

Учреждение образования
“БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ”

УДК 004.032.6 + 004.383.3

ЛИХАЧЁВ
Денис Сергеевич

**АНАЛИЗ И СИНТЕЗ УСТРОЙСТВ КОДИРОВАНИЯ РЕЧЕВОГО
СИГНАЛА НА ОСНОВЕ АНТРОПОМОРФИЧЕСКОЙ ОБРАБОТКИ И
СИНУСОИДАЛЬНЫХ МОДЕЛЕЙ**

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

специальность 05.13.05
«Элементы и устройства вычислительной техники и систем управления»

Минск 2006

Работа выполнена в Учреждении образования “Белорусский государственный университет информатики и радиоэлектроники”

Научный руководитель – Петровский Александр Александрович, д.т.н, профессор, зав. кафедрой электронных вычислительных средств, Учреждение образования “Белорусский государственный университет информатики и радиоэлектроники”

Официальные оппоненты: Тормышев Юрий Иванович, д.т.н., профессор, профессор кафедры электронных вычислительных машин, Учреждение образования “Белорусский государственный университет информатики и радиоэлектроники”

Бовбель Евгений Иванович, к.ф.-м.н., доцент, доцент кафедры радиофизики, Учреждение образования “Белорусский государственный университет”

Оппонирующая организация – Государственное научное учреждение “Объединенный институт проблем информатики НАН Беларуси”.

Защита состоится 21 декабря 2006 г. 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: dissovvet@bsuir.unibel.by, тел. 293-89-89.

С диссертацией можно ознакомиться в библиотеке Учреждения образования “Белорусский государственный университет информатики и радиоэлектроники”.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

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 “Антропоморфические методы и алгоритмы обработки речи и аудиосигналов” при финансовой поддержке Белорусского Республиканского Фонда фундаментальных исследований (ГБЦ 04-7041), а также НИР “Разработка низкоскоростного речепреобразующего устройства с препроцессором редактирования шума и компенсатором эхо-сигнала для аппаратуры М-461” (х/д 05-1160).

Цель и задачи исследования

Целью работы является разработка методов и алгоритмов анализа и синтеза устройств кодирования речи с высокой степенью компрессии, использующих антропоморфическую обработку речевого сигнала.

Для достижения данной цели необходимо решить следующие задачи.

1. Проанализировать подходы построения синусоидальных кодеров речи с целью выбора методов анализа речевого сигнала на основе синусоидальной модели, обеспечивающих высокую эффективность системы компрессии.

2. Разработать специальные методы и алгоритмы антропоморфической обработки речевых сигналов, адаптированных для применения в устройствах синусоидального кодирования речи, с целью достижения высокой степени компрессии и повышения качества синтезируемого сигнала.

3. Провести анализ существующих методов и подходов синтеза речевого сигнала на основе синусоидальных генераторов и выбрать наиболее подходящий метод с точки зрения не только высокого качества восстановленного речевого сигнала, но и хорошей совместимости с алгоритмами антропоморфической обработки.

4. Необходимо разработать эффективный метод квантования параметров синусоидальной модели речевого сигнала для их представления минимальным количеством бит.

5. Разработать аппаратно-программную реализацию устройства компрессии речи на основе синусоидальной модели с антропоморфической обработкой, предназначенного для работы в реальном масштабе времени.

Объектом исследования является устройство синусоидального кодера речи с антропоморфической обработкой речевого сигнала. Предмет исследования – методы компрессии речевых сигналов на основе синусоидальной модели, использующие антропоморфический анализ для выбора наиболее значимой для восприятия информации.

Решение рассматриваемых в диссертации задач требует применения следующих методов исследования: методов цифровой обработки сигналов, психоакустики, акустики, математического моделирования, теории вероятностей и математической статистики, методов проектирования микропроцессорных и проблемно-ориентированных вычислительных средств.

Положения, выносимые на защиту

1. Структурная схема устройства компрессии речи на основе совмещения синусоидальной модели с процедурой антропоморфического анализа речевого сигнала.

2. Метод и алгоритм антропоморфической обработки речевого сигнала в частотной области на основе трёх слуховых моделей: модели внешнего и среднего уха, комбинированной кохлеарной модели и модели слухового нерва человека, позволяющие производить выделение доминирующих (наиболее важных для слуха человека) частотных составляющих анализируемого сигнала.

3. Метод анализа речевого сигнала в частотной области на основе комбинированной кохлеарной модели (активной и пассивной) и преобразования Фурье с неравномерным частотным разрешением.

4. Метод квантования синусоидальных параметров на основе векторного квантования с частотной коррекцией, учитывающий особенности статистического распределения параметров (амплитуд, частот и фаз) и степень их влияния на качество синтезированной речи.

5. Особенности аппаратно-программной реализации устройства компрессии с антропоморфической обработкой речевого сигнала на основе процессора обработки сигналов TMS320C6713, функционирующего в реальном масштабе времени в составе комплекса мультимедиа устройств 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, с

разработкой математического описания модели улитки уха человека. 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. Для получения объективных оценок использовалось специальное программное обеспечение, разработанное доц. каф. ЭВМ Петровским А.А.

Апробация результатов диссертации

Результаты, полученные в ходе выполнения исследований, докладывались и обсуждались на: 4-й и 5-й международных конференциях “Современные средства связи” (Нарочь, Беларусь, 2001, 2002); the 6-th international workshop “Systems, Signals and Image Processing” (Manchester, UK, 2002); 5-й Международной конференции “Цифровая обработка сигналов и ее применение” (Москва, Россия, 2002); Российско-белорусских научно-технических конференциях “Технические средства защиты информации” (Нарочь, Беларусь, 2003, 2004); international conference “Computer Information Systems and Industrial Management Applications” (Elk, Poland, 2003); the 3-rd international conference “Neural Networks and Artificial Intelligence” (Minsk, Belarus, 2003); the 9-th international conference “Speech and Computer” (Saint-Petersburg, Russia, 2004); the 8-th international conference “Pattern Recognition and Information Processing” (Minsk, Belarus, 2005); научно-технических конференциях БГУИР (Минск, Беларусь, 2002–2006).

