

Учреждение образования
“БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ”

УДК 621.371.39:681.322.01

АНОШЕНКО
Алексей Евгеньевич

**Психоакустически мотивированные методы и устройства повышения
качества речевого сигнала в мультимедиа системах**

05.13.05 – Элементы и устройства вычислительной техники и систем управления

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Минск 2005

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

Научный руководитель – д.т.н., проф. Петровский А.А
(Учреждение образования “Белорусский государственный университет информатики и радиоэлектроники”, кафедра ЭВС)

Официальные оппоненты: д.т.н., проф. Ярмолик В.Н.
(Учреждение образования “Белорусский государственный университет информатики и радиоэлектроники”, кафедра ПОИТ)

к.т.н., с.н.с. Левковская Т.В.
(Объединенный институт проблем информатики НАНБ)

Оппонирующая организация – Проектное и научно-исследовательское
Республиканское унитарное предприятие
“Гипросвязь”

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы диссертации

Проблема повышения качества речевого сигнала: компенсация эхо-сигнала и редактирование шумов окружающей среды, многие годы вызывает интерес исследователей в области проектирования мультимедиа систем. Это объясняется наличием большого числа возможных приложений таких систем, работающих в определенной акустической обстановке: громкоговорящая связь (телеконференции и телефоны, установленные в автомобиле), верификация и идентификация диктора, распознавание речи, низкоскоростное кодирование речи. В данных системах микрофоны располагаются на достаточном удалении от ближнего диктора, а громкоговорители размещаются произвольно, например, на приборной панели автомобиля. В таком случае сигнал, снимаемый с микрофона, кроме сигнала ближнего диктора содержит шумы окружающей среды и акустическое эхо, обусловленное сигналом громкоговорителя (дальнего диктора), которые значительно ухудшают восприятие речи дальним диктором.

Развитие работ по данной задаче в настоящее время идет по пути построения на одном кристалле процессора цифровой обработки сигналов одноканальных, комбинированных систем, включающих подсистему компенсации эхо-сигнала на основе оптимального Винеровского фильтра, адаптация которого осуществляется по методу нормализованных наименьших квадратов (НМНК), и подсистему редактирования шума согласно методу спектрального вычитания или его модификаций. Однако методы редактирования шума, основанные на правилах спектрального взвешивания (вычитания), приводят к возникновению артефактов в реконструированном речевом сигнале в виде так называемых «музыкальных тонов». С другой стороны, длительность акустического эха может в некоторых случаях достигать нескольких сотен миллисекунд, компенсация которого требует использования Винеровского фильтра очень высокого порядка, что обуславливает значительно более высокую вычислительную сложность по сравнению с алгоритмами редактирования шума. Обработка сигнала в частотной области позволяет снизить вычислительную сложность алгоритмов. Использование же принципов психоакустики в комбинированных системах повышения качества речевого сигнала приводит к устранению как возмущения, обусловленного эхо-сигналом и шумом, так и искажения, вносимого в результат работы самой системой.

Вышесказанное подчёркивает актуальность исследований для разработки психоакустически мотивированных устройств и методов повышения качества речевых сигналов в мультимедиа системах. Важность проблемы, совершенствование методов и алгоритмов обработки речевых сигналов в реальном времени и их ото-

бражения на архитектуру микропроцессора определили цель работы, решаемые задачи и методы исследований.

Связь работы с крупными научными программами, темами

Диссертационная работа выполнена в соответствии с научно-техническими заданиями и планами работ кафедры «Электронных вычислительных средств» и НИЛ 3.1 «Микропроцессорные системы реального времени» Белорусского государственного университета информатики и радиоэлектроники в рамках фундаментальных исследований Министерства образования Республики Беларусь: ГБЦ 99-3008 «Разработка методов подавления эхо-сигналов и редактирования шумов на основе комбинированной время-частотной обработки: глобальная оптимизация», а также при финансовой поддержке Фонда фундаментальных исследований Республики Беларусь Т99-308 «Новый подход к низкоскоростному кодированию речевых сигналов в условиях запумления с учетом особенностей слуха человека», хозяйственного договора х/л 01-1089 «Разработать процедуры сжатия речевой информации, обеспечивающие коммерческое качество восстановленной речи», выполняемого в соответствии с Государственной научно-технической программой «Развитие методов и средств систем комплексной защиты информации».

Цель и задачи исследования

Целью работы является разработка методов и устройств подавления акустического эхо-сигнала и снижения шумов окружающей среды в речевом сигнале мультимедиа систем реального времени, базирующихся на обработке в частотной области и психоакустических особенностях восприятия человеком акустической информации.

Поставленная цель работы определяет следующие задачи исследования:

1. Анализ психоакустически мотивированных подходов построения комбинированных систем компенсации эхо-сигнала и редактирования шумов окружающей среды.
2. Разработка метода подавления эхо-сигнала и редактирования шумов окружающей среды, учитывающего психоакустические особенности восприятия человеком акустической информации.
3. Разработка метода оценки спектральной плотности мощности шумов окружающей среды для одномикрофонной системы повышения качества речевого сигнала.
4. Разработка метода оценки спектральной плотности мощности эхо-сигнала на основе сигналов микрофона и дальнего диктора.
5. Разработка алгоритма детектора наличия речевой активности.

6. Аппаратно-программная реализация перцептуального метода повышения качества частотно-ограниченного речевого сигнала: одноканальный вариант, комбинированная система, частотный подход, реальный масштаб времени.

Объект и предмет исследования

Объектом исследования являются устройства компенсации эхо-сигнала и редактирования шумов окружающей среды мультимедиа систем. Предмет исследования – психоакустически мотивированные методы повышения качества речевого сигнала в устройствах мультимедиа.

Гипотеза

В диссертации выдвигается гипотеза, что подавление акустического эхо-сигнала и редактирование шумов окружающей среды в устройствах мультимедиа с использованием перцептуально-мотивированных подходов обработки сигнала в частотной области может привести к повышению качества выходного речевого сигнала при одновременном значительном снижении вычислительной сложности алгоритмов обработки.

Методология и методы проведенного исследования

Совокупность использованных в диссертационной работе средств исследований базируется на аппарате линейной алгебры, математического анализа, теории дифференциальных уравнений, численных математических методов, теории вероятностей и математической статистики, методах проектирования проблемно-ориентированных вычислительных средств. Экспериментальная проверка основных положений и результатов работы проводилась при помощи компьютерного моделирования.

