

На правах рукописи



БАЛАНДИН Иван Васильевич

**АЛГОРИТМЫ ПОИСКА КЛЮЧЕВЫХ СЛОВ
В РАДИОТЕХНИЧЕСКИХ УСТРОЙСТВАХ ОБРАБОТКИ
РЕЧЕВОЙ ИНФОРМАЦИИ, УСТОЙЧИВЫЕ К
ВОЗДЕЙСТВИЮ МЕШАЮЩИХ ФАКТОРОВ**

Специальность: 05.12.04 -

«Радиотехника, в том числе системы и устройства телевидения»

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата технических наук

Рязань 2013

Работа выполнена в ФГБОУ ВПО
«Рязанский государственный радиотехнический университет»

- Научный руководитель - заслуженный работник Высшей школы РФ,
доктор технических наук, профессор
Кириллов Сергей Николаевич
- Официальные оппоненты - Паршин Юрий Николаевич,
доктор технических наук, профессор,
ФГБОУ ВПО «Рязанский государственный
радиотехнический университет», заведующий
кафедрой Радиотехнических устройств
- Шустиков Олег Евгеньевич,
кандидат технических наук, Рязанское региональное
отделение центрального филиала ОАО «Мегафон»,
руководитель по инфраструктуре
- Ведущая организация - ГКОУ ВПО «Академия Федеральной службы
охраны Российской Федерации», г.Орел

Защита состоится « 22 » ноября 2013 г. в 12 часов на заседании
диссертационного совета Д 212.211.04 в ФГБОУ ВПО «Рязанский
государственный радиотехнический университет» по адресу: 390005, г. Рязань,
ул. Гагарина, д. 59/1.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВПО
«Рязанский государственный радиотехнический университет».

Автореферат разослан « 7 » октября 2013 г.

Ученый секретарь
диссертационного совета



Овечкин Г.В.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Разработка алгоритмов поиска ключевых слов в потоке слитной речи является одной из наиболее сложных задач в области обработки речевых сигналов (РС) и требует использования особых методов повышения показателей качества радиотехнических устройств. Наиболее существенный вклад в развитие теории речеобразования и методов обработки, передачи и распознавания РС внесли работы М.А. Сапожкова, А.А. Пирогова, Т.К. Винцюка, М.В. Назарова, Ю.Н. Прохорова, В.Н. Сорокина, Л.В. Златоустовой, В.Г. Михайлова, Р.В. Гудонавичюса, Ю.К. Калининца, а также работы Дж.Д. Маркела, Л.Р. Рабинера, Р.В. Шафера, Дж.Д.Фланагана, Г. Фанта, А.Х. Грейя, У. Ли, Дж.Р. Доддингтона, С.Е.Левинсона и др.

Технологии поиска ключевых слов лежат в основе радиотехнических устройств голосового управления, интерактивного телевидения, аудио индексации, поиска речевой информации по образцу в мультимедиа-архивах, автоматического контроля речевых сообщений в системах безопасности и т.д. К алгоритмам автоматического поиска ключевых слов (ААПКС) предъявляют ряд специфических требований, выделяющих их в отдельный класс систем автоматического распознавания речи (АРР), такие как:

- малый объем словаря;
- работа в потоке спонтанной слитной речи;
- дикторонезависимость;
- обработка речи с выраженными дефектами;
- повышенные требования к использованию вычислительных и аппаратных ресурсов.

В ряде основных проблем при разработке алгоритмов первичной обработки РС и поиска ключевых слов наиболее актуальными являются:

1. Определение информативной системы акустических параметров первичных элементов речи, устойчивой к изменению голоса диктора и воздействию акустических помех;
2. Уменьшение влияния вариативности речи на показатели качества радиотехнических устройств обработки РС и поиска ключевых слов;
3. Повышение вычислительной эффективности алгоритмов формирования первичных признаков РС и поиска ключевых слов в интересах обеспечения дикторонезависимой обработки в реальном масштабе времени при заданном объеме словаря.

Впервые задача ААПКС была сформулирована в 40-е годы прошлого века, но разработанные алгоритмы обладали низкой эффективностью и надежностью. Несмотря на предъявляемые специфические требования, в настоящее время, задача поиска ключевых слов в основном решается системами АРР, также не обладающими заданной эффективностью. Известны различные методы АРР, но в последнее время основным стал метод сопоставления с эталоном на основе вероятностных моделей. Главным образом это связано с прогрессом в области электронных компонентов, в частности, с увеличением вычислительной мощности процессоров и объемов памяти. При этом нерешенными остались вопросы,

связанные с работой ААПКС в условиях акустических шумов, обеспечению работы в потоке слитной речи, а также адаптацией к изменению голоса диктора.

Для снижения вычислительных затрат при обработке речевой информации в радиотехнических устройствах могут использоваться алгоритмы вейвлет-анализа, позволяющие обеспечить дикторонезависимость алгоритмов и компактное хранение эталонов. Но особенности построения быстрых алгоритмов вейвлет анализа не позволяют использовать психофизические свойства слуховой системы человека, что значительно снижает качество формируемого пространства признаков в интересах построения ААПКС.

Поиск ключевых слов по методу сопоставления с эталоном определяет и необходимость решения задачи хранения образцов, которая усугубляется необходимостью увеличения их числа в целях обеспечения дикторонезависимости.

Кроме того, разработка ААПКС требует решения проблемы создания и обучения моделей ключевых слов, для решения которой не существует аналитических способов. Использование известных итеративных алгоритмов обуславливает необходимость обоснования оптимального критерия качества обучения.

Таким образом, актуальна задача разработки новых дикторонезависимых моделей и алгоритмов предварительной обработки РС, формирования признаков и решающих алгоритмов в интересах повышения надежности и эффективности радиотехнических устройств поиска ключевых слов в потоке речевой информации.

Цель работы. Основной целью диссертационной работы является разработка алгоритма автоматического дикторонезависимого поиска ключевых слов в потоке слитной речи, обеспечивающего повышение показателей надежности обнаружения при воздействии мешающих факторов.