Результаты диссертационной работы внедрены и используются в:

- 1) инженерно-техническом ООО “КВИНТЕЛ”;
- 2) Государственном научном учреждении “ОИПИ НАН Беларуси” (лаборатория “Распознавания и синтеза речи”);
- 3) Государственном центре безопасности информации при Президенте Республики Беларусь;
- 4) в учебном процессе специальности “Электронные вычислительные средства” (БГУИР).

Опубликованность результатов диссертации

По тематике представленной диссертационной работы было опубликовано 15 печатных работ, включая 2 тезиса докладов на научных конференциях и 13 статей в журналах и материалах международных научных конференций и симпозиумов. Суммарный объем публикаций составляет около 6,42 авторских листов. Результаты работы включены в 7 отчетов о НИР.

Структура и объём диссертации

Диссертационная работа состоит из введения, пяти глав, заключения и приложений. В **главе 1** проведён анализ существующих методов синусоидального кодирования речи и выполняется предварительный выбор аппаратной платформы для реализации устройства компрессии. В **главе 2** осуществляется разработка метода антропоморфической обработки речевого сигнала в частотной области на основе трёх модифицированных моделей слуха человека и предлагается структурная схема устройства компрессии речи на его основе. **Глава 3** посвящена разработке метода квантования синусоидальных параметров. В **4-й главе** дано описание структуры и функционирования устройства компрессии речи, обсуждаются проблемы и особенности аппаратной реализации данной системы в реальном масштабе времени. В **5-й главе** приведены результаты субъективных и объективных оценок восстановленной в декодере речи. В **приложениях** диссертации даны описание кохлеарной модели, алгоритм расчёта банка кохлеарных фильтров и акты внедрения результатов диссертационной работы.

Работа представлена на 178 страницах машинописного текста, в том числе основная часть – на 129 страницах. Она содержит 105 рисунков, 8-таблиц, список использованных источников из 158 наименований, в том числе – 15 авторских и 2 приложения объёмом 21 страница.

СОДЕРЖАНИЕ РАБОТЫ

В **главе 1** дан краткий обзор развития концепции синусоидального кодирования речи, дано определение синусоидальной модели речевого сигнала, отличительной особенностью которой является то, что речь $s(n)$, как на вокализованных, так и на невокализованных участках, представляется в виде:

$$s(n) = \sum_{i=1}^L A_i \sin(\omega_i n + \varphi_i), \quad (1)$$

где n – номер отсчета;

L – количество синусоидальных составляющих;

A_i , ω_i и φ_i – амплитуда, частота и фаза i -й синусоидальной составляющей соответственно.

Главным достоинством данной модели является её способность обеспечивать более высокое качество синтезированной речи по сравнению с существующими подходами.

Основными недостатками систем обработки речи на основе синусоидальной модели является то, что для достижения хорошего качества

восстанавливаемой речи сигнал должен быть представлен достаточно большим количеством параметров, что непригодно для эффективной компрессии речевого сигнала. При ограничении количества синусоидальных компонент резко деградирует качество речи.

На основе анализа существующих подходов синусоидального кодирования речевых сигналов делается вывод, что для достижения высокой степени компрессии требуется ограничить объём кодируемой информации за счёт отбора только наиболее перцептуально значимых (доминирующих) составляющих сигнала. Для решения данной проблемы предлагается использовать процедуру *антропоморфической обработки речевого сигнала* (рисунок 1).

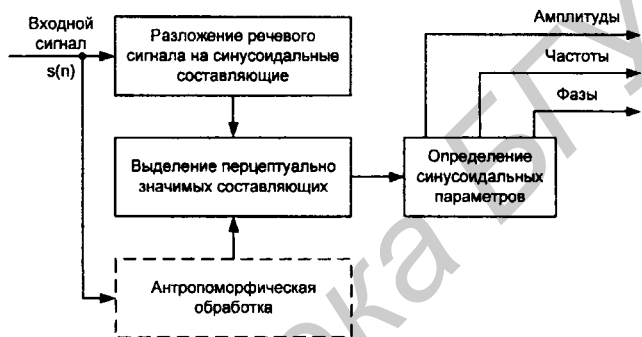


Рисунок 1 – Обобщённая схема синусоидального кодера с антропоморфической обработкой речи

Применение принципа «антропоморфической обработки сигнала» в данном случае предполагает использование таких устройств и алгоритмов обработки информации, когда вычислительный процесс организовывается по «образу и подобию» человека, т.е. применяемые способы и алгоритмы моделируют какие-либо процессы, происходящие в его слуховых и речеобразующих системах. При этом предполагается, что при достаточно точном моделировании используемые модели будут иметь те же полезные свойства, что и их физиологический аналог.

Возникает задача дифференциации элементов слуховой системы человека по значимости с точки зрения механизма обработки звуковой информации и применимости для задач компрессии. Моделирование «второстепенных» элементов не только сильно увеличит алгоритмическую и вычислительную сложность, но и затруднит процесс интерпретации полученных результатов.

Основной проблемой является то, что существующие в настоящее время алгоритмы антропоморфической обработки речевого сигнала предназначены для использования в задачах распознавания и идентификации диктора и не

адаптированы для применения в системах компрессии сигналов на основе синусоидальной модели. К тому же, существующие алгоритмы либо предъявляют очень высокие требования к вычислительным ресурсам системы, что не в лучшую сторону сказывается на сложности, стоимости системы компрессии и возможности эффективного функционирования в реальном масштабе времени, либо они не вполне адекватны происходящим в системе слуха человека реальным физиологическим процессам, что сказывается на качестве синтезируемой речи.