Научная новизна и значимость полученных результатов

Новые научные результаты диссертационной работы следующие:

1. На основании анализа подходов построения перцептуально мотивированных методов и устройств повышения качества речевого сигнала в мультимедиа системах показано, что вариант одноканальной комбинированной системы компенсации эхо-сигнала и редактирования шума на основе обработки в частотной области с учетом психоакустических особенностей восприятия человеком акустической информации позволяет значительно снизить вычислительную сложность алгоритмов, осуществлять параллельно-поточную обработку на однокристальном VLIW процессоре с поддержкой SIMD инструкций в реальном масштабе времени.
2. Психоакустически мотивированное правило спектрального взвешивания, которое является модификацией метода спектрального вычитания, но в отличие от последнего, позволяет организовать одновременное подавление акустического

эхо-сигнала и шумов окружающей среды. Учет при разработке правила психоакустического эффекта маскирования позволяет значительно снизить искажения вносимые методом, проявляющиеся в виде эффекта “музыкального тона”.

3. Новая структура комбинированной системы повышения качества речевого сигнала мультимедиа систем на основе единого перцептуального правила спектрального вычитания в частотной области, которая характеризуется по сравнению с известными компромиссом между вычислительной сложностью алгоритмов, малой задержкой трансмиссии и высоким перцептуальным качеством реконструированного речевого сигнала.

4. Метод оценки спектральной плотности мощности (СПМ) акустического эхо-сигнала по оценкам СПМ сигнала дальнего диктора, отличительной особенностью которого является то, что используется блочная обработка входного сигнала и корректировка весовых коэффициентов фильтра-предсказателя эхо-сигнала на основе нормализованного метода наискорейшего спуска осуществляется с использованием значений вероятности речевой активности детектора речевых пауз и двойного разговора, что позволяет повысить скорость сходимости фильтра-предсказателя в переходные моменты между периодом тишины и речевой активностью, а также степень подавления эхо-сигнала большой длительности.

5. Модифицированный метод оценки СПМ шумов окружающей среды с учетом вероятностных оценок речевой активности детектора речевых пауз и двойного разговора, что в отличие от классических подходов позволяет повысить достоверность измерений в моменты перехода от начала речевой активности к паузе и наоборот.

Практическая значимость полученных результатов

Полученные результаты формируют теоретическую и практическую базу для синтеза и анализа методов и устройств повышения качества речевого сигнала на основе перцептуального подхода обработки сигналов, функционирующих в реальном времени как для стационарных, так и мобильных мультимедиа систем. При этом практически реализованы следующие устройства и соответствующее инструментальное программное обеспечение:

Осуществлены аппаратные реализации устройств повышения качества речевого сигнала в реальном времени на базе ЦПОС TMS320C541 и аппаратного модуля TMS320C6711 DSP Starter Kit. Устройства обеспечивают подавления эхо-сигнала большой длительности (до 500мс), при этом уровень ослабления возвращаемого в канал эхо-сигнала равен 35-38 дБ, а уровень подавления шума составляет 15-20 дБ и реконструируемый речевой сигнал не имеет артефактов, что соответствует лучшим разработкам подобных устройств.

Предлагаемые методы и алгоритмы могут быть использованы при создании как самостоятельных устройств компенсации эхо-сигнала и редактирования шума, так и для построения препроцессоров повышения качества входного сигнала различных мультимедиа систем. Так в системах распознавания и компрессии речевых сигналов данные препроцессоры позволят значительно повысить их производительность, что будет способствовать скорейшему внедрению речевых интерфейсов в различные системы управления и связи.

Базовые результаты диссертационной работы переданы для внедрения в конструкторское бюро ОАО "Минский приборостроительный завод", а также, используются в учебном процессе на кафедре Электронных вычислительных средств Белорусского государственного университета информатики и радиоэлектроники в дисциплинах "Теория и применение ЦОС" и "Речевые интерфейсы ЭВС".

Основные положения диссертации, выносимые на защиту

На защиту выносятся следующие научные положения:

1. Психоакустически мотивированный метод повышения качества речевого сигнала в частотной области, характеризующийся по сравнению с известными на порядок меньшей вычислительной сложностью и единым для редактирования шума и эхо-сигнала правилом спектрального взвешивания.
2. Метод оценки СПМ акустического эхо-сигнала по оценкам СПМ сигнала дальнего диктора на основе блочной обработки входного сигнала и с учетом вероятности речевой активности детектора речевых пауз и двойного разговора при корректировке весовых коэффициентов фильтра-предсказателя эхо-сигнала.
3. Модифицированный метод оценки СПМ шумов окружающей среды с использованием вероятностных оценок речевой активности детектора речевых пауз и двойного разговора.
4. Структура комбинированной системы повышения качества речевого сигнала мультимедиа систем на основе единого психоакустически мотивированного правила спектрального вычитания в частотной области, характеризующаяся компромиссным решением между вычислительной сложностью алгоритмов, малой задержкой трансмиссии и высоким перцептуальным качеством реконструированного речевого сигнала.
5. Аппаратно-программная реализация перцептуальных одноканальных устройств повышения качества речевого сигнала в мультимедиа системах с обработкой в частотной области на основе параллельно-поточной обработки на однокристалльном VLIW процессоре с поддержкой SIMD инструкций в реальном масштабе времени.

Личный вклад соискателя

Все вошедшие в диссертацию результаты получены при непосредственном личном участии автора. Все предлагаемые методы разработаны в соавторстве с научным руководителем и программно реализованы лично автором. В публикациях с соавторами вклад соискателя определяется рамками излагаемых в диссертации результатов. Научный руководитель принимал участие в постановке задачи и определении возможных путей их решения.

Апробация результатов диссертации

Результаты исследований, включенные в диссертацию докладывались и обсуждались на III Международной научно-технической конференции “Современные средства связи” (Нарочь, Беларусь, 1998); Международной научно-технической конференции “Kommunikationstechnik-2000” (Mittweida, Germany, 1998); I Международной научно-технической конференции “Техника и технология связи”: (Минск, Беларусь, 1999); IV Международной научно-технической конференции “Современные средства связи” (Нарочь, Беларусь, 1999); 2-й Международной конференции “Цифровая обработка сигналов и ее применение” (Москва, Россия, 1999); 3-й Международной конференции “Цифровая обработка сигналов и ее применение” (Москва, Россия, 2000); 4-й Международной конференции “Цифровая обработка сигналов и ее применение” (Москва, Россия, 2002); 5-й Международной конференции “Цифровая обработка сигналов и ее применение” (Москва, Россия, 2003).

Опубликованность результатов

По материалам диссертационной работы опубликовано 10 печатных работ, включая 5 статей в научно-технических журналах и научных сборниках и 5 статей в материалах международных конференций. Суммарный объем публикаций составляет примерно 62 печатных страницы.

