На правах рукописи

Новоселов Сергей Александрович

ВЫДЕЛЕНИЕ И ПРЕДОБРАБОТКА СИГНАЛОВ В СИСТЕМАХ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧЕВЫХ КОМАНД

Специальность 05.12.04
Радиотехника, в том числе системы и устройства телевидения

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Владимир – 2011

Работа выполнена на кафедре динамики электронных систем Ярославского государственного университета им. П.Г. Демидова

Научный руководитель: доктор технических наук
Приоров Андрей Леонидович

Официальные оппоненты: доктор физико-математических наук, профессор
Рау Валерий Георгиевич

кандидат технических наук
Меньшиков Борис Николаевич

Ведущая организация: ОАО «Ярославский радиозавод»

Защита диссертации состоится «30» декабря 2011 г. в 14.00 часов на заседании диссертационного совета Д 212.025.04 при Владимирском государственном университете имени Александра Григорьевича и Николая Григорьевича Столетовых по адресу: 600000, г. Владимир, ул. Горького, д. 87, ВлГУ, корп. 3, ФРЭМТ, ауд. 301.

С диссертацией можно ознакомиться в библиотеке Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых.

Автореферат разослан « 29» ноября 2011 г.

Отзывы на автореферат, заверенные печатью, просим направлять по адресу: 600000, г. Владимир, ул. Горького, д. 87, ВлГУ, корп. 3, ФРЭМТ.

Ученый секретарь диссертационного совета
доктор технических наук, профессор А.Г. Самойлов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В настоящее время системы распознавания речи получают все большее распространение, особенно в тех приложениях, где речевой диалог является наиболее удобным средством управления и обмена информацией с техническими средствами. Но чем выше достоверность распознавания, тем сложнее такая система, и тем выше ее стоимость. Получение эффективной системы голосового управления в настоящее время является важной задачей, требующей создания методов, позволяющих получать высокую достоверность распознавания речевых команд.

Речевые сигналы, с которыми приходится иметь дело на практике, всегда в той или иной степени зашумлены. Начальные этапы выделения и фильтрации речевой команды являются важными и определяющими в решении задачи организации системы управления. Ошибки в выделении команды, а также наличие посторонних шумов в ней, приводят к значительному снижению вероятности правильного распознавания. Для разработки системы голосового управления необходимо уделить особое внимание процессу выделения «чистого» речевого сигнала из входного зашумленного. Сложным моментом является также выделение команды на фоне нестационарных шумов.

Для решения задачи выделения команд используют различные методы детектирования речевой активности (ДРА). Алгоритм ДРА обеспечивает классификацию сегментов речевого сигнала по типу «речь» или «не речь». В большинстве случаев используют простые и быстрые алгоритмы, построенные на основе пороговых сравнений кратковременных энергий, количества переходов через ноль, корреляционных параметров, энергий спектральных подполос и т.п. На практике чаще имеют дело с нестационарными фоновыми шумами (паразитные хлопки, щелчки и др.), иногда – с шумами значительной интенсивности, например, шум в кабине самолета, автомобиля. В этих случаях задача правильной сегментации речевого сигнала на команды значительно осложняется. Установлено, что простой детектор речевой активности на основе пороговой классификации не способен качественно решить проблему.

Алгоритмы распознавания незашумленных речевых команд уже сегодня показывают хорошие результаты. Но, при наличии внешних шумов результаты автоматического распознавания существенно ухудшаются. Это обстоятельство ограничивает сферу применения систем распознавания речи и приводит к постановке задачи предобработки речевого сигнала до стадии его распознавания.

На сегодняшний момент известно множество методов повышения качества и разборчивости речи. Но дело в том, что алгоритмы, обеспечивающие повышение качества звучания речи и ее разборчивости для восприятия человеком, могут оказаться неподходящими для решения задачи повышения вероятности верного распознавания в современных системах голосового управления.

Таким образом, проблема разработки новых алгоритмов выделения и фильтрации речевых команд в системах голосового управления является актуальной.

Основополагающие работы по обработке и анализу речевых сигналов связаны с именами таких известных зарубежных ученых, как Рабинер Л., Шафер Р., Янг Б., Мермелштейн П., Левинсон С. и др. Большой вклад в развитие статистического и

регрессионного анализа речевых сигналов внесли работы зарубежных и отечественных ученых Парзена Э., Розенблатта М., Репина В.Г., Тартаковского Г.П., Прохорова Ю.Н., Санникова В.Г. и др.

В настоящее время в радиотехнике широкое распространение получили методы цифровой обработки сигналов, использующие различные варианты вейвлет-преобразований. Это объясняется тем, что вейвлет-функции обеспечивают частотную и временную локализацию, а так же возможность обрабатывать сигнал на разных масштабах. В этой области широко используются работы Малла С., Добеши И., Чуи К., Блаттера К. Метод главных компонент, предложенный Пирсоном К., так же широко применяется в решении задач обработки и распознавания речевых сигналов.

Работы по обнаружению речевых сигналов связаны с именами таких ученых и исследователей как Самбур М., Жао Ю., Мекурла Ф., Рабинер Л., Крашенинников В.Р., Хвостов А.В. и др. Статистические методы детектирования речи тесно связаны с решением задачи об обнаружении разладки. Основополагающие работы в этой области принадлежат отечественным ученым Колмогорову А.Н., Ширяеву А.Н.

В области шумоподавления в речевых сигналах наибольшую известность получили работы ученых Ефрайма Я., Малла Д., Скаларта П., Коэна И. Наиболее применяемыми в этой области являются способы коррекции спектра сигнала, основанные на фильтрации Винера и минимизации среднеквадратичной ошибки.

Необходимым условием эффективной работы систем голосового управления является их устойчивость к воздействию внешних шумов. Данная работа посвящена исследованию ряда задач, связанных с правильным выделением речевых команд и шумоподавлением в них для повышения вероятности верного распознавания.

Целью работы является разработка и исследование методов анализа и обработки речевых сигналов, позволяющих эффективно решать задачи выделения и распознавания речевых команд на фоне внешних акустических шумов.

В соответствии с указанной целью в работе поставлены и решены следующие основные задачи:

- исследование влияния ошибок в определении границ команд на вероятность их верного распознавания в системах голосового управления;
- исследование влияния наличия шумов в командах на вероятность верного распознавания в системах голосового управления;
- исследование помехоустойчивости информативных параметров речевого сигнала и разработка помехоустойчивого метода параметризации речевых сигналов;
- разработка алгоритмов детектирования речевой активности и выделения речевых команд на фоне стационарных и нестационарных шумов;
- разработка алгоритма шумоподавления в речевых командах методом нелокального усреднения;
- разработка метода поиска похожих фрагментов на интервалах стационарности речевого сигнала.

Методы исследования. При решении поставленных задач использованы методы цифровой обработки сигналов, теории вейвлет-преобразований, линейной алгебры, теории факторизации матриц, теории вероятностей и

математической статистики. Широко использовались методы компьютерного моделирования.