Поставленная цель работы включает решение следующих задач:

- исследовать методы нормирования РС по амплитуде в целях повышения устойчивости ААПКС к изменению уровня РС;
- обосновать выбор базиса представления РС для обеспечения устойчивости алгоритмов обработки речевых сигналов ААПКС к воздействию мешающих факторов;
- предложить процедуру адаптации базиса представления РС в интересах повышения робастных свойств ААПКС к изменению голоса диктора;
- разработать алгоритм обработки РС, устойчивый к воздействию мешающих факторов, включающий оценку частоты основного тона (ОТ) и сегментацию непрерывного РС на речевые единицы;
- проанализировать возможность явного моделирования акустического окружения в целях обеспечения работы ААПКС в потоке слитной речи и уменьшения расхода вычислительных ресурсов;
- обосновать введение акустического контекста в алгоритм поиска ключевых слов в потоке слитной речи для повышения робастных свойств ААПКС;
- проанализировать процедуру явного моделирования изменчивости РС во временной области в интересах увеличения устойчивости ААПКС к изменению голоса диктора;

- определить возможности сокращения вычислительных затрат и уменьшения требуемого объема памяти для хранения эталонов в целях обеспечения дикторонезависимости;
- разработать автоматический алгоритм обучения предложенного ААПКС;
- проанализировать возможность аппаратной реализации ААПКС.

Методы проведения исследований. В работе использовались методы статистической радиотехники и математической статистики, распознавания образов, динамического программирования, кластерного анализа, вычислительной математики, а также новые достижения в области цифровой обработки речевой информации. Данные теоретические методы сочетались с экспериментальными исследованиями на основе имитационного моделирования разработанных алгоритмов.

Научная новизна. В рамках диссертационной работы были получены следующие новые научные результаты:

1. Разработан алгоритм формирования первичных признаков РС на основе непрерывного вейвлет преобразования (НВП) с адаптацией вейвлет фильтра (ВФ) к изменению голоса диктора на основе оценки частоты основного тона (ОТ), обеспечивающий повышение робастных свойств ААПКС в условиях воздействия мешающих факторов.

2. Разработан алгоритм адаптивной сегментации непрерывного РС, основанный на использовании модифицированной фонетической функции речи (ФФР) А.А.Пирогова, позволяющий уменьшить вероятность ложной сегментации, и тем самым снизить вычислительные затраты на работу ААПКС.

3. Предложен ААПКС на основе модифицированной гибридной модели восприятия речи, использующий акустический контекст и явное моделирование акустического окружения для обеспечения работы в потоке слитной речи и снижения вычислительных затрат на поиск границ ключевого слова.

Достоверность. Достоверность результатов и выводов полученных в диссертационной работе обеспечивается корректностью численных экспериментов, качественным и количественным сопоставлением с известными положениями теории обработки и распознавания РС.

Практическая ценность. Представленные в работе алгоритмы предварительной обработки и распознавания РС могут быть использованы в таких радиотехнических устройствах, как системы передачи речевой информации, системы интерактивного взаимодействия человека и ЭВМ, информационно-справочные системы, системы экономного хранения РС, медицинские системы диагностики, системы автоматической идентификации и верификации информанта по голосу, системы криминалистической фоноскопии, системы конфиденциального доступа и закрытия речевой информации, системы открытого образования для лиц с ограниченными возможностями здоровья и т.д. Результаты диссертационной работы нашли применение в действующей системе автоматического контроля безопасности связи войсковой части 67240 и в учебном процессе военной кафедры ФГБОУ ВПО «Комсомольский-на-Амуре государственный технический университет», что подтверждается соответствующими актами внедрения.

Основные положения, выносимые на защиту:

1 Алгоритм формирования вектора первичных признаков РС, включающий НВП РС с адаптацией ВФ к изменению голоса диктора и адаптивную сегментацию артикуляционных событий РС на основе модифицированной ФФР А.А.Пирогова, применение которого позволило уменьшить вероятность ложной сегментации с 0,15 до 0,08 при вероятности правильной сегментации 0,92 в случае ОСШ 20 дБ.

2 Процедура обеспечения работы ААПКС в потоке слитной речи с явным моделированием акустического окружения с помощью состояния заполнителя, позволяющая исключить из алгоритма этап поиска границ ключевого слова и уменьшить расход вычислительных ресурсов в шесть раз.

3 Модифицированная гибридная модель речевосприятия с использованием акустического контекста обеспечивающая увеличение вероятности правильного обнаружения на 0,05 и снижение вероятности ложного обнаружения на 0,02 по сравнению с известными системами. При этом средняя вероятность правильного обнаружения после оптимизации параметров составила 0,98, а средняя вероятность ложного обнаружения 0,05.

Апробация работы. Результаты работы докладывались на следующих научно-технических конференциях (НТК), семинарах и сессиях:

1. Научная сессия МИФИ -2008 г., 2009 г., г. Москва
2. 15-я МНТК «Проблемы передачи и обработки информации в сетях и системах телекоммуникаций». 2008 г., г. Рязань.
3. Всероссийская НТК "Биотехнологические, медицинские и экологические системы и комплексы. Биомедсистемы -2009". г. Рязань, 2009.
4. 14-я, 15-я, 16-я Всероссийская НТК студентов, молодых ученых и специалистов "Новые информационные технологии в научных исследованиях и образовании". 2009, 2010, 2011 гг., г. Рязань.

Публикации. По теме диссертации опубликовано 11 работ. 2-е статьи в научно-технических журналах рекомендованных ВАК и 9 тезисов докладов на конференциях.

Структура и объем работы. Диссертационная работа состоит из введения, трех глав, заключения, списка литературы из 145 наименований и 3-х приложений. Диссертация содержит 130 стр. основного текста, 11 таблиц и 42 рисунка.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность выбранной темы, определены цель и решаемые в работе задачи. Изложены новые научные результаты, полученные в работе, показаны ее практическая ценность и апробация. Сформулированы основные положения, выносимые на защиту.

В первой главе разработаны алгоритмы первичной обработки речевого сигнала устойчивые к изменению голоса диктора и воздействию мешающих факторов.

Предложено для обеспечения постоянного отношения сигнал-шум (ОСШ) квантования в широком диапазоне дисперсий РС, снижения вычислительных затрат связанных с флуктуациями шума пауз и обеспечения робастных свойств алгоритма

использовать на входе аналогово-цифрового преобразователя адаптивный усилитель с каналом обнаружителя пауз РС, на основе подсчета числа переходов через ноль. В качестве порогового уровня использовался управляющий параметр

$$g_l(n) = \alpha_l g_l(n-1) + |s(n-1)|, \quad (1)$$

где α_l - коэффициент импульсной характеристики.

Адаптация параметров усилителя производилась во время обнаруженной речевой активности.

Для сохранения временной структуры РС, учитывая постоянную времени слуха, экспериментально определены значения коэффициента импульсной характеристики $\alpha_l = 0,95$. Показано, что при ОСШ 20 дБ предложенный алгоритм обеспечивает выигрыш в динамическом диапазоне 10 дБ и среднюю ошибку выделения участков речевой активности не более 2,5 мс.