Структура и объем диссертации

Диссертационная работа состоит из общей характеристики работы, пяти глав, заключения, списка используемых источников и приложений. Она содержит 100 страниц основного текста; 49 рисунков и 8 таблиц, расположенных в тексте диссертации; 2 приложения на 15 страницах; в списке использованных источников на 10 страницах представлено 110 наименований.

СОДЕРЖАНИЕ РАБОТЫ

В главе 1 осуществляется анализ методов и устройств подавления эхо-сигнала и шума в мультимедиа системах. Типовая схема одноканальной комбинированной системы компенсации акустического эха и редактирования шумов окружающей среды состоит из модуля адаптивной компенсации эхо-сигнала (АКЭС) и модуля редактирования шума (РШ) (рис. 1).

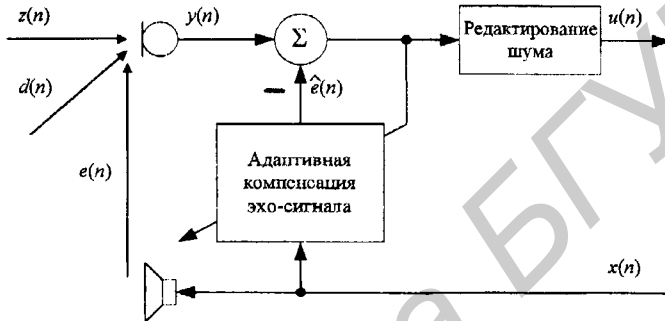


Рис. 1. Схема одноканальной комбинированной системы подавления эхо-сигнала и редактирования шумов окружающей среды

Сигнал микрофона $y(n)$ в заданной системе состоит из следующих компонентов

$$y(n) = z(n) + d(n) + e(n),$$

где $z(n)$ – речевой сигнал ближнего диктора; $d(n)$ – шумы окружающей среды; $e(n)$ – акустический эхо-сигнал, возникающий между громкоговорителем и микрофоном в результате многократного отражения и затухания сигнала дальнего диктора $x(n)$. Путь прохождения акустического эха можно представить как линейный канал, характеризующийся импульсной характеристикой $h(n)$. Следовательно, эхо-сигнал есть результат свертки сигнала $x(n)$ с $h(n)$: $e(n) = x(n) * h(n)$.

При построении одноканальной системы компенсации эхо-сигнала и редактирования шума принимаются следующие допущения: шум окружающей среды $d(n)$ независим от речевого сигнала $z(n)$, более того, сигнал ближнего диктора $z(n)$ и эхо-сигнал $e(n)$ так же независимы; речевые сигналы ближнего $z(n)$ и дальнего $x(n)$ дикторов, а также шумы окружающей среды $d(n)$ локально стационарны на некотором отрезке времени, что позволяет теоретически применить концепцию оптимальных фильтров для того, чтобы определиться с комбинированной системой редактирования шума и эхо-сигнала; используется свойство локальной стационарности речевого сигнала на отрезках времени от 20 до 60 мс.

Устройства АКЭС моделируют акустическую характеристику помещения $h(n)$, $n = \overline{0, N-1}$, N – длина характеристики, в котором используется громкоговорящая система связи. На базе полученной модели пути распространения эха и сигнала дальнего диктора $x(n)$ формируется предсказанный эхо-сигнал $\hat{e}(n)$, который затем вычитается из сигнала микрофона $y(n)$. Реализация подсистемы АКЭС на основе адаптивного нормализованного метода наименьших квадратов с конечной импульсной характеристикой (НМНК КИХ) обеспечивает хорошие свойства стабильности системы и сходимости к оптимальному решению. Однако алгоритм НМНК КИХ фильтра обладает высокой вычислительной сложностью (пропорциональна $\sim N^2$), что усложняет реализацию устройств компенсации эхо-сигнала большой длительности. Использование в подсистеме АКЭС реализации НМНК КИХ фильтра в частотной области, как показано в работе, позволяет значительно снизить вычислительную сложность алгоритмического обеспечения (пропорциональна $\sim N \cdot \log_2 N$) при достижении требуемого уровня ослабления эхо-сигнала.

При обработке m -го фрейма речевого сигнала $u(m, n)$ снизить уровень шума окружающей среды можно на основе спектрального вычитания спектральной плотности мощности (СПМ) сигнала микрофона $|Y(m, \omega)|^2$ и оценки СПМ шума $|\hat{D}(m, \omega)|^2$:

$$|U(m, \omega)|^2 = |Y(m, \omega)|^2 - |\hat{D}(m, \omega)|^2.$$

С другой стороны, результат спектрального вычитания $|U(m, \omega)|^2$ (СПМ выходного сигнала $u(m, n)$) рассматривается как продукт фильтрации некого, изменяемого от фрейма к фрейму, линейного фильтра $|G(m, \omega)|$:

$$|U(m, \omega)| = |G(m, \omega)| \cdot |Y(m, \omega)|.$$

Легко доказывается, что

$$|G(m, \omega)| = \sqrt{1 - |\hat{D}(m, \omega)|^2 / |Y(m, \omega)|^2}.$$

а с учетом отношения сигнал/шум $SNR(m, \omega) = |Y(m, \omega)|^2 / |\hat{D}(m, \omega)|^2$ весовая функция $|G(m, \omega)|$ определяется как

$$|G(m, \omega)| = \sqrt{1 - 1/SNR(m, \omega)}.$$

Таким образом, подсистема РШ обычно реализуется в виде спектрального взвешивания (рис. 2):

$$u(m, n) = \text{ОДПФ}(|Y(m, \omega)| \cdot G(m, \omega) \cdot e^{j \arg Y(m, \omega)}),$$

где ОДПФ – функция обратного дискретного преобразования Фурье.

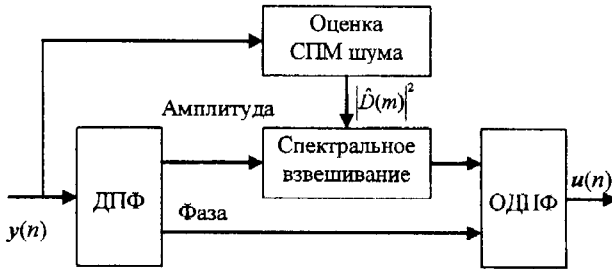


Рис. 2. Блок-схема алгоритма спектрального взвешивания

Метод спектрального вычитания не требует наличия отдельного входного шумового канала и обладает достаточно низкой вычислительной сложностью. Однако он имеет существенный недостаток, который заключается в появлении в выходном сигнале $u(n)$, так называемого, “музыкального тона” в виде неестественных “железных ноток” и слабого бульканья.