Объектом исследования является помехоустойчивая система распознавания речевых команд, применяемая в системах голосового управления техническими устройствами.

Предметом исследований являются методы, обеспечивающие правильное выделение речевых команд на фоне стационарных и нестационарных шумов, а также методы предобработки речевых команд с целью шумоподавления, обеспечивающие повышение вероятности их верного распознавания в условиях стационарных помех.

Научная новизна

1. Разработан метод параметризации речевых сигналов с помощью адаптированного к мел-шкале вейвлет-пакетного преобразования, оператора вычисления энергии Тегера-Кайзера и метода главных компонент.
2. Разработан алгоритм детектирования речевой активности на фоне стационарных и нестационарных шумов с помощью предложенного метода параметризации речевого сигнала и смесей гауссовских распределений.
3. Разработан алгоритм шумоподавления в речевых сигналах методом нелокального усреднения.
4. Разработан метод поиска похожих фрагментов на интервалах стационарности речевого сигнала.

Практическая значимость

1. Предложенный метод параметризации речевого сигнала является помехоустойчивым и позволяет решать задачу выделения речевой активности на фоне интенсивных шумов.
2. Разработанный детектор речевой активности позволяет эффективно проводить классификацию сегментов сигнала по типу «речь» и «не речь» на фоне стационарных и нестационарных помех при отношении сигнал/шум равном -5дБ.
3. Разработанный алгоритм выделения речевых команд на основе предложенного ДРА обеспечивает качественное выделение команд на фоне стационарных и нестационарных помех и позволяет снизить вероятность появления ошибок I-го и II-го родов по сравнению с существующими методами.
4. Предложенный алгоритм шумоподавления в речевых сигналах позволяет улучшить вероятность правильного распознавания в системе голосового управления в условиях стационарных шумов. Оценка вероятности правильного распознавания цифр при стационарном шуме в 10 дБ составляет 93%.

Результаты работы внедрены в соответствующие разработки ОАО «СеверТрансКом» и МОО «Союз криминалистов» г. Ярославль. Отдельные результаты диссертационной работы внедрены в учебный процесс ЯрГУ в рамках дисциплин «Цифровая обработка речевых сигналов», «Цифровые фильтры», а также в научно-исследовательские работы при выполнении исследований в рамках грантов «Развитие теории цифровой обработки сигналов и изображений в технических системах» (грант РФФИ № 06-08-00782, 2006–2008 гг.), «Развитие нелинейной теории обработки сигналов и изображений в радиотехнике и связи» (Программа «Развитие научного потенциала высшей школы (2009–2010 годы)»,

№ 2.1.2/7067). Все результаты внедрения подтверждены соответствующими актами.

Достоверность материалов диссертационной работы подтверждена результатами компьютерного моделирования, демонстрирующими эффективность предложенных алгоритмов в задачах выделения и распознавания речевых команд на фоне шумов.

Апробация работы. Результаты работы докладывались и обсуждались на следующих научно-технических конференциях и семинарах:

- 9–13 Международной конференции «Цифровая обработка сигналов и ее применение», Москва, 2007–2011.
- 61, 64–65 Научной сессии, посвященной Дню радио, РНТОРЭС им. А.С. Попова, Москва, 2006, 2009, 2010.
- VI Всероссийской научно-технической конференции «Информационные технологии в электротехнике и электроэнергетике», Чебоксары, 2004.
- 16 Международной научно-технической конференции «Проблемы передачи и обработки информации в сетях и системах телекоммуникаций», Рязань, 2009.
- XVIII Международной научно-технической конференции «Информационные средства и технологии», Москва, МЭИ, 2010.
- XVI Международной научно-технической конференции «Радиолокация, навигация, связь», Воронеж, 2010.
- Всероссийской конференции «Радиоэлектронные средства передачи и приема сигналов и визуализации информации», Таганрог, 2011.
- Международной научно-практической конференции студентов и молодых ученых «Молодежь и наука: модернизация и инновационное развитие страны», Пенза, 2011.
- IX Международной научно-технической конференции «Перспективные технологии в средствах передачи информации», Суздаль, 2011.

Публикации. По теме диссертации опубликована 21 научная работа, из них 5 статей в рецензируемых журналах, в том числе три статьи в журналах из перечня ВАК, и 1 свидетельство о регистрации программного обеспечения.

Структура и объем работы. Диссертация состоит из введения, трех глав, заключения, списка литературы. Содержание работы изложено на 142 страницах. Список литературы включает 139 наименований. В работе представлено 30 рисунков и 29 таблиц.

Основные научные положения и результаты, выносимые на защиту:

1. Метод параметризации речевых сигналов с помощью адаптированного к мел шкале вейвлет-пакетного преобразования, оператора вычисления энергии Тегера-Кайзера и метода главных компонент;
2. Алгоритм выделения речевой активности на основе разработанного метода параметризации и классификации с помощью статистического метода смесей гауссовских распределений;
3. Алгоритм шумоподавления в речевых сигналах на основе метода нелокального усреднения;
4. Метод поиска похожих фрагментов на интервалах стационарности речевого сигнала.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность выбранной темы, сформулированы цель и задачи исследования, изложены основные положения, выносимые на защиту, показаны научная новизна и практическая значимость работы.

В первой главе сформулирована задача автоматического распознавания речи. При анализе подходов, существующих в области физиологии слуха, выявлено, что все существующие теории восприятия речи могут быть классифицированы по следующим параметрам: первый параметр – это моторный или сенсорный принцип восприятия; второй – его активный или пассивный характер. Данная работа, как и подавляющее большинство алгоритмов обработки и распознавания речи, основывается на сенсорной или акустической модели восприятия речевых сигналов. Утверждается, что в самих акустических параметрах речи сосредоточена вся необходимая информация для проведения распознавания.

Рассмотрены основные этапы эволюции систем автоматического распознавания речи. На основе анализа современного этапа развития речевых технологий сделан вывод о том, что одним из наиболее важных направлений совершенствования систем голосового управления является повышение эффективности работы алгоритмов выделения речевых команд, а также методов предобработки речевых сигналов, позволяющих производить очистку входного речевого сигнала от шума. Этим двум направлениям посвящены второй и третий разделы данной работы.

Рассмотрены два наиболее общих и известных алгоритма детектирования речевой активности, применяемые на практике: алгоритм G.729 Annex B и алгоритм ДРА, предложенный Европейским институтом по стандартизации в области телекоммуникаций ETSI (European Telecommunications Standards Institute), который используется в мобильных системах связи стандарта GSM.

Как показывает практика, классификация сегментов по типу «речь» – «не речь» с помощью простых алгоритмов порогового принятия решения (даже с динамическим порогом) не способна эффективно решать проблему разделения сегментов в условиях нестационарных помех.