Следовательно, возникает необходимость в создании новых или модификации существующих антропоморфических алгоритмов для обеспечения возможности их использования в системах компрессии речевых сигналов на основе синусоидальной модели. Кроме того, устройство для синусоидального кодирования речи должно базироваться на практически реализуемых в реальном масштабе времени и в тоже время адекватных алгоритмах антропоморфического анализа.

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 и т.п.

В главе 2 выполняется разработка метода антропоморфической обработки речевого сигнала на основе спектрального анализа, совмещённого с моделями слуха человека. В качестве основной задачи для антропоморфической обработки ставится выработка такого критерия, который позволял бы отбирать наиболее важные для восприятия доминирующие частотные компоненты анализируемого сигнала. В этом случае предполагается, что речь будет синтезироваться как сумма синусоид с параметрами, которые определяются на доминирующих частотах.

На основе анализа существующих подходов делается вывод, что для успешного решения задачи по выделению доминирующих частотных составляющих необходимо не просто смоделировать процессы, происходящие в периферической части слуховой системы человека (главным образом в улитке уха человека), но и проанализировать информацию, которая циркулирует на уровне слухового нерва человека.

Поэтому в качестве основы для антропоморфической обработки предлагается использовать гибридную слуховую модель, включающую следующие части: модель наружного и среднего уха, комбинированную

кохлеарную модель (активную и пассивную) и модель слухового нерва (рисунок 2).

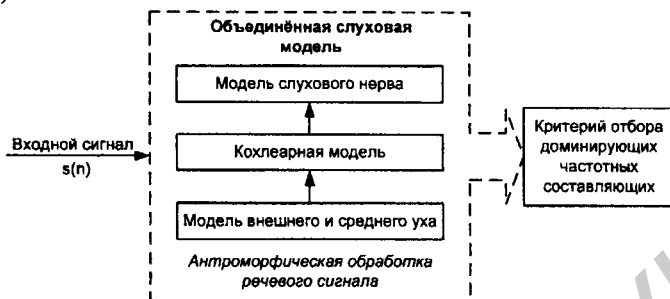


Рисунок 2 – Обобщённая схема выделения доминирующих частотных составляющих на основе объединённой слуховой модели

Результаты работы данных слуховых моделей необходимо представить в таком виде, который позволял бы эффективно их использовать совместно с синусоидальной моделью, поэтому для определения синусоидальных параметров предлагается использовать спектральный анализ и построить обработку сигнала главным образом в частотной области.

Для имитации эффектов изменения формы спектра сигнала, проходящего через внешнее и среднее ухо, используется набор специальных фильтров: фильтр направления и фильтр предобработки.

В качестве основной модели периферической части слуховой системы человека предлагается использовать SDCM-модель – Second Order Difference Cochlea Model (разностная кохлеарная модель второго порядка). Данная модель представляет собой двумерную математическую модель улитки человеческого уха, основанной на описании параметров движения вязкой несжимаемой жидкости. Она описывает процесс движения базилярной мембраны под воздействием акустических колебаний и объясняет механизм спектрального разложения сигнала. Согласно данной модели функционирование улитки уха описывается следующим разностным уравнением:

$$y_m(n) + b_{1m}y_m(n-1) + b_{2m}y_m(n-2) = A_m a_{0m}[u_s(n) - u_s(n-2)], \quad (2)$$

где $y_m(n)$ – перемещение или так называемая пучность базилярной мембраны в позиции x_m ;

b_{1m} , b_{2m} , A_m и a_{0m} – параметры, которые определяются физическими свойствами базилярной мембраны в позиции x_m и изменяются вдоль базилярной мембраны;

m – номер сегмента базилярной мембраны после дискретизации;

$u_s(n)$ – входной синусоидальный сигнал, характеризующий скорость перемещения стремечка.

Передаточную функцию модели улитки в дискретном пространстве и времени $H_m(z)$ можно записать в виде:

$$H_m(z) = \left[A_m \frac{a_{0m}(1-z^{-2})}{1+b_{1m}z^{-1}+b_{2m}z^{-2}} \right]^{NS}, \quad (3)$$

где NS – количество секций фильтра (в соответствии с данной моделью $NS=2$).

С целью повышения степени адекватности реальным физиологическим процессам предлагается произвести коррекцию кохлеарной модели с учётом передаточной характеристики внешнего и среднего уха:

$$H_m^E(z) = H_{E1}(z) \cdot H_m(z), \quad m = \overline{1, M_F}, \quad (4)$$

где $H_m^E(z)$ – передаточная функция эквивалентного (модифицированного) фильтра;

$H_{E1}(z)$ – передаточная функция фильтра внешнего и среднего уха;

$H_m(z)$ – передаточная функция m -го кохлеарного фильтра;

M_F – количество кохлеарных фильтров в банке.

Амплитудно-частотные характеристики для 32 модифицированных кохлеарных фильтров изображены на рисунке 3.

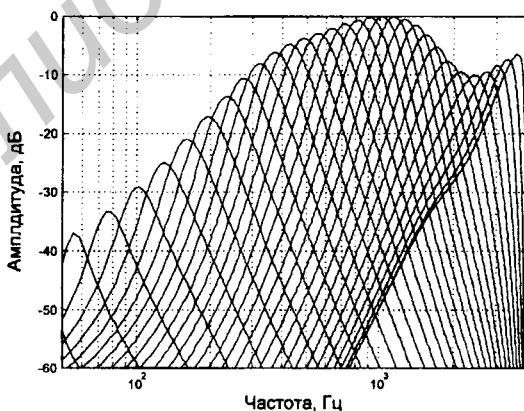


Рисунок 3 – Амплитудно-частотные характеристики для 32 модифицированных кохлеарных фильтров

В главе предложен метод реализации *комбинированной кохлеарной модели* в частотной области путём совмещения пассивной и активной кохлеарных моделей, позволяющий более корректно обрабатывать сигналы с различным уровнем.