Психоакустически мотивированные модификации метода спектрального взвешивания, учитывающие психоакустические особенности восприятия человека, позволяют значительно снизить искажения данного вида. Отличие от классического алгоритма заключается в способе вычисления весовой функции $G(m, \omega)$. При этом, следует отметить, что оценка порога маскирования в зашумленной акустической обстановке более точна, чем $SNR(m, \omega)$; порог маскирования эволюционирует медленнее, чем $SNR(m, \omega)$; адаптация алгоритмов на базе порога маскирования намного лучше коррелируется с перцептуальным восприятием человеком акустической информации, чем SNR -ориентированные алгоритмы.

В главе выполняется анализ аппаратных платформ реализации устройства, среди которых выделяются следующие: специализированные микросхемы, специализированные цифровые процессоры обработки сигналов и программируемые логические матрицы. На специализированных микросхемах, как правило, реализуется только определенный алгоритм с ограниченными возможностями регулировки, что затрудняет использование их в широком классе мультимедиа систем. На основе выполненного анализа архитектур цифровых процессоров обработки сигналов делается вывод, что при аппаратной реализации устройств компенсации эхо-сигнала и редактирования шумов, оптимальным подходом, с точки зрения производительности системы, возможности распараллеливания вычислений и стоимости решения, будет использование VLIW процессоров с поддержкой SIMD инструкций.

В главе 2 выполняется разработка единого психоакустически мотивированного правила компенсации эхо-сигнала и редактирования шумов окружающей среды. При построении систем подавления эхо-сигнала и редактирования шумов

предлагается основываться на следующем подходе: сигнал шума $d(n)$ и эхо сигнал $e(n)$ рассматриваются как некий сигнал помехи $p(n) = d(n) + e(n)$, искажающий речевой сигнал $z(n)$ ближнего диктора, и для его ослабления применяется техника редактирования шума (пост-фильтрация $G(m, \omega)$) (рис. 3). Использование модуля первичного РШ позволяет сократить шумовое влияние на подсистему АКЭС (фильтр $W_L(m, \omega)$, где L – число отсчетов импульсной характеристики фильтра) и, как следствие, повысить степень подавления эхо-сигнала. Весовую функцию $G(m, \omega)$ единого правила спектрального взвешивания в одноканальной системе редактирования шумов и компенсации эхо-сигнала предлагается конструировать следующим образом:

$$G(m, \omega) = \sqrt{1 - \frac{1}{\text{SNR}(m, \omega)} - \frac{1}{\text{SER}(m, \omega)}}$$

где $\text{SER}(m, \omega)$ -- отношение сигнал/эхо-сигнал.

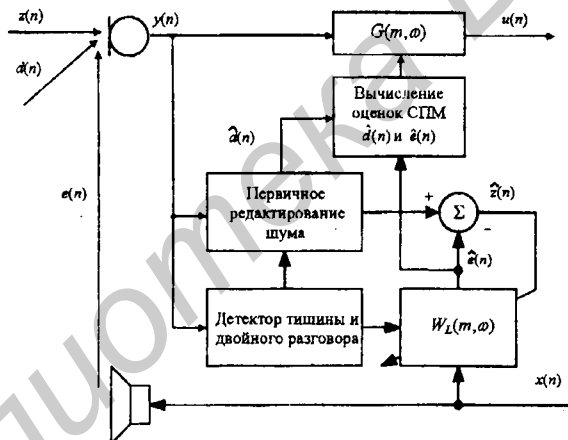


Рис. 3. Комбинированная система повышения качества речевого сигнала с постфильтром $G(m, \omega)$

Комбинация АКЭС $W_L(m, \omega)$ с постфильтром $G(m, \omega)$ не требует идентифицировать достаточно точно канал прохождения акустического эха (импульсную характеристику $h(n)$, $n = 0, N-1$), $L < N$ и позволяет получить показатели качества сравнимые со стандартной системой (АКЭС с импульсной характеристикой размера N), что приводит к снижению вычислительной сложности системы.

В разделе показано, что в результате использования классического метода спектрального вычитания в выходном сигнале присутствует ошибка $\varepsilon(\omega)$ зависящая от разности фаз сигналов $z(n)$ $d(n)$ и $e(n)$. Если $\varepsilon(\omega) > 0$ для всех ω , то ослаб-

ленная шумовая компонента присутствует в выходном сигнале $u(n)$, а речевая компонента $z(n)$ не меняется. Однако если $\varepsilon(\omega) < 0$ хотя бы для некоторых частот, то одновременно с подавлением шумовой компоненты осуществляется также на данных частотах изменение речевой компоненты $z(n)$, т.е. происходит как бы “перерегулирование” системы РПП, которое ощущается в выходном сигнале $u(n)$ в виде так называемого “музыкального тона”. Следовательно, для повышения качества выходного сигнала в метод спектрального вычитания необходимо ввести механизм компенсации отрицательной ошибки. Для этого предлагается использовать избирательное подавление шумов окружающей среды и эхо-сигнала. Главная идея данного подхода заключается в том, что в выходном сигнале $u(n)$ искажающие составляющие $d(n)$ и $e(n)$ удаляются не полностью, а оставляется некоторый комфортный, небольшой по уровню шум и остаточное эхо. Для определения данного уровня предлагается использовать порог маскирования T_M . В главе разработано новое правило психоакустически мотивированного спектрального вычитания, единое для компенсации эхо-сигнала и редактирования шумов окружающей среды:

$$G(m, \omega) = \begin{cases} 1, & |\hat{D}(m, \omega)|^2 + |\hat{E}(m, \omega)|^2 \leq T_M(m, \omega), \\ \sqrt{1 - \frac{1}{SNR(m, \omega)} - \frac{1}{SER(m, \omega)} + \frac{\alpha}{SMR(m, \omega)}}, & |\hat{D}(m, \omega)|^2 + |\hat{E}(m, \omega)|^2 > T_M(m, \omega), \end{cases}$$

где $SMR(m, \omega)$ – отношение сигнал/порог маскирования, коэффициент $\alpha < 1$ определяет, на сколько остаточный шум и эхо-сигнал должен быть ниже порога маскирования.

В разделе предложена новая структура комбинированной системы компенсации эхо-сигнала и редактирования шума на основе единого перцептуального правила спектрального взвешивания в частотной области (рис. 4), характеризующегося компромиссом между вычислительной сложностью, малой задержкой трансмиссии и высоким перцептуальным качеством реконструированного речевого сигнала $u(n)$.