В целях формирования компактного пространства признаков, устойчивого к воздействию мешающих факторов при выборе представления РС рассматривалась возможность использования психофизических свойств слуха, учтенных в Mel-частотной шкале,.

Эффективность формируемого пространства признаков оценивалась с помощью комбинированного критерия качества

$$F_a = \alpha_F K_a + (1 - \alpha_F) K_{stat} L / N_{кадр}, \quad (2)$$

где K_{stat} - статистический и K_a - артефактный критерии качества; $\alpha_F = \overline{0,1}$ - параметр комбинированного критерия; L - размерность пространства признаков; $N_{кадр}$ - размер кадра преобразования.

Критерии качества определялись из выражений

$$K_{stat} = \bar{r} / \bar{d}; \quad (3)$$

$$K_a = r_{max} / d_{min}, \quad (4)$$

где $\bar{d}$ - среднее расстояние между кластерами; $\bar{r}$ - средний радиус кластеров; d_{min} - минимальное расстояние между центрами кластеров; r_{max} - максимальный радиус кластера.

Экспериментально показано, что оптимальное, по критерию минимума вероятности ошибки классификации акустических образов, значение параметра комбинированного критерия $\alpha_F = 0,51$. На рисунке 1 представлены результаты исследования устойчивости пространства признаков, сформированного в базисах НВП, вейвлет-пакетного разложения (ВПР), оконного преобразования Фурье (ОПФ) и кепстра при размере кадра $N_{кадр} = 128$ отсчетов, к воздействию акустических шумов в виде гауссовского шума.

С целью сокращения вычислительных затрат выбор базового вейвлета для формирования первичных признаков РС осуществлялся из условий формирования требуемых частотных полос и минимизации площади частотно-временного окна:

$$S_{x\omega} = 4\Delta_x \Delta_\omega, \quad (5)$$

где Δ_x , Δ_ω - радиусы частотно-временного окна, соответственно во временной и

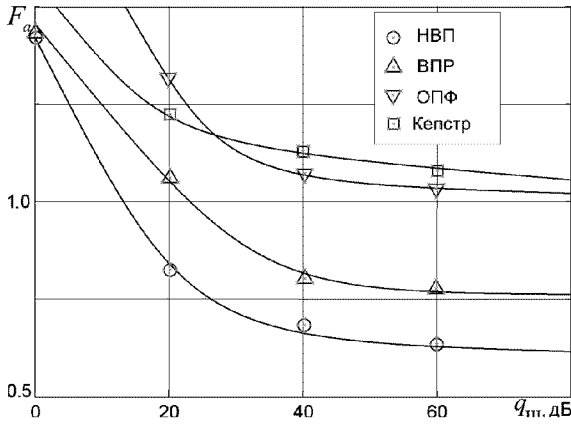


Рисунок 1

В гармонической модели РС импульсы ОТ являются модулирующим колебанием:

$$s(t) = M \sin(2\pi F_0 t) \sum_{k=0}^{\infty} A_k \cos(2\pi f_k t), \quad (7)$$

где M - индекс модуляции; F_0 - частота ОТ; A_k, f_k - амплитуда и частота k -й гармоники соответственно.

При этом формантные частоты, являющиеся гармониками ОТ, кратны частоте ОТ

$$f_k = kF_0, k = \overline{2, \infty}. \quad (8)$$

Предложена процедура нормализации РС в частотной области методом адаптации масштабов анализирующей ВФ НВП по изменению оценки частоты ОТ.

Масштабные коэффициенты ВФ i -го периода вокализованного участка определяются из выражения

$$a_k(i) = K_{OT}(i) a_k, k = \overline{1, N_a}, \quad (9)$$

где $N_a=14$ - число масштабов разложения; $K_{OT}(i)$ - коэффициент адаптации i -го периода вокализованного участка, определяемый из оценок частоты ОТ:

$$K_{OT}(i) = F_0(i) / F_0(1), \quad (10)$$

где $F_0(1), F_0(i)$ -оценки частоты ОТ 1-го и i -го периодов вокализованного участка РС соответственно.

На рисунке 2 представлены результаты сравнительных исследований качества признакового пространства для адаптивного и неадаптивного алгоритмов при воздействии гауссовского шума.

Алгоритм оценки частоты ОТ состоит из блока НВП, выделяющего импульсы ОТ из абсолютных значений отсчетов РС $|s(n)|$ и вычислителя, который определяет сдвиги ВФ, соответствующие переходам через ноль коэффициентов НВП $W_\psi(a, b)$ по всем масштабам. Расчетным методом определено число масштабов разложения $N_a = 4$, кратность изменения масштабов $k_a = 1,61$ и значение добротности ВФ $Q_\omega = 2,13$.

частотной областях.

Показано, что наиболее полно данным критериям соответствует вейвлет Морле. Для формирования 14 критических полос, перекрывающих полосу частот 300...3625 Гц, рассчитаны кратность изменения масштабов НВП $k_a = 1,195$, значение первого масштаба преобразования $a_1 = 0,0063$. Остальные масштабы рассчитывались из выражения

$$a_k = a_1 k_a^{k-1}, k = \overline{2, N_a}, \quad (6)$$

где $N_a=14$ – число масштабов разложения.

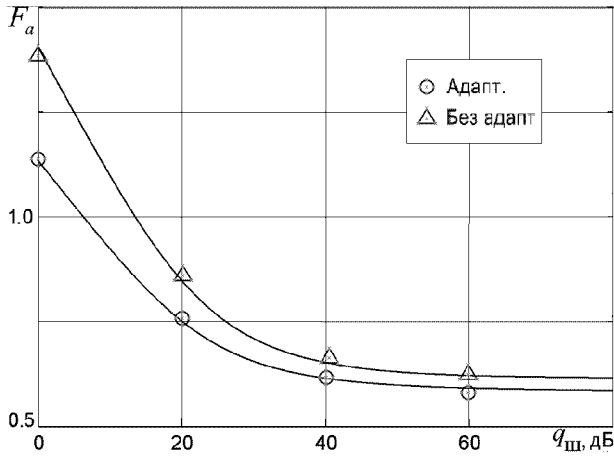


Рисунок 2

Выражение для оценки i -й границы периода ОТ:

$$b(i) = \bigcap_{k=1}^{N_a} (\text{sign}[W_\psi(a_k, b-1)] < \text{sign}[W_\psi(a_k, b)]), \quad (11)$$

где $W_\psi(a_k, b)$ - коэффициент НВП масштаба a_k , сдвига b .