На основе анализа существующих методов моделирования процессов в слуховом нерве человека показано, что наиболее целесообразным подходом является использование модели, основанной на вычислении групповой интервальной гистограммы (Ensemble Interval Histogram) – $G(t,f)$. Данная модель позволяет получить представление об акустической информации, циркулирующей на уровне слухового нерва. С её помощью можно дифференцировать частотные составляющие анализируемого речевого сигнала по степени их “важности” для человеческого слуха.

Однако непосредственное применение этой модели в данном случае затруднительно из-за её плохой “совместимости” с синусоидальной моделью. Кроме того, она обладает достаточно высокой вычислительной сложностью, что также усложняет обработку речевого сигнала на её основе в реальном масштабе времени.

Поэтому предлагается таким образом модифицировать процесс вычисления гистограммы $G(t,f)$, чтобы устранить вышеперечисленные недостатки и в тоже время сохранить её “полезные” свойства. Одним из возможных решений является перенос большинства действий из временной области в частотную, что позволяет не только резко снизить вычислительную сложность алгоритма анализа, но и даёт возможность корректно совместить его с применяемым для определения синусоидальных параметров спектральным анализом. Применяя обработку входного речевого сигнала в частотной области интервальную гистограмму $G(t,f)$ можно представить в виде дискретной функции частоты $G(k)$ (рисунок 4).

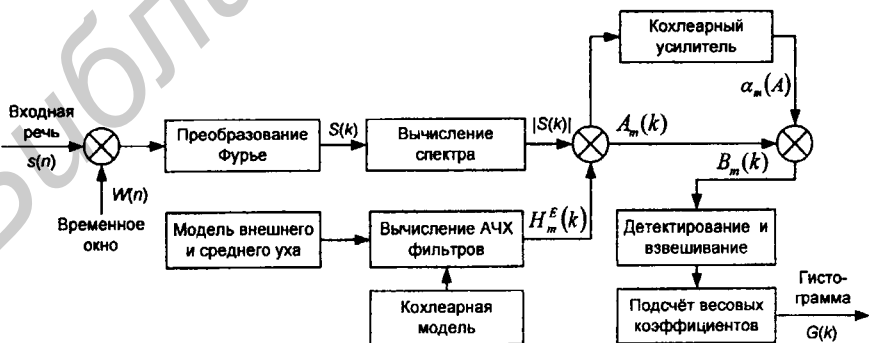


Рисунок 4 – Схема вычисления модифицированной слуховой гистограммы

В соответствии с предлагаемой моделью гистограмма $G(k)$ вычисляется с помощью следующего выражения:

$$G(k) = \sum_{m=1}^{M_F} G_m(k), \quad k = 1, \frac{N_F}{2}, \quad (5)$$

где m – номер обрабатываемого в текущий момент времени кохлеарного канала;

N_F – размер ДПФ;

$G_m(k)$ – k -й элемент гистограммы для m -го кохлеарного фильтра, он может быть вычислен на основе спектра сигнала $|S(k)|$ по следующей формуле:

$$G_m(k) = |S(k)| \cdot H_m^E(k) \cdot \alpha_m(A) \cdot \sum_{l=1}^L P_m(l), \quad (6)$$

где L – количество уровней;

$\alpha_m(A)$ – коэффициент усиления для m -го фильтра;

$P_m(l)$ – весовая функция, которая вычисляется по следующему условию:

$$P_m(l) = \begin{cases} \beta_l, & \text{если } B_m(k) \geq \text{Amp}_l, \\ 0, & \text{если } B_m(k) < \text{Amp}_l, \end{cases} \quad (7)$$

где β_l – величина, характеризующая степень нервного возбуждения;

Amp_l – амплитуда l -го уровня.

Работу активной модели (кохлеарного усилителя) при использовании для анализа речевого сигнала предлагается представить следующим образом:

$$\alpha_m(A) = \frac{100}{10^{40 \cdot \log_{10} [|S(k)| \cdot H_m^E(k)]}}. \quad (8)$$

Пример вычисления одного элемента гистограммы показан на рисунке 5. Из рисунка 5 видно, что вес выбирается в зависимости от величины произведения $|S(k)| \cdot H_m^E(k) \cdot \alpha_m(A)$.

Для повышения точности кохлеарного анализа предложено использовать ДПФ с неравномерным частотным разрешением. Это позволяет производить более детальный анализ в области низких частот, что соответствует физиологическим особенностям слуховой системы человека.

В главе описывается процедура выделения доминирующих частотных составляющих, даны соответствующие алгоритмы.

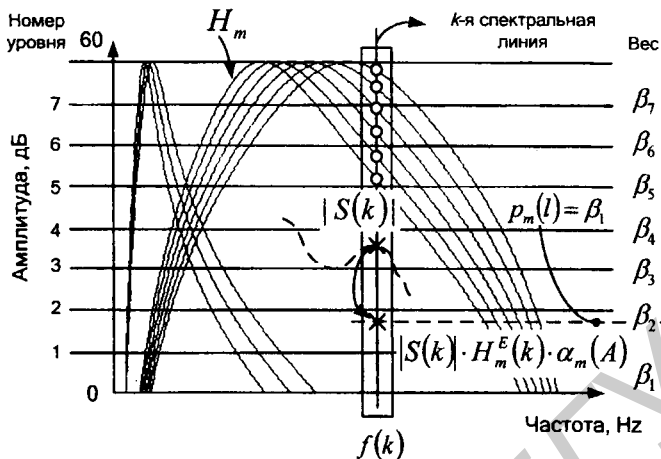


Рисунок 5 – Иллюстрация процесса вычисления k -го элемента гистограммы в частотной области

Предлагается структурная схема устройств анализа и синтеза речи на основе синусоидальной модели, совмещённой с моделями слуха человека. Данная система обладает рядом существенных преимуществ: имеет простую алгоритмическую реализацию, не требует процедур определения типа речевого фрейма и частоты основного тона в процессе анализа речи и позволяет получить речевой сигнал со спектром, который имеет большую часть необходимой для хорошего восприятия информации.