В главе 3 выполняется разработка методов измерения спектральных плотностей мощности эхо-сигнала и шумов окружающей среды с учетом вероятностных оценок детектора речевой активности. Предлагается алгоритм оценки вероятностей наличия речевой активности в сигналах микрофона $y(n)$ и дальнего диктора $x(n)$, основанный на анализе средней и текущей мощности сигналов микрофона $y(n)$, дальнего диктора $x(n)$ и реконструированного речевого сигнала $\hat{z}(n)$ (формула приведена только для $x(n)$):

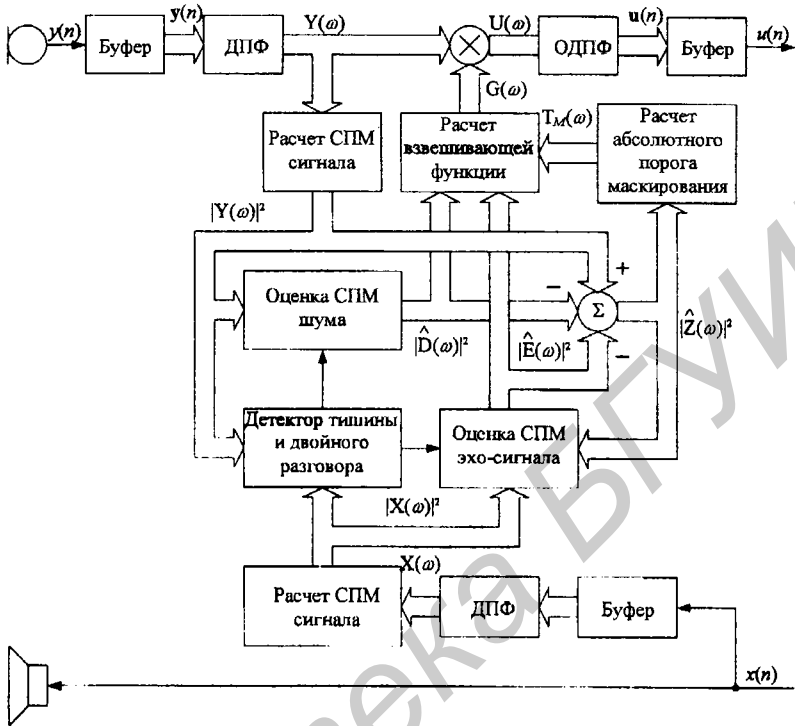


Рис. 4. Структурная схема новой комбинированной системы подавления эхо-сигнала и редактирования шума на основе психоакустически мотивированного правила спектрального вычитания

$$P_x(m) = \begin{cases} 0, & P_x(m) \leq P_{\alpha 1}(m); \\ \frac{P_x(m) - P_{\alpha 1}(m)}{P_{\alpha 2}(m) - P_{\alpha 1}(m)}, & P_{\alpha 1}(m) < P_x(m) < P_{\alpha 2}(m); \\ 1, & P_{\alpha 2}(m) \leq P_x(m), \end{cases}$$

где $P_x(m)$ – мощность сигнала, $P_{\alpha 1}(m)$ – нижний порог мощности, получаемый на основе оценки мощности шумов, $P_{\alpha 2}(m)$ – верхний порог мощности, который рассчитывается на базе средних значений мощности сигнала в периоды речевой активности.

Оценки СПМ эхо-сигнала для m -го фрейма обработки $|\hat{E}(m, \omega)|^2$ основана на блочном подходе вычислений:

$$|\hat{E}(m, \omega)|^2 = \sum_{k=0}^{Q-1} |W(m, k, \omega)|^2 \cdot |X(m-k, \omega)|^2,$$

и секционировании импульсной характеристики фильтра эхо-компенсатора (фильтра предсказателя):

$$|W(m+1, i, \omega)|^2 = |W(m, i, \omega)|^2 + 2\eta(p_x - p_{dt}) \frac{\left(|Y(m, \omega)|^2 - |\hat{E}(m, \omega)|^2 \right) \cdot |X(m-i, \omega)|^2}{\|X(m, \omega)\|^2}, \quad i = \overline{0, Q-1},$$

где $|X(m, \omega)|^2$ – оценка СПМ сигнала дальнего диктора $x(m, n)$, $|X(m, \omega)|^2 = \{X(m-i, \omega)\}^2, \quad i = \overline{0, Q-1}$ – вектор оценок, $|W(m, k, \omega)|$ – ЛЧХ k -й секции импульсной характеристики фильтра, Q – число секций, B – размер блока данных, $N = Q \cdot B$ – длина импульсной характеристики, p_x – оценка вероятности присутствия речевой активности дальнего диктора, p_{dt} – оценка вероятности присутствия двойного разговора. Использование в алгоритме корректировки параметров фильтра-предсказателя эхо-сигнала $|W(m, k, \omega)|^2$ вероятностных оценок присутствия речевой активности p_x и p_{dt} позволяет повысить скорость сходимости фильтра-предсказателя в переходные моменты между периодом тишины, речевой активностью и двойным разговором.

В главе предлагается модификация метода измерения СПМ шумов окружающей среды $|\hat{D}(m, \omega)|^2$ на основе экспоненциального усреднения и использовании значений вероятности речевых пауз:

$$\begin{aligned} |\hat{D}(m, \omega)|^2 &= \alpha(p_y) \cdot |Y(m, \omega)|^2 + (1 - \alpha(p_y)) \cdot |\hat{D}(m-1, \omega)|^2, \\ \alpha(p_y) &= (1 - \alpha_0) \cdot p_y + \alpha_0, \end{aligned}$$

где $\alpha(p_y)$ – сглаживающий коэффициент, являющийся функцией оценки вероятности присутствия речевой активности p_y . Введение в метод переменного коэффициента $\alpha(p_y)$, зависящего от вероятности присутствия во входном сигнале речевой активности, в отличие от классических подходов, позволяет повысить достоверность измерений в моменты перехода от начала речевой активности к паузе и наоборот.

В главе 4 представлена аппаратная реализация устройства повышения качества речевого сигнала в мультимедиа системах. Процессор TMS3206711 компании Texas Instruments с VLIW архитектурой был использован для построения одноканальной комбинированной системы подавления эхо-сигнала и редактирования шума. Анализ емкостной, а также вычислительной сложности алгоритмического обеспечения системы, представленной на рис. 4, показывает, что при размере входного буфера $B = 8$ отсчетов требуется не более $7,5 \cdot 10^7$ машинных циклов в секунду, а при выборе блока входных данных в 32 и более отсчетов число машинных циклов в секунду снижается до $1,5 \cdot 10^7$. Так как ЦПОС TMS320C6711 при частоте 100 МГц обеспечивает $1 \cdot 10^8$ машинных циклов в секунду, то для аппарат-

ной реализации устройства достаточно одного процессора. Сравнительный анализ вычислительной сложности перцентуально мотивированного метода повышения качества речевого сигнала в частотной области и метода, в основу работы которого положен НМНК КИХ фильтр во временной области, показывает, что при порядках фильтра более 1000, что соответствует типовой длительности эхо-сигнала в офисном помещении, выигрыш составляет в 10 и более раз.