Другой проблемой распознавателей команд является то, что речевые команды всегда в той или иной степени зашумлены. При наличии внешних шумов значительной интенсивности результаты их анализа и автоматического распознавания существенно искажаются. Поэтому отдельное внимание в рамках первой главы уделяется рассмотрению цифровых методов повышения качества и разборчивости речи. Отмечается, что алгоритмы, обеспечивающие повышение качества звучания речи и ее разборчивости для восприятия человеком, могут оказаться неподходящими для решения задачи повышения вероятности верного распознавания в современных системах голосового управления.

Выделены основные классы современных методов подавления шума, среди которых отмечены:

1. Методы, основанные на статистическом моделировании речевых сигналов.
2. Методы, основанные на использовании моделей искусственных нейронных сетей.

3. Методы, основанные на оценке параметров шума и пороговой обработке.

Каждой выделенной группе алгоритмов шумоподавления дана характеристика. На основе анализа литературы и проведенных исследований сделан вывод, что наиболее распространенным и эффективным методом шумоподавления в речи является алгоритм спектрального вычитания.

Во второй главе рассмотрена задача выделения речевых команд из общего потока звуковых колебаний. Под «командой» в данном случае понимается отдельно произнесенное слово. Проведено исследование влияния ошибок в выделении команд на вероятность верного распознавания. Отмечена важность эффективного решения задачи выделения команд для всей системы распознавания в целом.

Выделение команд осуществляется с использованием методов детектирования речевой активности.

Речевой сигнал является примером нестационарного процесса, в котором информативным является сам факт изменения его частотно-временных характеристик. Для выполнения анализа таких процессов требуются базисные функции, обладающие способностью выявлять в анализируемом сигнале как частотные, так и временные характеристики. Другими словами, сами функции должны обладать свойствами частотно-временной локализации. Поэтому в работе для вычисления информативных параметров предлагается использовать методы дискретного вейвлет-преобразования (ДВП). Методы вейвлет-анализа положительно зарекомендовали себя в решении многих практических задач обработки и сжатия сигналов. Они хорошо вписываются в архитектуру современных радиотехнических вычислительных систем и легко реализуются на практике в виде цифровых нерекурсивных фильтров.

Идея дискретного вейвлет-анализа состоит в представлении сигнала последовательностью образов с разной степенью детализации. Как показано на схеме (рис. 1), ДВП осуществляется с использованием цифровых вейвлет-фильтров H , G и блоков децимации.

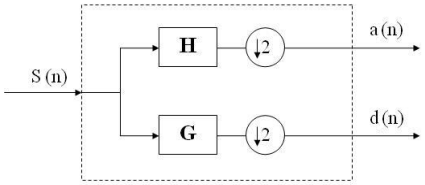

Рис. 1. Одноуровневое вейвлет-разложение

Таким образом, многомасштабный вейвлет-анализ сводится к нахождению коэффициентов аппроксимации $a(n)$ и детализирующих коэффициентов $d(n)$ в разложении сигнала $S(n)$.

Ортогональное вейвлет-преобразование вычисляется с помощью набора цифровых зеркально-сопряженных вейвлет-фильтров, на которые жестко наложены определенные требования в частотной области:

$$G_f(\omega)\overline{H(\omega)} + G_f(\omega+\pi)\overline{H(\omega+\pi)} = 0,$$

$$\left|H(\omega)\right|^2 + \left|H(\omega+\pi)\right|^2 \equiv 1.$$

Здесь $H(\omega)$ и $G(\omega)$ – частотные характеристики фильтров анализа, а $\overline{H(\omega)}$ и $\overline{G(\omega)}$ – частотные характеристики фильтров синтеза.

Для решения задачи детектирования речевой активности предложен алгоритм вычисления новых информативных параметров речевых сигналов – коэффициентов главных компонент мел-вейвлет-пакетных коэффициентов (ГК МВП). Эти параметры рассчитываются с учетом восприятия звуков человеком и, как показали исследования, обладают большей помехоустойчивостью, чем спектрограммы и мелкепстральные коэффициенты.

Вычисление ГК МВП осуществляется с использованием вейвлет-пакетного разложения сигнала. В данном случае предложено адаптировать дерево вейвлет-пакетного преобразования с учетом нелинейной шкалы восприятия частот мел (рис. 2, рис. 3). Предлагается также рассчитывать для каждой полученной полосы d

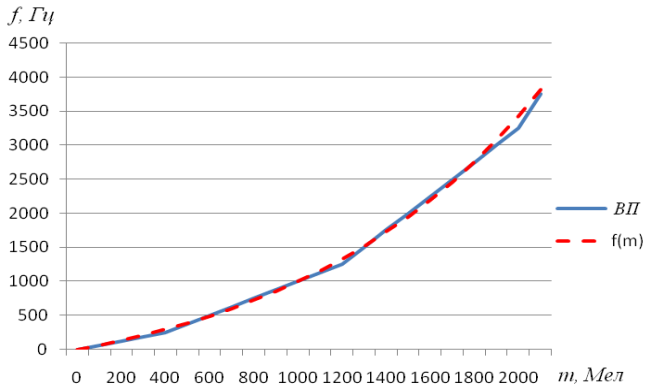

Рис. 2. Кусочно-линейная аппроксимация зависимости шкалы мел с помощью вейвлет-пакетного (ВП) разложения

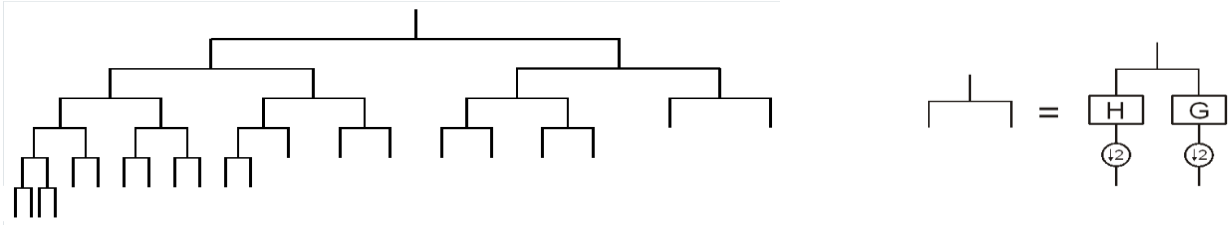

Рис. 3. Дерево вейвлет-пакетного разложения, адаптированное к мел-шкале

мел-вейвлет-пакетного разложения следующий параметр Te:

$$Te(d) = \log\left(\frac{1}{I_d}\sum_{i=1}^{I_d}\left|\Psi(s_d(i))\right|\right),$$

где s_d – коэффициенты подполосы d; I_d – количество этих коэффициентов; $\Psi(s_d(i)) = s_d(i)^2 - s_d(\mathrm{i}-1)\mathrm{s_d}(\mathrm{i}+1)$ – оператор вычисления энергии Тегера-Кайзера i-го отсчета s_d.

