яўляецца распрацоўка метадаў і алгарытмаў аналізу і сінтэзу для прылад параметрычнай апрацоўкі мовы, якія выкарыстоўваюць дэкампазіцыю маўленчага сігналу на перыядычную і аперыядычную складаючыя і псіхаакустычную аптымізацыю.

Аб'ектам даследавання з'яўляецца прылада параметрычнага кодэра мовы на аснове дэкампазіцыі мовы на перыядычную і аперыядычную кампаненты і псіхаакустычнай аптымізацыі. Прадмет даследавання – метады кампрэсіі мовы на аснове асобнай апрацоўкі перыядычнай і аперыядычнай кампанент маўленчых сігналаў, якія ўжываюць псіхаакустычную аптымізацыю для больш якаснага квантавання.

Навуковая навізна атрыманых аўтарам вынікаў складаецца ў наступным: распрацаваны новы метад параметрычнай апрацоўкі маўленчага сігналу на аснове гарманічнага аналізу з сачэннем, які дазваляе больш дакладную адзнаку параметраў перыядычнай кампаненты мовы пры непастаяннай фундаментальнай частаце маўленчага сігналу; распрацаваны метад і алгарытм адначасовага вызначэння частаты асноўнага тона і параметраў гармонік, якія забяспечваюць павышаную дакладнасць вызначэння контура частаты асноўнага тона за кошт ўжывання цыкла са зваротнай сувяззю з вызначэннем амплітуд і фаз гармонік; распрацаваны метад квантавання амплітуд гармонік мовы як вектараў пераменнай даўжыні, які ўлічвае асаблівасці слыхавай сістэмы чалавека і дазваляе павялічыць якасць рэканструяванай мовы за кошт больш дакладнага квантавання перцэптуальна важных амплітуд гармонік; распрацаваны спосаб кадавання мовы на аснове гібрыднай мадэлі, якая выкарыстоўвае дэкампазіцыю маўленчага сігналу, які дазваляе павялічыць ступень адпаведнасці мадэлі прыродзе апрацоўваемага адрэзку мовы; створана апаратна-праграмавая рэалізацыя прылады кадавання мовы, якая забяспечвае хуткасць перадачы 2,4 і 4,8 кбіт/с.

Вынікі дысертацыйнай працы ўкаранёныя і выкарыстоўваюцца у Дзяржаўным цэнтры бяспекі інфармацыі пры Прэзідэнце Рэспублікі Беларусь, Упраўленні ўрадавай сувязі КДБ Рэспублікі Беларусь, інжэнерна-тэхнічным ЗАА “Ай-Ти Мобайл”(г. Масква), НДЛ 5.3 БДУІР “Матэрыялы і элементы электроннай і звышправадніковай тэхнікі” і у навучальным працэсе спецыяльнасці “Электронныя вылічальныя сродкі”.

РЕЗЮМЕ

Павловец Александр Николаевич

Методы и алгоритмы параметрической обработки речевых сигналов с психоакустической оптимизацией в системах мультимедиа реального времени

Ключевые слова: компрессия речи, параметрическая обработка, гибридная обработка, декомпозиция речи, гармонический следящий анализ.

Целью работы является разработка методов и алгоритмов анализа и синтеза для устройств параметрической обработки речи, использующих декомпозицию речевого сигнала на периодическую и аperiodическую компоненты и психоакустическую оптимизацию.

Объектом исследования является устройство параметрического кодера речи на основе декомпозиции речевого сигнала на периодическую и аperiodическую компоненты и психоакустической оптимизации. Предмет исследования – методы компрессии речевых сигналов, основанные на отдельной обработке периодической и аperiodической компонент речи и использующие психоакустическую оптимизацию для более качественного квантования.