Оценка длительности i -го периода ОТ производилась по выражению

$$t_0(i) = b(i) - b(i-1), k=1, \overline{N_a}. \quad (12)$$

Математическое ожидание периода ОТ j -го вокализованного участка РС определялось по формуле

$$T_0(j) = \frac{1}{n-k} \sum_{i=k}^n t_0(i), \quad (13)$$

а оценка частоты ОТ j -го вокализованного участка РС

$$F_0(j) = F_s / T_0(j). \quad (14)$$

Для уменьшения вероятности ложной сегментации предложен двухканальный алгоритм сегментации непрерывного РС на акустически однородные участки речи, использующий как статические, так и динамические параметры речи.

Сегментирующая функция на основе модифицированной ФФР А.А. Пирогова:

$$W(b) = \sum_{k=1}^{N_a} \ln([W_\psi(a_k, b)^2] / [W_\psi(a_k, b-1)W_\psi(a_k, b+1)]). \quad (15)$$

Решение о наличии артикулярного события принималось при выполнении условия:

$$W(b) > \beta_w W^0(b), \quad (16)$$

где $W^0(q)$ - пороговая функция; β_w - коэффициент, определяющий риск потерь обнаружения границ сегментов.

Пороговая функция вычислялась рекуррентно из выражения

$$W^0(q) = \alpha_w W^0(b-1) + W(b), \quad (17)$$

где $\alpha_w \leq 1$ - параметр адаптации.

Второй канал, использующий статические параметры РС, выделял межфонемные паузы в результате сравнения логарифма энергии коэффициентов НВП

$$E(b) = 10 \lg \sum_k W_\psi^2(a_k, b), \quad (18)$$

Исходя из требования минимизации площади частотно-временного окна, в целях уменьшения вычислительных затрат, и условия наличия центрального максимума ВФ для выделения формы импульса ОТ, выбрана ВФ Гаусса 6-го порядка, для которой рассчитаны масштабные коэффициенты при частоте дискретизации $F_s = 8000$ Гц:

$$\begin{aligned} a_1 &= 1,1 \cdot 10^{-3}; & a_2 &= 6,83 \cdot 10^{-4}; \\ a_3 &= 4,24 \cdot 10^{-4}; & a_4 &= 2,63 \cdot 10^{-4}. \end{aligned}$$

с адаптивным порогом E_{nor} , учитывающим наличие постоянной времени слухового восприятия τ_0 , составляющей в среднем 150...200 мс, вычисляемым из условия:

$$E_{nor}(b) = \begin{cases} E(b), E_{nor}(b-1) \leq E(b): \\ E_{nor}(b-1) - 10/(\tau_0 F_s), E_{nor}(b-1) > E(b). \end{cases} \quad (19)$$

Структурная схема алгоритма сегментации РС на АОУ представлена на рисунке 3.

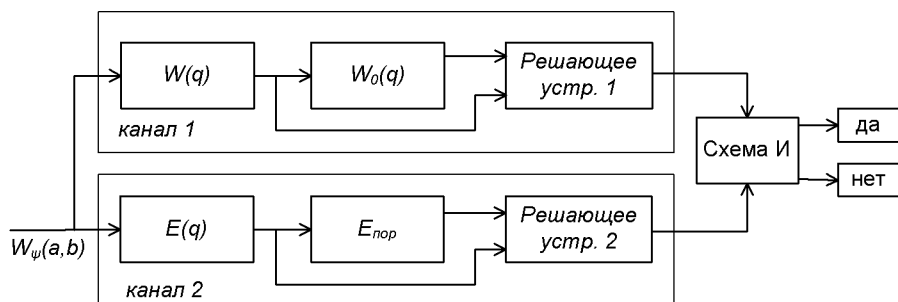


Рисунок 3

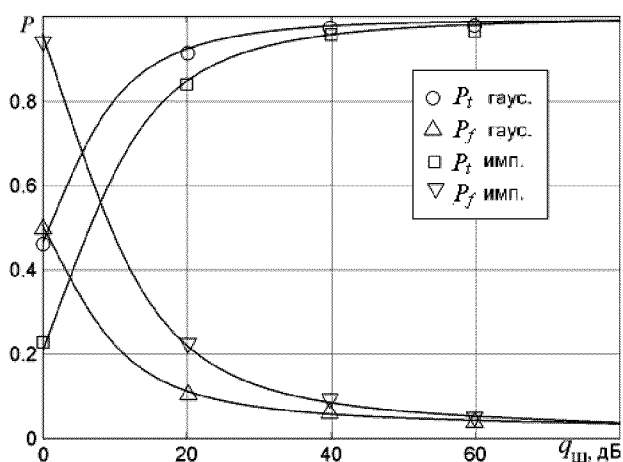


Рисунок 4

На рисунке 4 представлены зависимости правильной сегментации P_t и ложной сегментации P_f от ОШС для гауссовского шума и импульсной помехи при оптимальных значениях параметров сегментирующей функции $\alpha_w=0,95$, $\beta_w=0,04$, определенных экспериментально.

Во второй главе обоснован дикторонезависимый алгоритм автоматического поиска ключевых слов в потоке слитной речи на основе модифицированной гибридной модели восприятия речи.

Сформулирована задача поиска ключевых слов как задача классификации образов по критерию максимума апостериорной вероятности

$$\hat{Q} = \arg \max_Q P(Q | H), \quad (20)$$

где $\{Q\}^T = \{q_1, q_2, \dots, q_T\}$ последовательности состояний генерируемых источником; $\{H\}^K = \{h_1, h_2, \dots, h_K\}$ - последовательность символов наблюдения; q_t - состояние источника в момент времени $t = \overline{1, T}$; h_k - символ наблюдения поступивший в момент времени $k = \overline{1, K}$.

Анализ известных алгоритмов АРР показал, что в целях реализации ААПКС наибольшим потенциалом обладает гибридная модель восприятия речи, в которой для моделирования временных зависимостей используется скрытая марковская модель (СММ), а искусственная нейронная сеть (ИНС) решает задачу классификации акустических образов.