Применение метода главных компонент позволяет приводить ковариационные матрицы полученных выше параметров $Te_t(d)$ (здесь t – зависимость от времени) к

диагональному виду и уменьшает размерность информативных векторов. Сделано предположение о том, что базисы главных компонент $Te_t(d)$ являются информативными с точки зрения описания различных классов звуковых колебаний, и каждый класс звуков может быть характеризован «своим собственным» базисом. Это позволило значительно снизить вычислительную сложность расчета информативных параметров и использовать при обработке сигнала базисы главных компонент, построенные на этапе обучения.

В работе предложен алгоритм выделения речевой активности на основе применения статистических моделей смесей гауссовских распределений. Детектор речевой активности, реализуемый в данном подходе, основан на представлении функций плотности распределения вероятностей информативных параметров сигнала моделями смесей нормальных распределений (рис. 4).

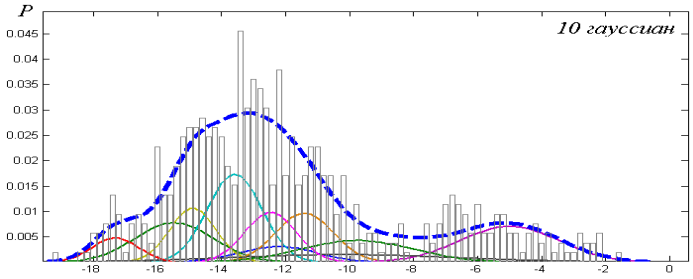

Рис. 4. Моделирование распределений одного элемента вектора информативных параметров речевого сигнала

Модель смесей гауссовских распределений представляет собой метод описания функции плотности распределения D–мерного вектора параметров $\mathbf{x}(t)$ сигнала с помощью взвешенной суммы N гауссовских распределений:

$$p(\mathbf{x}(t) \mid \lambda) = \sum_{i=1}^{N} C_i b_i(\mathbf{x}(t) \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),$$

где λ – обозначение модели; $b_i(\mathbf{x}(t) \mid \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, $i = 1..N$, – функции плотности распределения вероятностей составляющих модели и C_i, $i = 1..N$ – веса компонентов модели. Каждый компонент является D–мерной гауссовской функцией распределения.

Такой алгоритм является обучаемым и дает возможность учитывать несколько типов фоновых шумов и несколько типов голосов дикторов одновременно для обеспечения требуемой надежности обнаружения речи.

На основе предложенного алгоритма ДРА разработан алгоритм автоматического выделения голосовых команд из общего потока звуковых колебаний (рис. 5). Начало команды фиксируется в момент детектирования перехода типа «не речь» – «речь», а окончание определяется на основе детектирования перехода типа «речь» – «не речь» при учете максимальной и минимальной длительности команды.

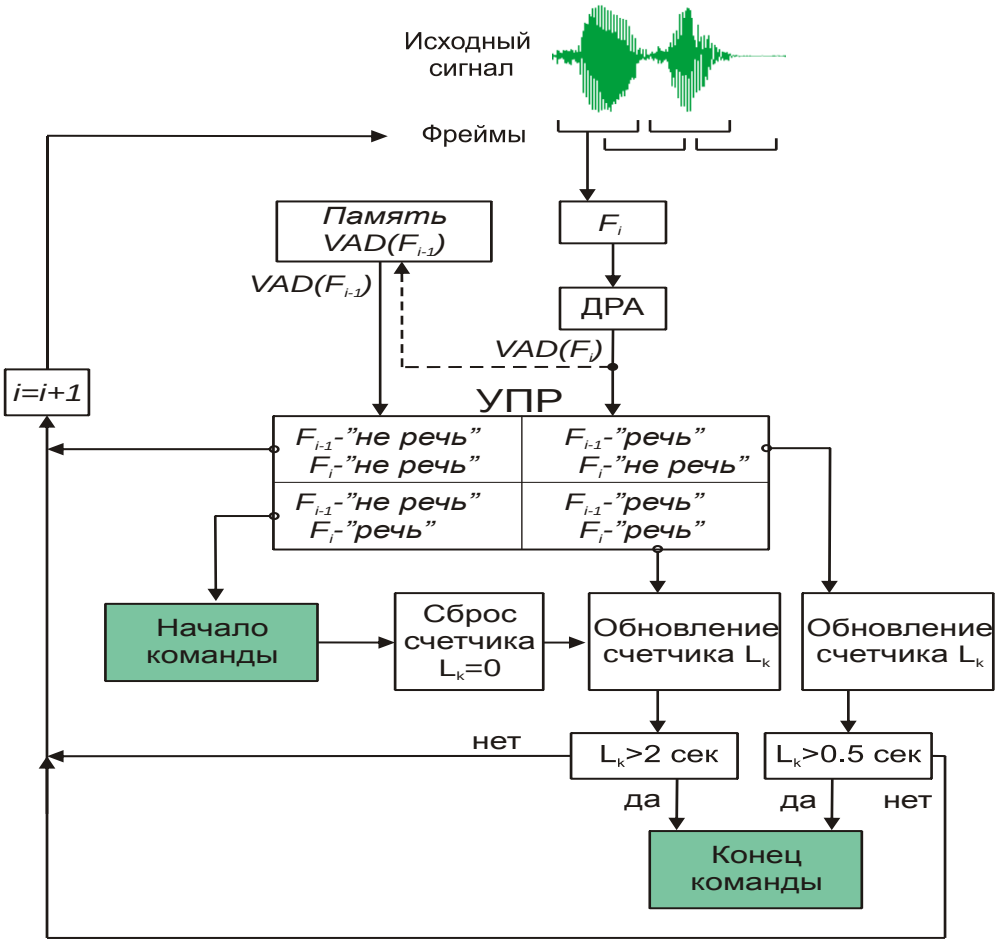

Рис. 5. Схема алгоритма выделения речевых команд

На схеме: F_i – текущий сегмент, F_{i-1} – предыдущий сегмент, L_k – счетчик длительности команды, $VAD(F_i)$ – решение детектора речевой активности о принадлежности сегмента F_i к классу «речь» или «не речь»; УПР – устройство принятия решения.

Проведены исследования алгоритма выделения команд и показана его эффективность для случаев стационарных и нестационарных шумов.

<u>В третьей главе</u> проведено исследование влияния стационарных шумов на вероятность правильного распознавания речевых команд, а также разработка и исследование метода предобработки, позволяющего повысить эту вероятность. Алгоритмы распознавания незашумленных речевых сигналов уже сегодня показывают хорошие результаты, но при наличии внешних шумов результаты автоматического распознавания речи существенно ухудшаются. Это обстоятельство ограничивает сферу применения систем распознавания речи и приводит к постановке задачи предобработки речевого сигнала до стадии распознавания. Основная цель предобработки – провести шумоподавление в речевом сигнале для повышения вероятности его верного распознавания.