Научная новизна полученных автором результатов заключается в следующем: разработан метод параметрической обработки речи на основе гармонического следящего анализа, обеспечивающий более точную оценку параметров периодической компоненты при изменяющейся фундаментальной частоте речи; разработаны метод и алгоритм одновременного определения частоты основного тона и параметров гармоник, повышающий точность оценки контура частоты основного тона за счет применения цикла с обратной связью с определением амплитуд и фаз гармоник; предложен метод квантования амплитуд гармоник как векторов переменной длины, повышающий качество реконструированной речи за счет более точного квантования перцептуально значимых амплитуд; разработан способ кодирования речи на основе гибридной модели, использующей декомпозицию речевого сигнала, повышающий степень соответствия модели природе речевого отрезка; разработана аппаратно-программная реализация устройства кодирования речевого сигнала, обеспечивающая скорость передачи 2,4 и 4,8 кбит/с.

Результаты диссертационной работы внедрены и используются в Государственном центре безопасности информации при Президенте Республики Беларусь, Управлении правительственной связи КГБ Республики Беларусь, инженерно-техническом ЗАО “Ай-Ти Мобайл”(г. Москва), НИЛ 5.3 БГУИР “Материалы и элементы электронной и сверхпроводниковой техники” и в учебном процессе специальности “Электронные вычислительные средства”.

SUMMARY

of the dissertation by Pavlovets Alexander Nikolaevich

The methods and algorithms of parametric speech processing with psychoacoustical optimization in real-time multimedia systems

Keywords: speech compression, parametric processing, hybrid processing, speech decomposition, harmonic tracking analysis.

The purpose of the work is a development of methods and algorithms of analysis and synthesis for parametric speech processing devices based on the speech decomposition into periodic and aperiodic components.

The object of the research is a parametric speech coder based on the speech signal decomposition into periodic and aperiodic components. The subjects of the research are methods of speech compression based on the separate processing of periodic and aperiodic components and psychoacoustical optimization for quantization.

The scientific novelty of the results received by the author consists of the following issues: method of parametric speech processing based on harmonic tracking analysis that gives more accurate speech periodic component parameters estimation at varying pitch conditions and more accurate periodic and aperiodic components extraction; method and algorithm of simultaneous pitch and periodic component parameters estimation, that gives higher accuracy estimation of pitch contour because of closed-loop cycle with harmonic amplitudes and phases estimation; method for harmonic vector quantization based on variable-dimension vector quantization that takes into consideration human auditory system properties, and lets to increase synthesized speech quality at the cost of more accurate perceptually important harmonic amplitudes quantization; speech coding method based on the hybrid model, that utilizes speech signal decomposition at periodic and aperiodic components and allows to increase degree of compliances of the model used with the speech frame nature; real-time hardware and software implementation of speech coding device with rates of 2,4 и 4,8 kbps.

Results of the dissertational work are implemented into practice and used in: in the State Center of Information Security at the President of the Republic of Belarus, in the Government communication department of CGS of the Republic of Belarus, in the "IT Mobile" company (Moscow), in the scientific research laboratory "Materials and elements of electronic and superconductor devices" № 5.3 at BSUIR and in the educational process at the speciality "Electronic computing devices".

Научное издание

ПАВЛОВЕЦ Александр Николаевич

**МЕТОДЫ И АЛГОРИТМЫ
ПАРАМЕТРИЧЕСКОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ
С ПСИХОАКУСТИЧЕСКОЙ ОПТИМИЗАЦИЕЙ
В СИСТЕМАХ МУЛЬТИМЕДИА РЕАЛЬНОГО ВРЕМЕНИ**

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 08.10.2010.

Гарнитура «Таймс».

Уч.-изд. л. 1,3.

Формат 60x84 1/16.

Отпечатано на ризографе.

Тираж 60 экз.

Бумага офсетная.

Усл. печ. л. 1,63.

Заказ 711.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники»
Ч371 от 16.03.2009. ЛП №02330/0494175 от 03.04.2009.
220013, Минск, П. Бровки, 6