Глава 3 посвящена разработке метода квантования параметров используемой модели речевого сигнала. Ставится задача – оптимальным образом заквантовать параметры отобранных синусоидальных составляющих, т.е. данные параметры должны быть представлены минимальным количеством бит, необходимым для сохранения хорошего качества синтезированного сигнала.

Как показали проведённые эксперименты, фазы имеют равномерное распределение, а амплитуды и частоты – нет. Кроме того, наблюдается значительное отличие в распределении амплитуд и частот для разных составляющих из набора, приходящимся на один фрейм. Исходя из этого, фазы предлагается кодировать, используя скалярное квантование, а поскольку частоты и амплитуды определяются по спектральной характеристике речевого сигнала, то для их кодирования предлагается применить векторное квантование в пространстве амплитуда-частота. При таком подходе, в декодер, в качестве параметра, вместо непосредственных значений амплитуды и частоты, передаётся только индекс найденного элемента в кодовой книге.

Кодовые книги тренируются независимо для каждой синусоидальной составляющей из набора для одного фрейма. Для тренировки кодовой книги используется итерационный кластерный алгоритм *K-средних*.

Экспериментальные результаты свидетельствуют, что величина ошибки квантования зависит главным образом от длины кодовой книги и от способа её формирования. Используемый метод квантования синусоидальных параметров даёт хорошие результаты, если длина кодовой книги равняется 4096 или больше. Однако в этом случае алгоритм квантования параметров имеет высокую вычислительную сложность, что создаёт трудности для его использования в системах реального времени. С другой стороны, если кодовая книга слишком мала, то выходная синтезированная речь имеет плохое качество, появляются значительные искажения. Экспериментальным образом установлено, что именно ошибка по частоте чрезвычайно сильно влияет на качество синтезируемой речи. Поэтому предлагается скомбинировать векторное квантование амплитуд и частот со скалярным квантованием ошибки по частоте (рисунок 6).



Рисунок 6 – Схема коррекции частоты

Анализ полученных экспериментальных результатов позволяет говорить о достаточно высокой эффективности метода векторного квантования с частотной коррекцией. Использование частотной коррекции позволяет уменьшить длину кодовой книги для квантования амплитуды и частоты.

Выбраны следующие характеристики квантования синусоидальных параметров: длина кодовой книги для квантования амплитуды и частоты – 256 элементов (8 бит), для кодирования ошибки квантования частоты используется 4 бита и для квантования фазы – 3 бита. Таким образом, для представления параметров одной синусоидальной компоненты требуется 15 бит. Если анализ-синтез речевого сигнала происходит с частотой 50 раз в секунду и количество синусоидальных компонент – от 5 до 10 штук, тогда использование описанного метода квантования параметров позволяет достичь скорости передачи данных от 3750 до 7500 бит/с.

В **4-й главе** дано описание структуры и функционирования устройства компрессии. Обсуждаются проблемы и особенности аппаратной реализации данной системы. Даны и детально описаны алгоритмы компрессии и декомпрессии речевого сигнала. Предъявляются требования к объёму и структуре системной памяти, даётся оценка вычислительной сложности применяемых алгоритмов.

В данном случае предполагается, что система компрессии может непосредственно использоваться в составе персонального компьютера как дополнительное мультимедиа 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).

На основе анализа вычислительной сложности применяемых 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.

С целью эффективного использования особенностей VLIW-архитектуры разработан и адаптирован для реализации на базе процессора TMS320C6713 модифицированный метод расчёта слуховой гистограммы с возможностью распараллеливания вычислений. Это позволяет повысить эффективность вычисления слуховой гистограммы почти в два раза.

Аппаратная реализация данного устройства в реальном масштабе времени с учётом распараллеливания вычислений требует не более 45% вычислительных ресурсов процессора. Это позволяет реализовать различные дополнительные алгоритмы, которые используются в мультимедиа системах обработки речи или реализовать многоканальную систему с двумя устройствами компрессии (два кодера и два декодера).

Глава 5 является экспериментальной частью диссертационной работы. В главе проанализированы существующие методики субъективной и объективной оценки реконструированного сигнала и проведены соответствующие исследования для оценки качества восстановленной в декодере речи.

В ходе НИР «Разработать процедуры сжатия речевой информации, обеспечивающие коммерческое качество восстановленной речи», которая проводилась в рамках Государственной научно-технической программы «Развитие методов и средств системы комплексной защиты информации» с помощью специально разработанной в среде MATLAB программы были

получены объективные оценки качества восстановленного речевого сигнала для данного устройства.

Полученные оценки описываются отношениями сигнал/шум (*SNR*), шум/порог маскирования (*NMR*) и параметрами искажения спектра барков, вычисленными по стандартному (*BSD*) и модифицированному (*MBSD*) алгоритмам.

Были получены следующие результаты:

- 1) отношение сигнал/шум (*SNR*): 8.36 дБ;
- 2) отношение шум/порог маскирования (*NMR*): -3.47 дБ;
- 3) оценка искажений спектра барков (*BSD*): 0.09;
- 4) модифицированная оценка искажений спектра барков (*MBSD*): 0.07.

Значения полученных объективных оценок являются типичными для низкоскоростных параметрических вокодерных устройств.