Предложенный метод шумоподавления базируется на хорошо зарекомендовавшем себя в области цифровой обработки изображений методе нелокального усреднения (Non-Local Means, NLM). При адаптации метода для задачи подавления аддитивного белого гауссовского шума (АБГШ) в речевом

сигнале учитываются особенности речи, благодаря чему получившийся алгоритм значительно отличается от своего двумерного аналога.

В разработанном методе обработка сигнала осуществляется во временной области. Можно выделить два основных этапа: поиск похожих фрагментов и усреднение этих фрагментов, позволяющие значительно снизить дисперсию аддитивного шума. Длина фрагмента выбирается в пределах интервала 40–80 отсчетов (установлено экспериментально). Поиск похожих фрагментов ведется в пределах окна длиной 400 отсчетов, так как при частоте дискретизации 8000 отсчетов в секунду это соответствует 50 мс, то есть сравнимо с интервалом стационарности речевого сигнала, который равен примерно 30 мс.

Одним из наиболее важных, определяющих факторов является выбор опорного сигнала, по которому проводится поиск похожих фрагментов на интервалах речи длительностью 50 мс. В работе предлагается использовать в качестве опорного сигнал, полученный из зашумленного сигнала путем шумоподавления с помощью метода спектрального вычитания (Spectral Subtraction, SS).

Для общности обозначим фрагменты текущего отрезка речевого сигнала векторами $\vec{S}_i$, соответствующие им шумовые фрагменты текущего отрезка опорного сигнала – векторами $\vec{n}_i$, а соответствующие им шумовые фрагменты текущего отрезка исходного обрабатываемого сигнала – векторами $\vec{N}_i$.

Похожесть речевых фрагментов-векторов определяется евклидовым расстоянием между ними. Речевые фрагменты $\vec{S}_i$ и $\vec{S}_j$ будем считать похожими, если евклидово расстояние между ними не превышает некоторого порогового значения $t_{исх}$:

$$\left\|\vec{S}_j - \vec{S}_i\right\| \le t_{исх}.$$

Порог $t_{исх}$ соответствует допустимым незначительным отклонениям похожих речевых блоков друг от друга.

Требуется для некоторого выбранного вектора $\vec{S}_j + \vec{n}_j,\ j \in [1; L]$, из всего набора фрагментов $\{M_{\vec{n}} : \vec{S}_i + \vec{n}_i\}$ опорного сигнала найти такое множество векторов $\{O_{\vec{n}} : \vec{S}_o + \vec{n}_o\}$, что $\forall \vec{S}_o : \left\|\vec{S}_j - \vec{S}_o\right\| \le t_{исх}$, т.е. любой $\vec{S}_o$ является похожим на $\vec{S}_j$. Рассмотрим евклидово расстояние между зашумленным фрагментом $\vec{S}_j$ и произвольным вектором $\vec{S}_i$. Используя неравенство треугольника, получим:

$$\left\|(\vec{S}_j + \vec{n}_j) - (\vec{S}_i + \vec{n}_i)\right\| \le \left\|\vec{S}_j - \vec{S}_i\right\| + \left\|\vec{n}_j - \vec{n}_i\right\|.$$

Если фрагменты $\vec{S}_j$ и $\vec{S}_i$ похожи, то положим, что $\left\|\vec{S}_j - \vec{S}_i\right\| \le t_{исх} << \left\|\vec{n}_j - \vec{n}_i\right\|$. Дополнительно, учитывая стационарность шума и используя решение детектора речевой активности, можно оценить максимальную амплитуду шумового вектора $\vec{n}_i$ на интервалах, где отсутствует речевая активность и $\left\|\vec{n}_i\right\| \le \left\|\vec{n}\right\|_{\max}$. Тогда имеем:

$$\left\|(\vec{S}_j + \vec{n}_j) - (\vec{S}_i + \vec{n}_i)\right\| \le \left\|\vec{n}_j - \vec{n}_i\right\| \le 2\left\|\vec{n}\right\|_{\max}.$$

Последнее неравенство определяет условия похожести двух зашумленных речевых фрагментов и задает порог, который необходимо применять для нахождения множества $\{O_{\vec{n}} : \vec{S}_o + \vec{n}_o\}$:

$$\vec{S}_i + \vec{n}_i \in \{O_{\vec{n}}\} : \left\|(\vec{S}_j + \vec{n}_j) - (\vec{S}_i + \vec{n}_i)\right\| \le 2\left\|\vec{n}\right\|_{\max}.$$

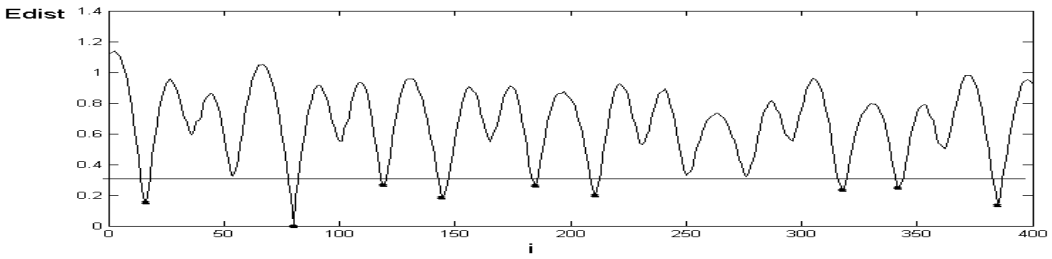

Рис. 6. Евклидово расстояние между текущим и другими фрагментами зашумленного отрезка речевого сигнала

Метод поиска, предложенный выше, приводит к некоторому приближению множества похожих фрагментов. Однако следует иметь в виду, что смежные блоки выбранного фрагмента речевого сигнала являются «ложно» похожими на сам фрагмент из-за высокой корреляции соседних отчетов сигнала. Если их не исключить из рассмотрения и произвести оценку незашумленного блока путем нелокального усреднения по найденному множеству зашумленных векторов, то получим сглаженную копию искомого, потеряв при этом часть информации. Для исключения «ложных» фрагментов из множества $\{O_{\vec{n}} : \vec{S}_o + \vec{n}_o\}$ предлагается находить позиции локальных минимумов функции расстояния между векторами. Эти позиции соответствуют положениям пригодных для усреднения фрагментов (рис. 6). Таким образом, из рассмотрения исключаются окрестности минимумов, соответствующие «ложным» фрагментам.