Процессор повышения качества речевого сигнала реального времени в мультимедиа системах имеет следующие технические характеристики: аппаратная платформа: TMS320C6711 DSP Starter Kit производства компании Texas Instruments; тактовая частота микропроцессора 150 МГц, 16 разрядный АЦП AD535, внешняя память: SDRAM - 16MB, Flash ROM - 128KB, параллельный порт связи с PC компьютером; частотный диапазон обрабатываемых сигналов 0-4000Гц; частота дискретизации 8000Гц; подавление эхо-сигнала длительности (до 500мс); уровень ослабления возвращаемого в канал эхо-сигнала (среднее *ERLE*) 35-38 дБ; уровень подавления шума 15-20 дБ; реконструируемый речевой сигнал не имеет артефактов.

Глава 5 отражает экспериментальную часть работы. Эксперименты проводились для двух основных случаев: шумы окружающей среды отсутствуют и входные сигналы системы зашумлены (присутствует акустическое эхо и шум). Пример работы устройства показан на рис. 5.

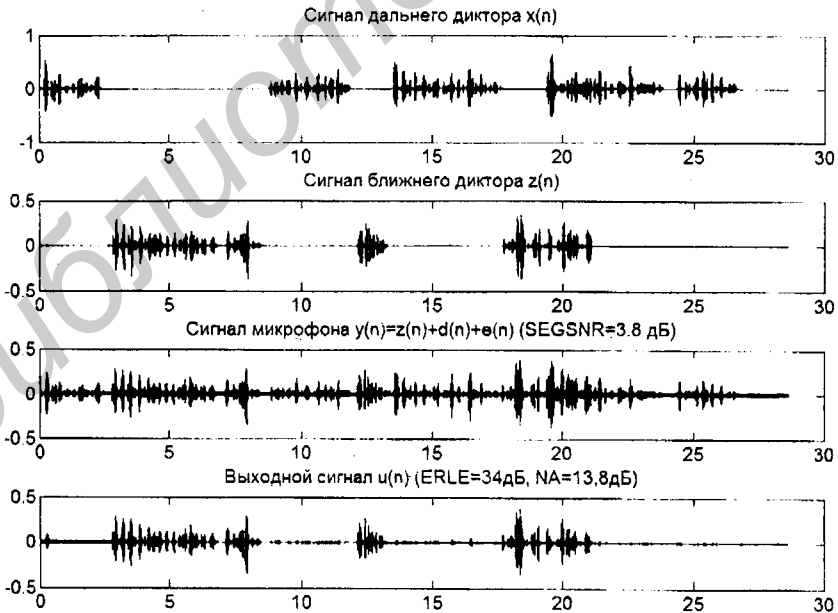


Рис. 5. Временные диаграммы работы устройства

Тестирование предложенных в диссертации алгоритмов (кривая 1 на рис. 6) компенсации эхо-сигналов различной длительности выполнялось в сравнении с алгоритмами на основе НМНК КИХ фильтра (кривая 2 на рис. 6). Анализ зависимостей, приведенных на рис. 6а, показывает, что скорость сходимости разработанных алгоритмов выше для больших N . Зависимость уровня ослабления возвращаемого в канал эхо-сигнала $ERLE_C$ от его длительности иллюстрируется на рис. 6б. Как видно из графиков, наблюдается ухудшение $ERLE_C$ при увеличении длительности эхо-сигнала при осуществлении обработки во временной области (кривая 2), а $ERLE_C$ психоакустически мотивированного метода слабо изменяется при длительности эхо-сигнала в промежутке от 40 до 400 мс (кривая 1).

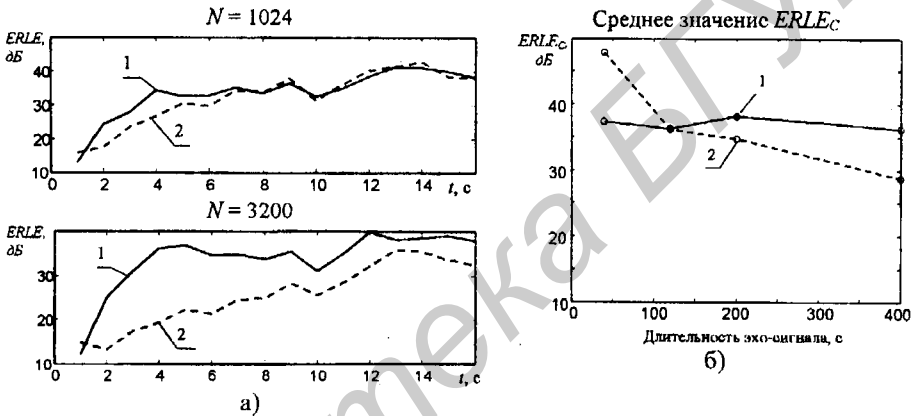


Рис. 6. Величина ослабления возвращаемого в канал эхо-сигнала $ERLE_C$

Анализ работы подсистемы редактирования шума показывает, что уровень ослабления шума NA в предлагаемом методе несколько меньше, чем у прототипа (рис. 7а). Однако, искажения, вносимые в речевой сигнал, при этом значительно меньше (рис. 7б), что делает восприятие человеком выходного сигнала более комфортным. Использовалась оценка качества реконструированного речевого сигнала BSD (Bark spectral distortion).

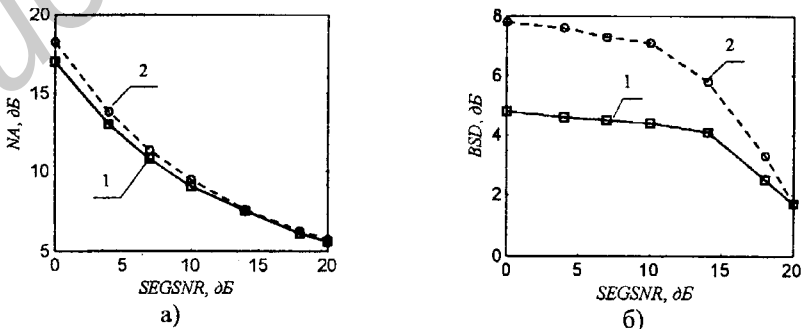


Рис. 7. Оценки степени подавления шумов окружающей среды

ЗАКЛЮЧЕНИЕ

Основные научные и практические результаты диссертационной работы следующие:

1. На основании анализа подходов построения перцептуально мотивированных методов и устройств повышения качества речевого сигнала в мультимедиа системах показано, что вариант одноканальной комбинированной системы компенсации эхо-сигнала и редактирования шума на основе обработки в частотной области с учетом психоакустических особенностей восприятия человеком акустической информации позволяет значительно снизить вычислительную сложность алгоритмов [1, 2, 3], осуществлять параллельно-поточную обработку на однокристальном VLIW процессоре с поддержкой SIMD инструкций в реальном масштабе времени [6, 8].