При тестировании устройств компрессии с помощью субъективных методик под качеством восстановленной речи понимается показатель, характеризующий общее впечатление слушателя от принимаемой речи. Качество речи определяется как параметр, представляющий субъективную оценку звучания речи в испытуемой 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 8000 бит/с):

- 1) слоговая разборчивость речи – 96 %;
- 2) субъективное качество речи – 3.8 баллов;
- 3) особенности восстановленной в декодере речи – некоторое нарушение естественности и узнаваемости, иногда присутствуют подзванивание и дребезжание;
- 4) узнаваемость голоса диктора – 95 %.

Результаты субъективной оценки качества синтезированной речи подтверждают, что речь отличается довольно высокой степенью разборчивости и хорошей узнаваемостью диктора даже при ограниченном числе синусоидальных компонент.

В приложениях диссертации даны подробное описание используемой кохлеарной модели, алгоритм расчёта банка кохлеарных фильтров и приведены акты внедрения результатов диссертационной работы.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

В ходе представленной диссертационной работы были получены следующие научные результаты.

1. На основании анализа существующих подходов синусоидального кодирования речевых сигналов применительно к системам компрессии было показано, что для достижения высокой степени компрессии требуется введение антропоморфического критерия для анализа входного сигнала [1, 2]. В ходе анализа практически доказана принципиальная возможность реализации низкоскоростной системы компрессии речевого сигнала на основе синусоидальной модели [3, 11].

2. Разработан новый метод антропоморфической обработки речевого сигнала на основе трёх модифицированных моделей слуха человека: модели внешнего и среднего уха, комбинированной кохлеарной модели [5, 10] и модели слухового нерва [4, 9, 14], позволяющий производить отбор доминирующих частотных составляющих сигнала и обладающий на порядки меньшей вычислительной сложностью по сравнению с подобными алгоритмами [5]. Применение моделей слуха человека позволяет получить синтезированный речевой сигнал со спектром, который имеет большую часть информации, необходимой для хорошего восприятия речи. Данный метод был адаптирован для использования в задачах компрессии речевого сигнала с использованием синусоидальной модели [11].

3. Предложен метод реализации комбинированной кохлеарной модели в частотной области путём совмещения пассивной и активной моделей, позволяющий более корректно обрабатывать речевые сигналы с различным уровнем. Основным достоинством комбинированной кохлеарной модели является её высокая степень адекватности реальным физиологическим процессам, происходящим в улитке уха человека. Для повышения точности кохлеарного анализа предложено использовать ДПФ с неравномерным частотным разрешением. Это позволяет производить более детальный анализ в области низких частот, что соответствует физиологическим особенностям слуховой системы человека [7].

4. Разработан метод квантования синусоидальных параметров модели, учитывающий особенности их общего статистического распределения и положения внутри набора параметров для каждого фрейма [12, 15]. Использование метода векторного квантования с частотной коррекцией позволило не только снизить вычислительную сложность алгоритма квантования, но и уменьшить объём памяти, необходимый для хранения кодовых книг. Метод тренировки кодовых книг с использованием алгоритма K -

средних адаптирован для квантования небольшого набора разнородных синусоидальных параметров.

5. Создано алгоритмическое обеспечение синусоидального вокодера с антропоморфической обработкой, особенностью которого является отсутствие процедур определения частоты основного тона и классификации фреймов, а при восстановлении речевого сигнала в декодере – использование небольшого числа синусоидальных компонент [11].

6. Осуществлена аппаратная реализация устройства компрессионно-декомпрессионного речевого сигнала [6, 13]. При использовании векторного квантования параметров и числе синусоидальных компонент не более 8 скорость потока данных составляет около 5 кбит/с [12]. Экспериментальные результаты показывают, что речь отличается довольно высокой степенью разборчивости и хорошей узнаваемостью голоса диктора [8].

Рекомендации по практическому использованию результатов

Концепция совмещения синусоидального представления речевого сигнала со слуховыми моделями с целью отбора наиболее важной для слуха человека информации позволяет получить восстановленный речевой сигнал с сохранением основных характеристических особенностей речи при ограниченном числе синусоидальных компонент. Это свидетельствует о перспективности дальнейших исследований в области синусоидального кодирования речи с использованием антропоморфического анализа.

Перспектива дальнейшего развития данной модели заключаться в поиске способов улучшения качества восстановленной речи при ограничении количества отбираемых “критичных” для слуха человека частотных составляющих до 8 штук. Одним из достоинств применяемого антропоморфического подхода является его гибкость и принципиальная возможность подстраиваться под индивидуальные характеристики слуха аудитора, что можно использовать для повышения качества персонализированного восприятия синтезированной речи.

Предполагается, что разработанная система компрессии может использоваться в составе персонального компьютера как дополнительное мультимедиа 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).

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в научных журналах и сборниках научных трудов

1 Гармонический вокодер на базе кохлеарной модели анализа речи и психоакустического критерия / Д.С. Лихачёв, А. Савицки, А.А. Петровский // Известия Белорусской инженерной академии. – 2001. – № 1(11)/2. – С. 38–40.

2 Лихачёв, Д.С. Использование модели слуха человека для высококачественного низкоскоростного кодирования речи / Д.С. Лихачёв, А.А. Петровский // Известия Белорусской инженерной академии. – 2002. – № 2(14)/1. – С. 159–162.

3 Лихачёв, Д.С. Низкоскоростной кодер речи на основе синусоидальной модели речеобразования и модели слуха человека / Д.С. Лихачёв, А.А. Петровский // Известия Белорусской инженерной академии – 2002. – № 2(14)/1. – С. 163–166.