После того как множество похожих фрагментов $O_{\vec{n}}$ для заданного фрагмента $\vec{S}_j + \vec{n}_j$ найдено, с помощью гауссовского ядра w_o реализуется процесс взвешенного усреднения множества векторов $\{O_{\vec{N}} : \vec{S}_o + \vec{N}_o\}$ отрезка обрабатываемого зашумленного сигнала, соответствующего векторам множества $\{O_{\vec{n}} : \vec{S}_o + \vec{n}_o\}$ опорного сигнала, с целью нахождения оценки $\vec{S}_j'$:

$$\vec{S}_j' = \sum_{\forall o} w_o (\vec{S}_o + \vec{N}_o),$$

где

$$w_o = \frac{e^{-\frac{\left\|\vec{S}_j + \vec{n}_j - (\vec{S}_o + \vec{n}_o)\right\|^2}{h^2}}}{\sum_o e^{-\frac{\left\|\vec{S}_j + \vec{n}_j - (\vec{S}_o + \vec{n}_o)\right\|^2}{h^2}}},$$

$$\vec{S}_j{}' = \sum_{\forall o} w_o(\vec{S}_o + \vec{N}_o) \approx \sum_{o=1}^{N} \frac{1}{N}(\vec{S}_j + \vec{N}_o) \approx \vec{S}_j + \sum_{o=1}^{N} \frac{1}{N}\vec{N}_o .$$

Здесь N – количество векторов множества $O_{\vec{n}}$, h – параметр, влияющий на степень фильтрации (в работе $h = 1$). При этом $h^2 >> 2\left\|\vec{n}_{\max}\right\|^2$. Полагая, что $\{\vec{N}_o\}$ – набор случайных независимых векторов с постоянным вектором математического ожидания, равным $\vec{\mu}$, получим $\vec{S}_j{}' \approx \vec{S}_j + \vec{\mu}$. Подобный вывод подтверждается тем, что отсчеты АБГШ независимы друг от друга и математическое ожидание шума не зависит от времени. Таким образом, можно получить оценку $\hat{\vec{S}}_j$ незашумленного фрагмента $\vec{S}_j$ с помощью выражения:

$$\hat{\vec{S}}_j = \vec{S}_j{}' - \vec{\mu} \approx \vec{S}_j .$$

После проведения оценки всех фрагментов незашумленного отрезка сигнала, можно получить оценку всего отрезка. Затем, по оценкам всех отрезков сигнала, можно получить оценку и самого незашумленного речевого сигнала. Все это осуществляется с помощью процедуры восстановления сигнала с применением усреднения перекрывающихся данных.

На основе идеализированной модели сигнала и неравенства Рао-Крамера проведена оценка предельной эффективности предложенного метода. Установлено, что с применением метода можно добиться снижения дисперсии, а значит и снижения мощности шума, в среднем в n раз. Таким образом, выигрыш в отношении сигнал/шум (ОСШ) по сравнению с начальным ОСШ для участка речевого сигнала составит:

$$\Delta SNR = 10\lg(n),$$

где n – количество усредняемых фрагментов. На практике для речевого сигнала $n \le 20$, поэтому можно полагать, что $\Delta SNR \le 13$дБ.

Субъективная оценка позволяет сделать вывод, что предложенный алгоритм способен значительно снизить уровень шума, при этом улучшается разборчивость речи и не возникает артефактов, получивших название «музыкальный шум». Сравнение спектрограмм речевого сигнала до добавления шума, после его подавления, а также при обработке с помощью предложенного алгоритма и алгоритма, реализующего метод спектральных вычитаний, подтверждают сделанные ранее выводы.

Для объективной оценки речевого сигнала на выходе алгоритма шумоподавления используется сегментное отношение сигнал/шум (СегОСШ, Segmental SNR, SSNR). Результаты моделирования показывают, что при высоких уровнях шума предложенный алгоритм обладает большей эффективностью по сравнению с алгоритмом, реализующим метод спектральных вычитаний.

Наибольший выигрыш достигается при обработке участков речевого сигнала, отвечающих гласным звукам. Это объясняется, исходя из особенностей алгоритма и структуры речевого сигнала: гласные звуки содержат больше похожих фрагментов (имеют структуру, близкую к периодической), на поиске которых и основан

алгоритм нелокального усреднения. Предложенный алгоритм NLM увеличивает СегОСШ по сравнению с зашумленным сигналом на 10–12 дБ. Для сравнения, метод спектрального вычитания SS уступает предложенному методу на 1–1.5 дБ (рис. 7).

Ограничениями предложенного алгоритма можно считать то, что участки речевого сигнала, соответствующие шипящим звукам и глухим согласным, не удовлетворяют предложенной модели речевого сигнала, обладают низкой энергией и практически не содержат похожих фрагментов. В случае воздействия сильных внешних шумов на речевые сигналы (ОСШ 0 дБ), такие участки рассматриваются алгоритмом как шум и обнуляются. Для случая воздействия маломощных шумов (ОСШ>5дБ), подобные отрезки сигнала фактически не обрабатываются, поскольку число похожих фрагментов в них не превышает 4.

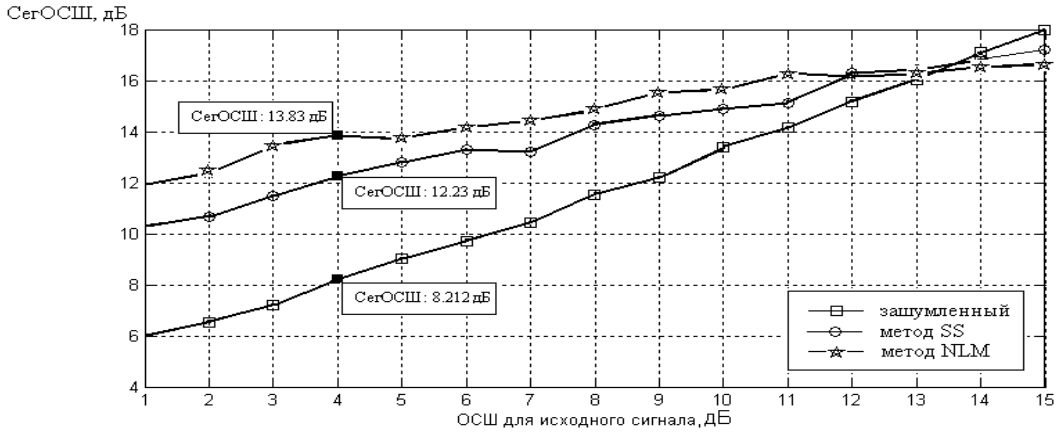

Рис. 7. Зависимость среднего значения сегментного ОСШ от ОСШ исходного сигнала

Также проведен анализ возможности применения разработанного алгоритма подавления шумов в речевых сигналах на этапе предобработки для систем распознавания голосовых команд. В ходе исследований использована дикторозависимая система распознавания 10 цифр («0»–«9») русского языка (мужской голос) при наличии АБГШ. Рассмотрено три варианта системы распознавания (рис. 8): без стадии предобработки; с использованием предобработки с помощью предложенного выше алгоритма шумоподавления методом нелокального усреднения; с использованием предобработки с целью шумоподавления методом спектрального вычитания.