2. Психоакустически мотивированное правило спектрального взвешивания, которое является модификацией метода спектрального вычитания, но в отличие от последнего, позволяет организовать одновременное подавление акустического эхо-сигнала и шумов окружающей среды. Учет при разработке правила психоакустического эффекта маскирования позволяет значительно снизить искажения вносимые методом, проявляющиеся в виде эффекта “музыкального тона” [5, 7, 10].

3. Новая структура комбинированной системы повышения качества речевого сигнала мультимедиа систем на основе единого перцептуального правила спектрального вычитания в частотной области, которая характеризуется по сравнению с известными компромиссом между вычислительной сложностью алгоритмов, малой задержкой трансмиссии и высоким перцептуальным качеством реконструированного речевого сигнала [5, 8, 9].

4. Метод оценки спектральной плотности мощности (СПМ) акустического эхо-сигнала по оценкам СПМ сигнала дальнего диктора, отличительной особенностью которого является то, что используется блочная обработка входного сигнала и корректировка весовых коэффициентов фильтра-предсказателя эхо-сигнала на основе нормализованного метода наискорейшего спуска осуществляется с использованием вероятности речевой активности детектора речевых пауз и двойного разговора, что позволяет повысить скорость сходимости фильтра-предсказателя в переходные моменты между периодом тишины и речевой активностью, а также степень подавления эхо-сигнала большой длительности [4, 5, 9].

5. Модифицированный метод оценки СПМ шумов окружающей среды с учетом вероятностных оценок речевой активности детектора речевых пауз и двойного разговора, что, в отличие от классических подходов, позволяет повысить дос-

товерность измерений в моменты перехода от начала речевой активности к паузе и наоборот [5, 7, 9, 10].

6. Осуществлена аппаратная реализация устройств повышения качества речевого сигнала в реальном времени на базе ЦПОС TMS320C541 и аппаратного модуля TMS320C6711 DSP Starter Kit. Устройства обеспечивают подавления эхо-сигнала большой длительности (до 500мс), при этом уровень ослабления возвращаемого в канал эхо-сигнала равен 35-38 дБ, а уровень подавления шума составляет 15-20 дБ и реконструируемый речевой сигнал не имеет артефактов, что соответствует лучшим разработкам подобных устройств [5, 8, 9, 10].

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в научных журналах и сборниках научных трудов

1. Petrovsky A., Bielawski K., Anoshenko A. Hands-free radiotelephony communication devices with combine front end processing systems: global approaches in the time and frequency domain // Journal of the University of Applied Sciences Mittweida (FH). – Mittweida, Germany, 1998. – Band C, Nr.3. – P.135-142.
2. Аношенко А.Е. Адаптивные фильтры в частотной области и их применение для подавления эхо-сигнала // Известия белорусской инженерной академии. – Минск, 1998. – № 2(6)/1. – С. 53-56.
3. Аношенко А.Е. Анализ вычислительной сложности алгоритмов подавления эхо-сигнала в мобильных средствах связи на базе адаптивной фильтрации в частотной области // Радиотехника и электроника. – Минск, 1999. – Вып. 23. С. 120-123.
4. Аношенко А.Е., Петровский А.А. Применение психоакустического подхода в системах совместного подавления шума и эха в мобильных средствах связи // Известия белорусской инженерной академии. – Минск, 1999. – № 1(7)/1. – С. 10-13.
5. Аношенко А.Е., Петровский А.А. Перцептуальный метод повышения качества частотно-ограниченного речевого сигнала: одноканальный вариант, комбинированная система, частотный подход // Доклады Белорусского государственного университета информатики и радиоэлектроники. – Минск, 2003. – №4. – С. 71-81.

Статьи в сборниках трудов научно-технических конференций

6. A.A. Petrovsky, A.E. Anoshenko, A.A. Petrovsky, Implementation of decoupled partitioned block frequency domain adaptive filter based on the pipeline multi-DSP systems // Техника и технология связи: Материалы I Международной научно-

технической конференции, Минск, 22-26 февраля 1999 / Вестник связи. – Минск, 1999. – № 1. – С.39-42.

7. Петровский А.А., Аношенко А.Е. Комбинированная система подавления эхо и шумового сигнала на основе обработки в частотной области с использованием психоакустического подхода // Цифровая обработка сигналов и ее применение : Сб. докладов 2-й Международной научно-технической конференции, Москва, 21-24 сентября 1999. – Москва, 1999. – Т. 1. – С. 166-169.

8. Петровский А.А., Аношенко А.Е. Система подавления акустического эхо и шумов окружающей среды, использующая алгоритмы с психоакустической мотивацией, на базе одного процессора TMS320C541 // Цифровая обработка сигналов и ее применение: Сб. докладов 3-й Международной научно-технической конференции, Москва, 29 ноября - 1 декабря 2000. – Москва, 2000. – Т. 2. – С. 236-241.

9. Петровский А.А., Аношенко А.Е. Метод подавления эхо-сигнала и шумов окружающей среды на основе спектрального вычитания с психоакустической мотивацией // Цифровая обработка сигналов и ее применение: Сб. докладов 4-й Международной научно-технической конференции, Москва, 27 февраля - 1 марта 2002. – Москва, 2002. – Т. 2. – С.450-453.

10. Петровский А.А., Аношенко А.Е. Повышение качества речевого сигнала на основе психоакустического подхода // Цифровая обработка сигналов и ее применение: Сб. докладов 5-й Международной научно-технической конференции, Москва, 12-14 марта 2003. – Москва, 2003. – Т. 2. – С.323-326.

А.Е.Аношенко

РЭЗЬЮМЭ

Ліошанка Аляксей Яўгенавіч

Псіхаакустычна матываваныя метады і ўстройства павышэння якасці маўленчага сігналу ў мультымедыя сістэмах

Ключавыя словы: кампенсаванне рэха-сігналу, рэдактаванне шума, спяктральнае вычытанне, мультымедыя сістэмы, павышэнне якасці маўленчага сігналу, псіхаакустыка, парог маскіравання.