Для построения моделей ключевых слов, учитывая особенности РС, выбрана лево-правая структура СММ, в которой запрещены обратные переходы. Представлено множество параметров описывающих СММ z -го ключевого слова

$$\lambda_z = \{A_{HMM}, B_{HMM}, \Pi\}, \quad (21)$$

где $A_{HMM} = \{a_{ij}\}$, $a_{ij} = P[q_t = s_j | q_{t-1} = s_i]$, $i, j = \overline{1, U}$ - распределение вероятностей переходов между состояниями; $B_{HMM} = \{b_j(y_k)\}$, $b_j(y_k) = P[h_t = y_k | q_t = s_j]$, $j = \overline{1, U}$, $k = \overline{1, V}$ - множество распределений вероятностей появления символов наблюдения y_k в состоянии s_j ; $\Pi = \{\pi_i\}$, $\pi_i = P[q_1 = s_i]$, $i = \overline{1, U}$ - начальное распределение вероятностей состояний; $\{S\}^U = \{s_1, s_2, \dots, s_U\}$ - множество состояний модели; U - число возможных состояний модели; $\{Y\}^V = \{y_1, y_2, \dots, y_V\}$ - множество символов наблюдения, которые могут порождаться моделью; V - количество возможных символов наблюдения модели.

С учетом правила Байеса, и ограничений СММ решающее правило ААПКС (21) запишется в виде

$$\hat{z} = \arg \max_z P(H | \lambda_z). \quad (22)$$

Предложено для обеспечения работы ААПКС в потоке слитной речи модифицировать гибридную модель восприятия речи, включив в СММ ключевого слова состояние-заполнитель для явного моделирования акустического окружения.

На рисунке 5 представлена СММ ключевого слова «Клад» с состоянием-заполнителем s_0 , которому соответствуют символы наблюдения y_8, y_9 .

Оценка вероятности последовательности наблюдений производилась по выражению

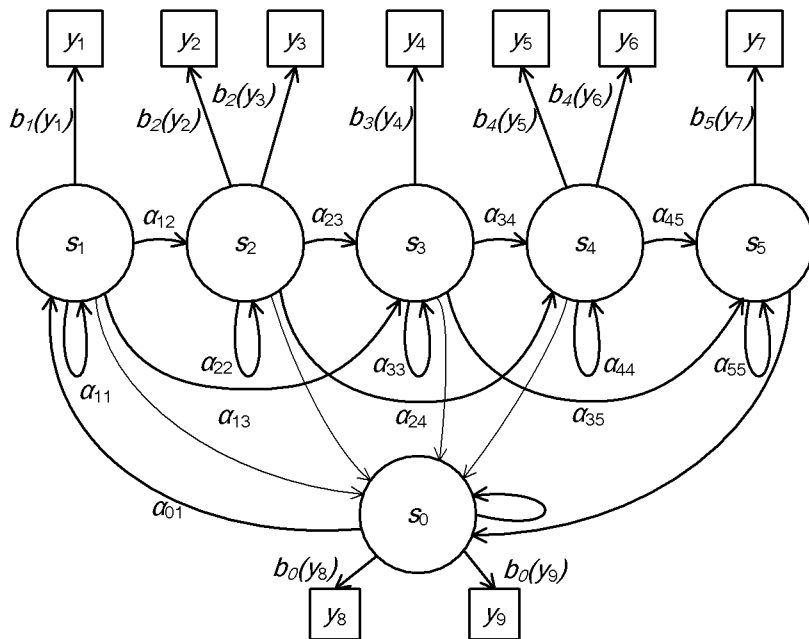


Рисунок 5

$$P(H | \lambda_z) = \sum_{j=1}^U \alpha_T(j), \quad (23)$$

где $\alpha_t(j) = P(q_t = s_j, H_1^t | \lambda_z)$, $t = \overline{1, T}$ - прямая переменная, вычисляемая рекуррентно алгоритмом прямого - обратного хода.

При появлении на входе системы символов наблюдения $h_t = y_8$ или $h_t = y_9$, соответствующих акустическому окружению, значение прямой переменной $\alpha_t(j)$, $j = \overline{1, U}$ станет равной нулю и система перейдет в

исходное состояние s_0 . Так как, в соответствии с выбранной лево-правой моделью, переход из состояния s_0 возможен только в состояние s_1 , отсутствует необходимость в дополнительном определении границ ключевого слова.

В таблице 1 представлены результаты сравнительных исследований предложенной модификации гибридной системы ААПКС с состоянием заполнителем и известных систем,

Таблица 1

Метод	Ошибка определения границ, %	Время работы алгоритма, мс
Предлож.алг.	10	2,5
Алг.Витерби	5	15
МОС	22	10

с поиском алгоритмом Витерби и моделями обобщающих слов (МОС) на наборе из 20 ключевых слов.

Наличие упрощающего предположения СММ о статистической независимости наблюдений

$$P(h_t | Q_1^T, H_1^T) = P(h_t | q_t) \quad (24)$$

не позволило уменьшить число ошибок обнаружения.

Предложено для преодоления этого ограничения, ввести в ААПКС акустический контекст в виде прямой переменной рассчитанной в предыдущий момент времени $u_t = \{\alpha_{t-1}(1), \dots, \alpha_{t-1}(U)\}$.

Оценка условной вероятности СММ состояния s_j на выходе ИНС акустического моделирования при поступлении символа наблюдения h_t определялась выражением

$$s_j^{MII}(y_k) = P(s_j | h_t = y_k, q_{t-1}, H_1^{t-1}), k = \overline{1, K}, j = \overline{1, U}. \quad (25)$$

Тогда предположение о статистической независимости наблюдений (24) запишется в виде:

$$P(h_t | Q_1^T, H_1^T) = P(h_t | q_t, H_1^{t-1}). \quad (26)$$

Что позволяет частично преодолеть недостаток СММ, связанный с предположением о статистической независимости наблюдений.

В таблице 2 приведены результаты исследований гибридных систем АРР с различным способом введения акустического контекста в виде вероятностей правильного P_t и

Таблица 2

Способ	P_t	P_f
Без контекста	0,82	0,16
С использованием задержки	0,92	0,08
Предл. способ	0,94	0,058

ложного обнаружения P_f .

Предложено для увеличения точности моделирования временных последовательностей в целях снижения ошибки поиска ключевых слов ввести в СММ дополнительную информацию о

длительности АОУ τ_t на интервале времени t в виде совместной вероятности текущего наблюдения и его длительности $P(h_t, \tau_t | q_t)$.

Для оценки совместной вероятности в гибридной модели на вход ИНС акустического моделирования подавался совместный вектор $\{h_t, \tau_t\}$.