4 Likhachov, D.S. Improved auditory-based speech coding using psychoacoustic model based on a cochlear filter bank and an average localized synchrony detection / D.S. Likhachov, A.A. Petrovsky // Computer information systems and industrial management applications / Eds.: K. Saeed, R. Mosdorf, Z. Sosnowski – Poland, Bialystok, 2003. – P. 11–19.

5 An anthropomorphic speech processing based on the cochlear model and its application for coding task / A.A. Petrovsky, D.S. Likhachov, W. Wan // International scientific journal of computing. – 2004. – Vol. 3, № 1. – P. 75–83.

6 Лихачев, Д.С. Особенности аппаратной реализации системы компрессии речевой информации на основе слуховой модели человека / Д.С. Лихачев // Известия Белорусской инженерной академии. – 2004. – №1(17)/3. – С. 122–125.

7 Лихачёв, Д.С. Антропоморфический анализ на основе дискретного преобразования Фурье с неравномерной частотной шкалой / Д.С. Лихачёв // Известия Белорусской инженерной академии. – 2005. – № 1(19)/2 – С. 177–180.

8 Лихачёв, Д.С. Анализ и синтез устройств кодирования речевого сигнала на основе антропоморфической обработки и синусоидальных моделей / Д.С. Лихачёв, А.А. Петровский // Доклады Белорусского государственного университета информатики и радиоэлектроники. – 2006. – № 3 (15). – С. 35–43.

Статьи в сборниках трудов научно-технических конференций

9 Spiking neuron auditory model for speech processing systems / A.V. Ivanov, D.S. Likhachov, A.A. Petrovsky // Systems, signals and image processing: proceedings of the 9-th intern. workshop, Manchester, UK, November 7–8, 2002. – Manchester, 2002. – P. 231–236.

10 Petrovsky, A.A. A digital cochlear model as a base of anthropomorphic speech processing / A.A. Petrovsky, D.S. Likhachov // Neural networks and artificial intelligence: proceedings of the 3-rd intern. conf., Belarus, Minsk, November 12–14, 2003. – Minsk, 2003. – P. 126–131.

11 Лихачёв, Д.С. Низкоскоростной кодер речи на основе модели слуха человека / Д.С. Лихачёв, А.А. Петровский // Цифровая обработка сигналов и ее применение: труды 5-й междунар. конф., Россия, Москва, 12–17 марта 2003 г. – Москва, 2003. – Т. 2. – С. 379–382.

12 Likhachov, D.S. Parameters quantization in sinusoidal speech coder on basis of human auditory model / D.S. Likhachov, A.A. Petrovsky // Speech and Computer: proceedings of the 9-th intern. conf., Russia, Saint-Petersburg, September 20–22, 2004. – Saint-Petersburg, 2004. – P. 195–202.

13 Likhachov, D.S. Details of the hardware implementation of sinusoidal vocoder with anthropomorphical speech processing based on the VLIW-architecture DSP / D.S. Likhachov // Pattern recognition and information processing: proceedings of the 8-th intern. conf., Belarus, Minsk, May 18–20, 2005. – Minsk, 2005. – P. 428–431.

Тезисы докладов

14 Лихачёв, Д.С. Использование групповой интервальной гистограммы в задачах компрессии и кодирования речи / Д.С. Лихачёв // Технические средства защиты информации: тезисы докладов Российско-белорусской научно-технической конференции, Беларусь, Нарочь, 19–23 мая 2003 г. / Доклады Белорусского государственного университета информатики и радиоэлектроники. – 2003. – Т. 1, № 2/1. – С. 44.

15 Лихачёв, Д.С. Квантование параметров в синусоидальных кодерах речи, основанных на слуховой модели человека / Д.С. Лихачёв, А.А. Петровский // Технические средства защиты информации: тезисы докладов Российско-белорусской научно-технической конференции, Беларусь, Нарочь, 17–21 мая 2004 г. / Доклады Белорусского государственного университета информатики и радиоэлектроники. – 2004. – № 5. – С. 83–84.



РЭЗЬЮМЭ

Ліхачоў Дзяніс Сяргеевіч

Аналіз і сінтэз прылад кадавання маўленчага сігналу на аснове антрапамарфічнай апрацоўкі і сінусоідных мадэляў

Ключавыя словы: кампрэсія мовы, антрапамарфічная апрацоўка, кахлеарная мадэль, мадэль слыхавога нерва, дамінуючыя частотныя складнікі маўленчага сігналу.

Мэтай працы з'яўляецца распрацоўка метадаў і алгарытмаў аналізу і сінтэзу прылад кадавання мовы з высокай ступенню кампрэсіі, якія выкарыстоўваюць антрапамарфічную апрацоўку маўленчага сігналу.

Аб'ектам даследавання з'яўляецца прылада сінусоіднага кодэра мовы з антрапамарфічнай апрацоўкай маўленчага сігналу. Прадмет даследавання – метады кампрэсіі маўленчых сігналаў на аснове сінусоіднай мадэлі, якія выкарыстоўваюць антрапамарфічны аналіз для выбару найбольш значнай для ўспрымання інфармацыі.

Навуковая навізна атрыманых аўтарам вынікаў складаецца ў наступным: распрацаваны новы метады антрапамарфічнай апрацоўкі маўленчага сігналу ў частотнай вобласці на аснове трох слыхавых мадэляў: мадэлі знешняга і сярэдняга вуха, кахлеарнай мадэлі і мадэлі слыхавога нерва чалавека, які дазваляе вырабляць адбор найбольш важных для добрага ўспрымання мовы чалавекам дамінуючых складнікаў; распрацаваны метады аналізу маўленчага сігналу ў частотнай вобласці на аснове камбінаванай кахлеарнай мадэлі (актыўнай і пасіўнай) і ператварэнні Фур'е з нераўнамерным шагам па частаце, асноўнай годнасцю якога з'яўляецца высокая ступень адэкватнасці рэальным фізіялагічным працэсам, што адбываюцца ва ўлітцы вуха чалавека; распрацаваны метады вектарнага квантавання сінусоідных параметраў з частотнай карэкцыяй, які ўлічвае асаблівасці статыстычнага размеркавання кадаваных параметраў і ступень іх уплыву на якасць сінтэзаванай мовы; створана праграма-апаратнае забеспячэнне прылады кампрэсіі, прызначанага для выкарыстання ў складзе персанальнага кампутара ў якасці дадатковай мультымедыя прылады для сціску маўленчых паведамленняў з мэтай іх захоўвання ў памяці або перадачы па кампутарных сетках (Інтэрнэт-тэлефанія).