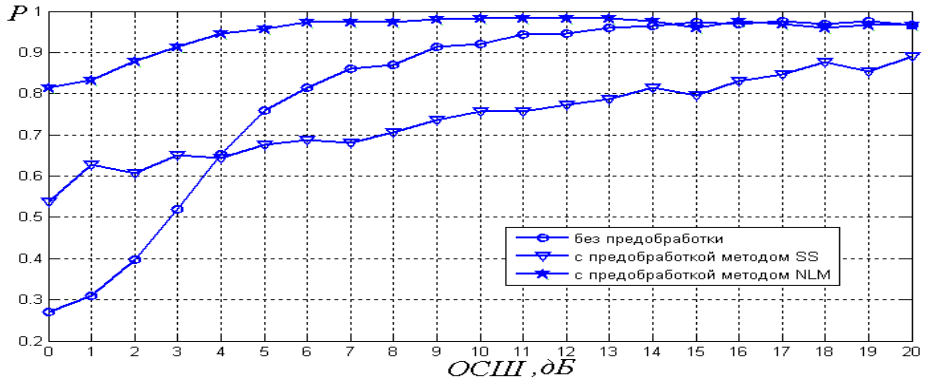

Рис. 8. Зависимость вероятности верного распознавания речевых команд на фоне шума от ОСШ

Из анализа зависимостей видно, что предложенный метод подавления шума позволяет повысить вероятность верного распознавания команд. Особенно заметен выигрыш для ОСШ 0–10 дБ. Например, вероятность правильного распознавания при стационарном шуме в 10 дБ составляет 98%. Демонстрируется также несостоятельность применения метода подавления шума с помощью алгоритма спектрального вычитания для решения такой задачи, поскольку в этом случае наблюдается снижение вероятности распознавания для ОСШ 4–20 дБ. При использовании подобного подхода выигрыш в распознавании достигается лишь для ОСШ менее 4 дБ. При этом вероятность верного распознавания не превышает 70%.

В качестве примера применения предложенных в работе алгоритмов выделения и предобработки речевых сигналов для систем распознавания разработана система голосового управления мобильным роботом Rovio, реализованная в пакете Matlab. Rovio является мобильным роботом с видеокамерой и обладает системой навигации Northstar – MINI GPS (Global Positioning System) (рис. 9).

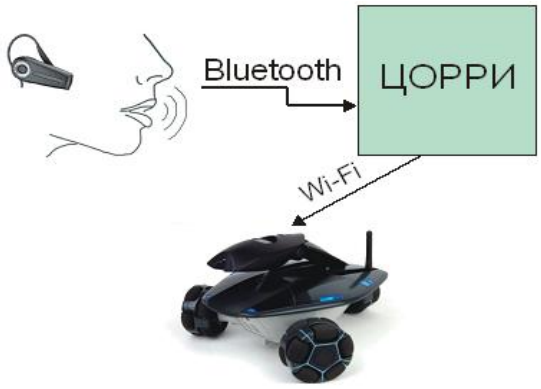

Рис. 9. Рассматриваемый способ голосового управления мобильным роботом

Разработанная система управления рассчитана на небольшое количество управляющих команд для голосового управления роботом и работает в реальном режиме времени. В набор из 8 команд входят навигационные команды управления: «вперед», «назад», «налево», «направо», «поворот» (робот осуществляет поворот вокруг оси в правую сторону на 40 градусов), «домой» (робот возвращается на платформу зарядки), а также команды управления положением видеокамеры «вверх», «вниз».

Голосовое управление роботом осуществляется посредством передачи речевого сигнала от беспроводной Bluetooth гарнитуры пользователя в центр обработки и распознавания речевой информации (ЦОРРИ), реализованного на базе компьютера с установленным пакетом Matlab. В ЦОРРИ производится выделение и предобработка команд, верификация диктора и последующее распознавание. Распознанная команда в виде управляющих сигналов передается мобильному роботу для выполнения определенного действия. Подобная система голосового управления основана на применении статистических методов классификации речевых сигналов и предполагает наличие этапов обучения.

Как известно, сегодня мировые лидеры в области создания систем распознавания речи (например, Google) представляют широкой общественности доступ к своим хорошо обученным системам распознавания через Интернет. Используя предложенные в работе алгоритмы автоматического выделения и предобработки команд, можно формировать «очищенный» речевой сигнал и

«отправлять» его для распознавания на сервер Google. В работе проведено исследование подобной схемы автоматического распознавания команд в условиях шумов для управления роботом. Распознавателю компании Google предлагалось распознать зашумленные БГШ (ОСШ = 10дБ) команды «налево», «направо», «вперед», «назад» (по 10 вариантов произнесения) и те же команды, обработанные методом нелокального усреднения. Эксперимент показывает, что точность распознавания во втором случае выше.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

На основании проведенных исследований в области цифровой обработки речевых сигналов в работе получены следующие результаты:

1. Представлен алгоритм вычисления новых информативных параметров для речевых сигналов – коэффициентов главных компонент мел-вейвлет-пакетных коэффициентов, произведена оценка их помехоустойчивости. Данные параметры являются более устойчивыми к шумам, чем спектрограммы и кепстральные коэффициенты, приведенные в мел-шкалу. Результаты проведенных исследований показывают эффективность использования параметров ГК МВП для описания различных классов звуковых колебаний (например, «речь» и «не речь»).
2. На основе использования ГК МВП и статистических моделей гауссовских смесей разработан алгоритм детектирования речевой активности. Алгоритм является обучаемым и дает возможность учитывать несколько типов фоновых шумов и несколько типов голосов дикторов одновременно для обеспечения требуемой надежности обнаружения речи. Предложенный алгоритм детектирования речевой активности при этом способен эффективно определять положение речевых и неречевых участков сигнала.
3. Предложен алгоритм выделения речевых команд из потока звуковых колебаний с использованием разработанного метода детектирования речевой активности.
4. Проведена оценка вероятности правильного выделения команд на фоне стационарных и нестационарных помех. Выполнено сравнение эффективности выделения команд с аналогичным алгоритмом, работающим на основе мелкепстральных информативных параметров. Оба алгоритма показывают хорошую надежность выделения команд, но применение параметров ГК МВП позволяет надежнее проводить выделение в случае нестационарного шума и на фоне шумов значительной интенсивности (ОСШ -5 дБ).
5. Разработан алгоритм шумоподавления в речевых сигналах на основе нелокального усреднения. Предложен метод поиска похожих фрагментов на интервалах стационарности речевого сигнала с помощью нахождения локальных минимумов евклидового расстояния между фрагментами опорного речевого сигнала. Опорный речевой сигнал при этом может быть получен из исходного зашумленного сигнала путем шумоподавления с помощью метода спектрального вычитания.
6. Проведена оценка качества шумоподавления в речевых сигналах с помощью предложенного алгоритма. При его использовании хорошо сохраняются все значимые детали спектра и отсутствуют нежелательные всплески, приводящие к

возникновению «музыкального шума», что характерно для методов фильтрации в спектральной области.