Предложено для снижения объема обучающей выборки для ИНС акустического моделирования, используя теорему умножения вероятностей и

приняв упрощающее допущение о независимости вектора наблюдения h_t и длительности наблюдения τ_t , преобразовать выражение совместной вероятности

$$P(h_t, \tau_t | q_t) \approx P(h_t | q_t)P(\tau_t | q_t). \quad (27)$$

где $P(\tau_t | q_t)$ - условная вероятность длительности наблюдения τ_t .

Выразив условную вероятность $P(\tau_t | q_t)$ в терминах множества состояний модели $\{S\}^U$

$$P(\tau_t | q_t) \rightarrow P(\tau_t | q_t = s_j) = p_j(\tau_t), j = \overline{1, U}, \quad (28)$$

где $p_j(\tau_t)$ - плотность вероятности длительности пребывания СММ в состоянии s_j в течение времени τ_t .

Тогда выражение для расчета прямой переменной в алгоритме прямого-обратного хода запишется в виде

$$\alpha_{t+1}(j) = b_j(h_{t+1})p_j(\tau_{t+1})\sum_{i=1}^U \alpha_t(i)a_{ij}, t = \overline{1, T-1}, j = \overline{1, U}. \quad (29)$$

Отсюда следует выражение для оценки правдоподобия совместной последовательности наблюдений и длительностей наблюдений

$$P(H, T | w_z, \lambda_z) = \sum_{i=1}^U \alpha_T(i). \quad (30)$$

Для параметрического описания функций плотности вероятности длительности состояний СММ, наиболее точно описывающего временные особенности РС, выбрано нормальное распределение, с математическим ожиданием μ_j и дисперсией σ_j^2 , $j = \overline{1, U}$. Распределение моделировалось однослойной ИНС с радиальной функцией активации. Параметры распределения оценивались при статистическом анализе обучающей выборки по сформированной СММ и сохранялись в весовых коэффициентах ИНС.

Структурная схема предложенной ААПКС на основе модифицированной гибридной модели речевосприятия представлена на рисунке 6.

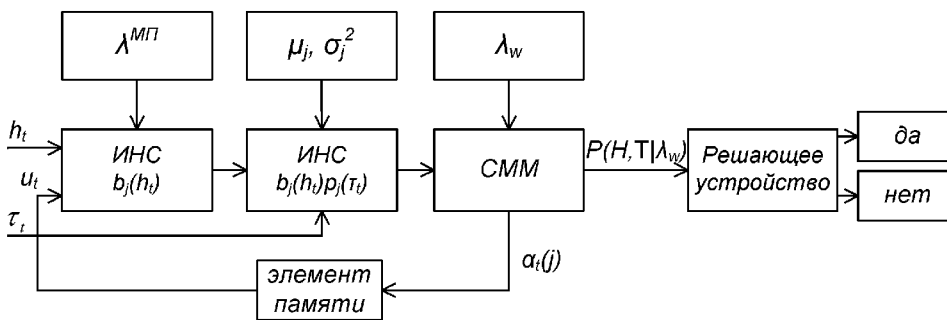


Рисунок 6

Обосновано применение векторного квантователя (ВК), на самоорганизующейся ИНС Кохонена для уменьшения размерности признакового пространства в целях сокращения требований к памяти для обеспечения дикторонезависимости.

Весовые коэффициенты ИНС содержат координаты кодовых векторов,

Предложено в целях повышения качества ВК и снижения вычислительных затрат на обучение ВК использовать предварительную кластеризацию обучающей выборки методом иерархической группировки Варда.

В таблице 3 представлены значения средней ошибки квантования D_l для различных способов задания начальных значений весов ИНС.

Таблица 3

Метод инициализации ИНС	$D_l, \times 10^{-3}$, при L		
	64	128	256
Случ. значениями	0,72	0,61	0,59
Средними значениями по обучающей выборке	0,84	0,72	0,66
Предв. кластеризацией обучающей выборки	0,51	0,43	0,41

Для оценки качества кластеризации применялся

комбинированный критерий качества

$$F_K = \alpha_k K_a + (1 - \alpha_k) K_{stat}, \quad (31)$$

где $\alpha_k = 0,1$ - параметр комбинированного критерия.

На рисунке 7 представлены зависимости комбинированного критерия качества F_K ВК от размера кодовой книги L при различных

уровнях гауссовского шума $q_{ш}$ и оптимальном, по критерию минимума вероятности ошибки квантования, значении параметра $\alpha_k = 0,42$, определенном экспериментально.

Для кодирования номеров кодовых векторов использовался слой Гроссберга. В результате размерность пространства признаков была уменьшена с 14 непрерывных координат до 7 двоичных разрядов.

В третьей главе рассмотрены вопросы обучения предложенного ААПКС и аппаратной реализации дикторонезависимой системы автоматического поиска ключевых слов в потоке слитной речи на его основе.

Предложено для повышения эффективности начальной оценки параметров СММ, уменьшения числа состояний, исключения ручной разметки обучающей выборки на акустические образы использовать 2-е кодовых книги ВК - обучающую W_o^K и полную W_l^K .

Оценка параметров СММ λ_z начиналась с определения числа состояний модели. Используя свойства принятой лево-правой модели СММ начальное число состояний U принималось равным числу символов наблюдения самой длинной обучающей последовательности. Кодовые векторы, использованные при ВК этой последовательности, составляли обучающую кодовую книгу W_o^K , далее использовавшуюся для формирования последовательности наблюдений очередного обучающего вектора и поиска оптимальной последовательности состояний алгоритмом Витерби.

Расчет эмиссионных вероятностей производился по полной кодовой книге W_l^K после определения оптимальной последовательности состояний каждого обучающего вектора.