Вынікі дысертацыйнай працы ўкаранёныя і выкарыстоўваюцца ў інжынерна-тэхнічным ТАА "КВІНТЭЛ", ў Дзяржаўнай навуковай установе "АІП НАН Беларусі" (лабараторыя "Распазнання і сінтэзу мовы") і ў Дзяржаўным цэнтры бяспекі інфармацыі пры Прэзідэнце Рэспублікі Беларусь.

РЕЗЮМЕ

Лихачёв Денис Сергеевич

Анализ и синтез устройств кодирования речевого сигнала на основе антропоморфической обработки и синусоидальных моделей

Ключевые слова: компрессия речи, антропоморфическая обработка, кохлеарная модель, модель слухового нерва, доминирующие частотные составляющие речевого сигнала.

Целью работы является разработка методов и алгоритмов анализа и синтеза устройств кодирования речи с высокой степенью компрессии, использующих антропоморфическую обработку речевого сигнала.

Объектом исследования является устройство синусоидального кодера речи с антропоморфической обработкой речевого сигнала. Предмет исследования – методы компрессии речевых сигналов на основе синусоидальной модели, использующие антропоморфический анализ для выбора наиболее значимой для восприятия информации.

Научная новизна полученных автором результатов заключается в следующем: разработан новый метод антропоморфической обработки речевого сигнала в частотной области на основе трёх слуховых моделей: модели внешнего и среднего уха, кохлеарной модели и модели слухового нерва человека, позволяющий производить отбор наиболее важных для хорошего восприятия речи человеком доминирующих составляющих; разработан метод анализа речевого сигнала в частотной области на основе комбинированной кохлеарной модели (активной и пассивной) и преобразования Фурье с неравномерным частотным разрешением, основным достоинством которого является высокая степень адекватности реальным физиологическим процессам, происходящим в улитке уха человека; разработан метод векторного квантования синусоидальных параметров с частотной коррекцией, учитывающий особенности статистического распределения кодируемых параметров и степень их влияния на качество синтезированной речи; создано программно-аппаратное обеспечение устройства компрессии, предназначенного для использования в составе персонального компьютера в качестве дополнительного мультимедиа устройства для сжатия речевых сообщений с целью их хранения в памяти или передачи по компьютерным сетям (Интернет-телефония).

Основные результаты диссертационной работы внедрены и используются в инженерно-техническом ООО “КВИНТЕЛ”, Государственном научном учреждении “ОИПИ НАН Беларуси” (лаборатория “Распознавания и синтеза речи”) и в Государственном центре безопасности информации при Президенте Республики Беларусь.

SUMMARY

of the dissertation by Likhachov Denis Sergeevich

The analysis and synthesis of speech coding devices on the basis of anthropomorphical processing and sinusoidal models

Keywords: speech compression, anthropomorphical processing, cochlear model, model of the auditory nerve, dominating frequency components of a speech signal.

The purpose of the work is a development of methods and analysis algorithms and synthesis of speech compression devices on the basis of sinusoidal model coupled with anthropomorphical procedure for speech analysis.

The object of the research is a sinusoidal speech coder with anthropomorphical processing of a speech signal. The subjects of the research are methods of speech compression based on sinusoidal model that use anthropomorphical analysis for selecting the most significant perceptual information.

The scientific novelty of the results received by the author consists of the following issues: new method of anthropomorphical speech processing in the frequency domain is developed on the basis of three auditory models – model of an external and inner ear, cochlear model and model of human auditory nerve, which allows making the selection of the most important for good speech perception dominating components; the method of speech analysis in the frequency domain is developed on the basis of combined cochlear model (active and passive) and non-uniform Fourier transform which basic advantage is the high degree of adequacy to the real physiological processes occurring in human cochlea; the method for vector quantization of sinusoidal parameters with the frequency correction is developed, features of statistical distribution of coded parameters and a degree of their influence on quality of the synthesized speech are considered; hardware-software implementation of the device is designed for use in personal computer as an additional multimedia device for compression of speech messages with the purpose of storing in the memory or transferring through computer networks (Internet-telephony).

Results of the dissertational work are implemented into practice and used in: "KVINTEL" company, in the State scientific organization "United Institute of Informatics Problems of the National Academy of Sciences of Belarus" (laboratory "Recognition and synthesis of speech") and in the State Center of Information Security at the President of the Republic of Belarus.

ЛИХАЧЁВ
Денис Сергеевич

**АНАЛИЗ И СИНТЕЗ УСТРОЙСТВ КОДИРОВАНИЯ РЕЧЕВОГО
СИГНАЛА НА ОСНОВЕ АНТРОПОМОРФИЧЕСКОЙ ОБРАБОТКИ И
СИНУСОИДАЛЬНЫХ МОДЕЛЕЙ**

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

специальность 05.13.05

«Элементы и устройства вычислительной техники и систем управления»

Подписано в печать	13.11.2006.	Формат 60x84 1/16.	Бумага офсетная.
Гарнитура «Таймс».	Печать ризографическая.		Усл. печ. л. 1,63.
Уч.-изд. л. 1,4.	Тираж 60 экз.		Заказ 672.

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. Бровка, 6.