7. Проведен анализ возможности применения разработанного алгоритма подавления шумов на этапе предобработки в дикторозависимой системе распознавания цифр русского языка. Предложенная схема подавления шума позволяет повысить вероятность верного распознавания цифр. Особенно заметен выигрыш для ОСШ 3–12 дБ. Например, вероятность правильного распознавания при стационарном шуме в 10 дБ составляет 93%.

8. На основе предложенных алгоритмов выделения и предобработки команд разработана и реализована система голосового управления мобильным роботом.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в журналах из перечня ВАК

1. Новоселов С.А., Ульдинович С.В., Новиков А.Е., Веселов И.А. Классификация речевых команд с использованием аппарата скрытых марковских моделей // Проектирование и технология электронных средств. 2009. №1. С. 40–44.

2. Новоселов С.А., Савватин А.А., Приоров А.Л. Применение банков фильтров для построения системы защищенной передачи речевой информации // Электросвязь. 2011. №9. С. 48–51.

3. Новоселов С.А., Савватин А.И., Приоров А.Л. Использование цифровых вейвлет-фильтров в задаче построения защищенного канала передачи речевой информации // Проектирование и технология электронных средств. 2009. №2. С. 39–43.

Статьи в рецензируемых журналах

4. Новоселов С.А., Ульдинович С.В. Распознавание изолированных фонем на основе согласованных вейвлет-фильтров и нейронной сети // Вестн. Яросл. гос. ун-та. Сер. Физика. Радиотехника. Связь. 2008. С. 152–155.

5. Новоселов С.А., Веселов И.А., Новиков А.Е. Метод иерархической классификации речевых команд на основе скрытых марковских моделей // Вестн. Яросл. гос. ун-та. Сер. Физика. Радиотехника. Связь. 2009. С. 81–86.

Статьи в сборниках статей

6. Новоселов С.А., Волохов В.А. Метод вейвлет-сжатия звука, учитывающий частотное маскирование // VI всерос. науч.-техн. конф. «Информационные технологии в электротехнике и электроэнергетике». Чебоксары, 2006. С. 346-347.

7. Новоселов С.А., Приоров А.Л. Согласованные одномерные вейвлет-фильтры в задаче распознавания речевых сигналов // Тр. LXII науч. сессии, посвященной Дню Радио. М., 2007. С. 160–161.

8. Новоселов С.А. Применение согласованных одномерных вейвлет-фильтров в задаче распознавания речевых сигналов // Докл. 9-й междунар. конф. «Цифровая обработка сигналов и ее применение». М., 2007. С. 147–149.

9. Новоселов С.А. Распознавание речевых сигналов с использованием вейвлет-преобразования // Тез. 60-й науч.-техн. конф. студентов и магистрантов. Ярославль: ЯГТУ, 2007. С. 62.

10. Новоселов С.А., Ульдинович С.В., Приоров А.Л. Распознавание фонем на основе согласованных вейвлет-фильтров // Докл. 10-й междунар. конф. «Цифровая обработка сигналов и ее применение». М., 2008. Т. 1. С. 242–245.

11. Новоселов С.А., Веселов И.А., Новиков А.Е., Топников А.И. Применение вейвлет-преобразования и скрытых марковских моделей в задаче распознавания речевых команд // Докл. 11-й междунар. конф. «Цифровая обработка сигналов и ее применение». М., 2009. Т. 1. С. 244–247.

12. Новоселов С.А., Максимов В.И., Кравцов С.А., Гречко Р.С. Алгоритм идентификации диктора с помощью метода динамического искажения времени и вейвлет-преобразования // Докл. 11-й междунар. конф. «Цифровая обработка сигналов и ее применение». М., 2009. Т. 1. С. 269–270.

13. Новоселов С.А., Веселов И.А., Новиков А.Е. Распознавание речевых команд с помощью скрытых марковских моделей на основе вейвлет-параметров сигналов // Тр. LXIV науч. сессии, посвященной Дню Радио. М., 2009. С. 210-212.

14. Новоселов С.А., Савватин А.И. Использование согласованных вейвлет-фильтров в задаче защиты речевой информации // Сб. матер. XVI междунар. науч.-техн. конф. «Радиолокация, навигация, связь». Воронеж, 2010. С. 388-396.

15. Новоселов С.А., Новиков А.Е., Веселов И.А., Топников А.И. Вейвлет-преобразование и скрытые марковские модели в задаче распознавания речевых команд // Матер. 16-й междунар. науч.-техн. конф. «Проблемы передачи и обработки информации в сетях и системах телекоммуникаций». Рязань, 2009. С. 124-125.

16. Новоселов С.А., Веселов И.А., Новиков А.Е. Расстояние между скрытыми марковскими моделями в задаче распознавания речевых команд // Тр. LXV науч. сессии, посвященной Дню Радио. М., 2009. С. 215-216.

17. Новоселов С.А., Веселов И.А., Новиков А.Е. Применение скрытых марковских моделей для схемы иерархической классификации речевых команд // Докл. 12-й междунар. конф. «Цифровая обработка сигналов и ее применение». М., 2010. Т. 1. С. 195–197.

18. Новоселов С.А., Топников А.И. Анализ независимых компонент в задаче разделения смесей речевых сигналов // Докл. 12-й междунар. конф. «Цифровая обработка сигналов и ее применение». М., 2010. Т. 1. С. 197–199.

19. Новоселов С.А., Топников А.И., Савватин А.И. Алгоритм шумоочистки речевых команд методом спектрального слежения // Докл. 13-й междунар. конф. «Цифровая обработка сигналов и её применение». М., 2011. Т. 2. С. 224-226.

20. Новоселов С.А., Приоров А.Л. Метод удаления шума из речевых команд методом спектрального слежения // Сб. матер. всерос. конф. «Радиоэлектронные средства передачи и приема сигналов и визуализации информации». Таганрог, 2011. С. 104–107.

21. Новоселов С.А., Топников А.И. Потенциальная эффективность подавления шума в речевых сигналах методом нелокального усреднения // Сб. тр. междунар. науч.-практ. конф. студентов и молодых ученых «Молодежь и наука: модернизация и инновационное развитие страны». Пенза, 2011. Ч. 2. С. 292–295.

Свидетельство о государственной регистрации программ для ЭВМ

22. Новоселов С.А., Топников А.И., Савватин А.И., Приоров А.Л. Научно-исследовательская программа для подавления шума в речевых сигналах Yar_SpeechCleaner // Свидетельство о регистрации в Реестре программ для ЭВМ № 2011618562 от 31.10.2011.

Подписано в печать 28.11.11.
Формат 60×84 1/16. Тираж 100 экз.

Отпечатано на ризографе
Ярославский государственный университет
150000 Ярославль, ул. Советская, 14.