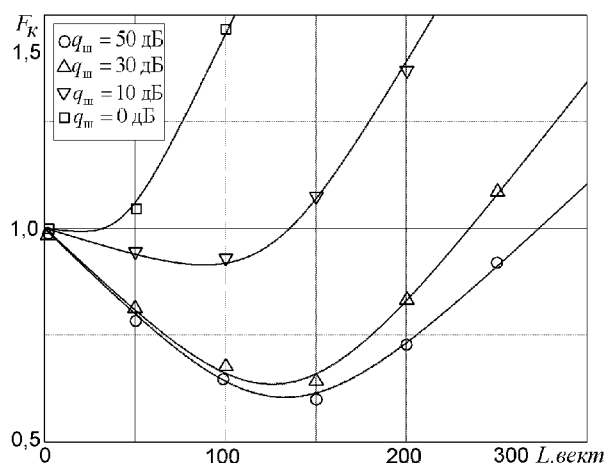


Рисунок 7

На рисунке 8 представлены результаты исследования эффективности методов оценки параметров СММ от величины обучающей выборки для предложенного

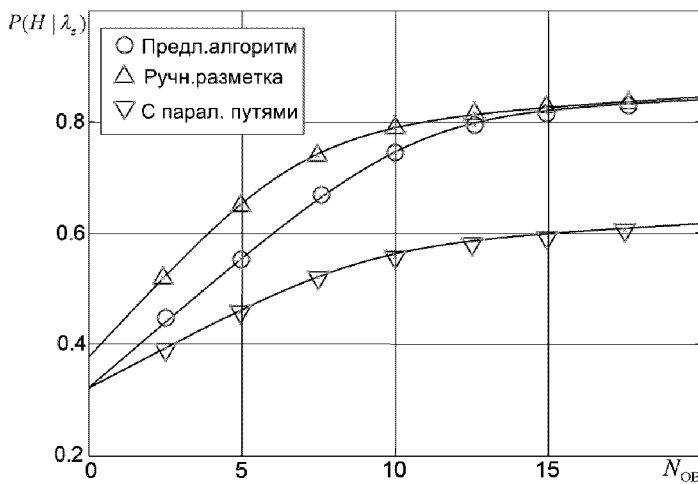


Рисунок 8

алгоритма, метода ручной разбивки обучающих векторов на состояния, метода автоматического формирования СММ с параллельными путями.

Переоценка значений параметров СММ с целью локальной максимизации вероятности $P(H | \lambda_z)$ и приближения к глобальному максимуму функции правдоподобия, выполнялась с помощью итеративного ЕМ - алгоритма.

Предложено для проверки условия остановки алгоритма использовать целевую функцию на основе информационной меры Кульбака — Лейблера:

$$I_n(\lambda_z, \bar{\lambda}_z) = -\sum_H P(H | \lambda_z) \ln[P(H | \bar{\lambda}_z)], \quad (32)$$

где $\bar{\lambda}_z$ - переоцененные параметры СММ.

Условие остановки ЕМ-алгоритма имеет вид:

$$I_{n-1}(\lambda_z, \bar{\lambda}_z) - I_n(\lambda_z, \bar{\lambda}_z) < \nu^I, \quad (33)$$

где ν^I - целевой порог, оптимальное, по критерию максимума вероятности $P(H | \lambda_z)$, значение которого было найдено экспериментально $\nu^I = 1,7 \times 10^{-6}$. Это значение было достигнуто при 84 итерациях оптимизации параметров СММ.

Предложено в алгоритме оптимизации параметров разработанного ААПКС методом обратного распространения ошибки использовать для вычисления прямой переменной рекуррентную ИНС, в которой веса нейронов входного слоя содержат значения транзитивных вероятностей СММ a_{ij} . Применение такой структуры ИНС позволило получать переоцененные значения a_{ij} .

Для определения целевого вектора, необходимого для работы алгоритма обратного распространения ошибки, использовался вспомогательный вычислитель обратной переменной $\beta_i(j)$. Промежуточные значения целевого вектора, с выполнением условия нормирования находились из выражения:

$$\alpha_i(j) = \beta_i(j) / \sum_{i=0}^U \beta_i(i), j = \overline{0, U}. \quad (34)$$

На рисунке 9 представлена структурная схема предложенного алгоритма оптимизации параметров системы ААПКС.

На рисунке 10 представлены зависимости вероятности правильного P_t и ложного P_f обнаружения ключевого слова от ОСШ для гауссовского шума при оптимальном, по критерию минимума ошибок обнаружения, значении порога обнаружения $\rho = 0,91$ полученные на тестовой выборке из 20 реализаций

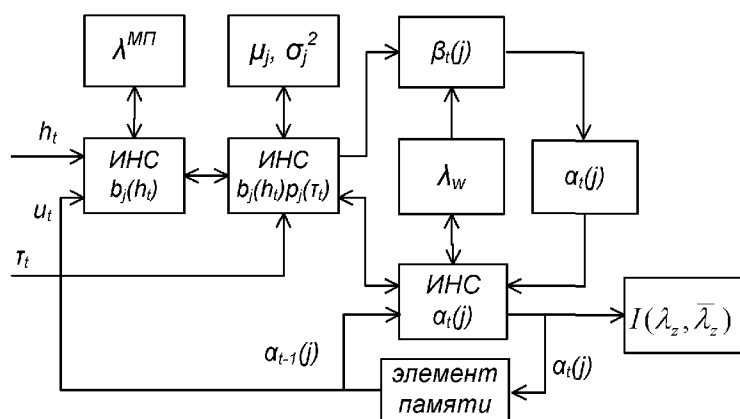


Рисунок 9

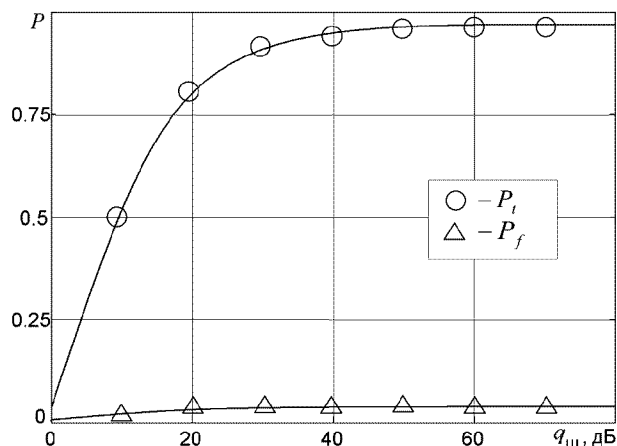


Рисунок 10

предложений, содержащих ключевое слово.

Произведен расчет вычислительной сложности аппаратной реализации ААПКС характеризуемой количеством используемых ячеек памяти и числом вычислительных операций, выполняемых за единицу времени (таблица 4). Показано, что наибольшую вычислительную сложность представляют алгоритмы предварительной обработки РС и формирования вектора признаков, а именно адаптивный алгоритм НВП.

Таблица 4

объем памяти, слов	ОЗУ 2374	ПЗУ 20006
Кол-во опер./с	умножения 2090592,32	сложения 2090592,32

Произведен выбор цифрового сигнального процессора по комбинированному критерию, учитывающему технические и экономические показатели качества, которому соответствовала микросхема TMS320C5532 фирмы Texas Instruments.