Мэтай працы з'яўляецца распрацоўка метадаў і прыладаў кампенсавання акустычнага рэха-сігналу і папярэжэнні шуму ў навакольнага асяроддзя маўленчага сігналу мультымедыя сістэм рэальнага часу, якія базуюцца на апрацоўцы ў частотнай вобласці і псіхаакустычных асаблівасцях успрымання чалавека.

Аб'ектам даследаванняў з'яўляюцца сістэмы кампенсавання рэха-сігналу і рэдактавання шуму ў навакольнага асяроддзя. Прадметам даследавання з'яўляюцца метады кампенсавання рэха-сігналу і рэдактавання шуму ў навакольнага асяроддзя.

Прапанаваны псіхаакустычна матываваны метады павышэння якасці маўленчага сігналу ў частотнай вобласці, які характарызуецца у параўнанні з вядомымі на парадак меншай вылічальнай складанасцю і адзіным для рэдактавання шуму і рэха-сігналу правілам спяктральнага узважвання; распрацаваны метады адзнакі спяктральнай шчыльнасці магутнасці (СШМ) акустычнага рэха-сігналу па адзнаках СШМ сігналу далёкага дыктара на аснове блочнай апрацоўкі ўваходнага сігналу і з улікам верагоднасці прысутнасці маўленчай актыўнасці; прапанаваны мадыфікаваны метады адзнакі СШМ шуму ў навакольнага асяроддзя з выкарыстаннем імавернасных адзнак моўнай актыўнасці дэтэктара маўленчых паўзаў; распрацаваная структура камбінаванай сістэмы падвышэння якасці маўленчага сігналу мультымедыя сістэм на аснове адзінага псіхаакустычна матываванага правіла спяктральнага выпічэння ў частотнай вобласці, выкананая апаратна-праграмавая рэалізацыя перацэнтральных аднаканальных прыладаў падвышэння якасці маўленчага сігналу ў мультымедыя сістэмах з апрацоўкай у частотнай вобласці на аснове паралельна-паточнай апрацоўкі на аднакрыштальным VLIW працэсары з падтрымкай SIMD інструкцый у рэальным маштабе часу.

Вынікі даследаванняў перададзены ў канструктарскае бюро ААТ "Мінскі прыборабудуначы завод".

РЕЗЮМЕ

Анощенко Алексей Евгеньевич

Психоакустически мотивированные методы и устройства повышения качества речевого сигнала в мультимедиа системах

Ключевые слова: компенсация эхо-сигнала, редактирование шума, спектральное вычитание, мультимедиа системы, повышение качества речевого сигнала, психоакустика, порог маскирования.

Целью работы является разработка методов и устройств подавления акустического эхо-сигнала и снижения шумов окружающей среды в речевом сигнале мультимедиа систем реального времени, базирующихся на обработке в частотной области и психоакустических особенностях восприятия человека.

Объектом исследований являются системы подавления эхо-сигнала и редактирования шумов окружающей среды. Предметом исследования являются методы подавления эхо-сигнала и редактирования шумов окружающей среды.

Научная новизна результатов заключается в следующем: предложен психоакустически мотивированный метод повышения качества речевого сигнала в частотной области, характеризующийся по сравнению с известными на порядок меньшей вычислительной сложностью и единым для редактирования шума и эхо-сигнала правилом спектрального взвешивания; разработан метод оценки спектральной плотности мощности (СПМ) акустического эхо-сигнала по оценкам СПМ сигнала дальнего диктора на основе блочной обработки входного сигнала и с учетом вероятности присутствия речевой активности; предложен модифицированный метод оценки СПМ шумов окружающей среды с использованием вероятностных оценок речевой активности детектора речевых пауз; разработана структура комбинированной системы повышения качества речевого сигнала мультимедиа систем на основе единого психоакустически мотивированного правила спектрального вычитания в частотной области, выполнена аппаратно-программная реализация перцептуальных одноканальных устройств повышения качества речевого сигнала в мультимедиа системах с обработкой в частотной области на основе параллельно-поточной обработки на однокристальном VLIW процессоре с поддержкой SIMD инструкций в реальном масштабе времени.

Результаты диссертационной работы переданы для внедрения в конструкторское бюро ОАО "Минский приборостроительный завод".

SUMMARY

of the dissertation thesis by Alexei E. Anoshenko

Psychoacoustically motivated methods and devices of speech signal improvement in multimedia systems.

Keywords: echo cancellation, noise reduction, spectral subtraction, multimedia systems, speech signal improving, psychoacoustics, masking threshold.

The purpose of work is the development of methods and devices of an acoustic echo cancellation and noise reduction in the real time speech multimedia systems based on frequency domain processing and on the psychoacoustics features of human perception.

Objects of the research are the systems of an echo cancellation and noise reduction. Subject of the research are the methods of echo cancellation and noise reduction.

The scientific novelty of the results consists of the following issues: new psychoacoustically motivated rule of spectral weighing that allows organizing simultaneous cancellation of an acoustic echo and noise along with reduction of distortions ("musical tone" effect) brought in by the spectral weighing is offered; the modified method of an estimation of the noise power spectral density during speech pauses using exponential averaging and probability of presence of speech activity is offered; the method of an estimation of power spectral density of an acoustic echo that uses block processing of a signal with separate estimation of spectrum components is developed; the modified method of updating the weight factors of the echo-signal power spectral density predictor created on the basis of the normalized least mean square method and using real values for the probability of presence of speech activity is offered; the combined system of echo cancellation and noise reduction on the basis of the psychoacoustically motivated method of spectral weighing with the additional block of preliminary noise suppression is developed.

The results of the research work are submitted for implementation to the Research and Development department of "Minskii Priborostroitelnyi zavod" joint-stock company.

АНОШЕНКО
Алексей Евгеньевич

**Психоакустически мотивированные методы и устройства повышения
качества речевого сигнала в мультимедиа системах**

Специальность 05.13.05 – Элементы и устройства вычислительной техники и
систем управления

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Подписано в печать 29.03.2005.	Формат 60x84 1/16	Бумага офсетная.
Гарнитура «Таймс».	Печать ризографическая.	Усл. печ. л. 1,63.
Уч.-изд. л. 1,4.	Тираж 70 экз.	Заказ 211.

Издатель и полиграфическое исполнение: Учреждение образования:

«Белорусский государственный университет информатики и радиоэлектроники»

Лицензия на осуществление издательской деятельности №02330/0056964 от 01.04.2004

Лицензия на осуществление полиграфической деятельности №02330/0133108 от 30.04.2004

220013, Минск, П. Бровки, 6