В заключении приведены основные научные и практические результаты диссертационной работы:

1. Обоснован алгоритм адаптации параметров квантователя для уменьшения вариативности РС по уровню. Применение этого алгоритма обеспечило выигрыш в динамическом диапазоне 10 дБ при ОСШ 20 дБ и среднюю ошибку выделения участков речевой активности не более 2,5 мс.

2. Получено компактное представление первичных признаков ключевых слов в базе кратномасштабного НВП с ВФ Морле, применение предложенного алгоритма адаптации параметров ВФ к изменению голоса диктора на основе оценки частоты ОТ, позволило уменьшить значение комбинированного показателя качества на 0,2 по сравнению с другими базами, достигнув значения 0,75 при ОСШ 20 дБ.

3. Предложен алгоритм предварительной обработки РС, включающий:

а) алгоритм оценки мгновенной частоты ОТ на основе НВП, который обеспечил величину нормированного среднеквадратического отклонения (СКО) оценки частоты ОТ меньше 0,15, а величину нормированного СКО оценки границ вокализованных участков не хуже 0,07 при ОСШ 20 дБ;

б) алгоритм сегментации непрерывного РС на основе модифицированной сегментирующей функции, применение которого позволило уменьшить вероятность

ложной сегментации с 0,15 до 0,08 при вероятности правильной сегментации 0,92 в случае ОСШ 20 дБ.

4. Предложена процедура, обеспечивающая работу ААПКС на основе модифицированной гибридной модели в потоке слитной речи, на основе явного моделирования акустического окружения с помощью состояния-заполнителя, которая позволила исключить из алгоритма этап поиска границ ключевого слова и уменьшить расход вычислительных ресурсов в шесть раз.

5. Обоснована необходимость введения акустического контекста в целях повышения робастных свойств ААПКС, что позволило увеличить вероятность правильного обнаружения на 0,05 и снизить вероятность ложного обнаружения на 0,02 по сравнению с известными системами. При этом средняя вероятность правильного обнаружения составила 0,94, а вероятность ложного обнаружения – 0,07.

6. Предложена процедура учета в СММ плотности вероятности длительности состояния для явного моделирования изменчивости РС во временной области, которая позволила увеличить вероятность правильного обнаружения на 0,05 и уменьшить вероятность ложного обнаружения на 0,02 по сравнению с известными алгоритмами. При этом вероятность правильного обнаружения составила 0,95, а вероятность ложного обнаружения – 0,05.

7. Доказана возможность применения ВК на основе самоорганизующейся сети Кохонена, обученной предложенным методом с предварительной кластеризацией обучающей выборки методом иерархической группировки Варда, что позволило уменьшить размерность пространства признаков с 14 непрерывных координат до 7 двоичных разрядов, снизить вычислительные затраты и уменьшить требуемый объем памяти для обеспечения дикторонезависимости в 1,75 раза.

8. Разработан алгоритм автоматической оценки параметров ААПКС для модифицированной гибридной модели без использования ручной разметки обучающей выборки, что обеспечило достижение среднего, по тестовой выборке, значения величины правдоподобия последовательности наблюдений ключевого слова равного 0,84 без оптимизации параметров ААПКС.

9. Предложена целевая функция на основе информационной меры Кульбака-Лейблера для оценки качества обучения ААПКС и экспериментально определено значение целевого порога для проверки условия остановки алгоритма равное $1,7 \times 10^{-6}$.

10. Предложен алгоритм оптимизации параметров ААПКС на основе модифицированной гибридной модели, включающий реализацию алгоритма вычисления прямой переменной на рекуррентной ИНС, что позволило увеличить надежность обнаружения ключевых слов на 3 %, достигнув вероятности правильного обнаружения 98 %.

11. Проанализирована возможность аппаратной реализации ААПКС на основе модифицированной гибридной модели. Вычислительные затраты составили около 4×10^6 вычислительных операций в секунду и около 2×10^4 слов памяти для хранения переменных и параметров. Обоснована возможность аппаратной реализации ААПКС на ЦСП TMS320C5532 фирмы Texas Instruments.

СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

1 Кириллов С.Н., Дмитриев В.Т., Баландин И.В. Дикторонезависимая система автоматического поиска ключевых слов в потоке слитной речи// Научная сессия МИФИ-2008. Сборник научных трудов Т. 11.- М., 2007. – С.177 -178.

2 Кириллов С.Н., Дмитриев В.Т., Баландин И.В. Исследование устойчивости к акустическим шумам дикторонезависимой системы автоматического поиска ключевых слов//15-я Международная научно-техническая конференция "Проблемы передачи и обработки информации в сетях и системах телекоммуникаций": Тез. докладов. Часть 1. – Рязань: РГРТУ, 2008. –С.54-56.

3 Дмитриев В.Т., Баландин И.В. Дикторонезависимая система автоматического поиска ключевых слов в потоке слитной речи устойчивая к акустическим шумам// Вестник РГРТУ 2008 № 2 (Выпуск 24) – С. 15-19.

4 Кириллов С.Н., Дмитриев В.Т., Баландин И.В. Алгоритм сегментации речевых сигналов на основе вейвлет-пакетного разложения// Научная сессия МИФИ-2009. Сборник научных трудов Т. 3.- М.: 2009. – С.140.

5 Дмитриев В.Т., Баландин И.В. Исследование алгоритмов сегментации речевого сигнала при определении патологии гортани// Биотехнические, медицинские и экологические системы и комплексы. Международная конференция с элементами научной школы для молодежи. Материалы конференции. Рязань. РГРТУ. 2009. С. 161 – 164.

6 Баландин И.В. Разработка алгоритма сегментации речевого сигнала на основе вейвлет-пакетного разложения// 14 ВНТК студентов молодых ученых и специалистов «Новые информационные технологии в научных исследованиях и образовании». Материалы конференции. Рязань 2009. –С. 114-115.

7 Баландин И.В. Разработка алгоритма обучения нейронной сети для классификации акустически однородных участков речи// 14 ВНТК студентов молодых ученых и специалистов «Новые информационные технологии в научных исследованиях и образовании». Материалы конференции. Рязань 2009. – С. 244-245.

8 Баландин И.В. Алгоритм сегментации речевых сигналов для автоматической системы поиска ключевых слов// Вестник РГРТУ 2010 № 2 (Выпуск 32) – С. 27-31.

9 Баландин И.В. Алгоритм дикторонезависимого поиска ключевых слов// 15 ВНТК студентов молодых ученых и специалистов «Новые информационные технологии в научных исследованиях и образовании». Материалы конференции. Рязань 2010. – С. 142-143.

10 Баландин И.В. Применение дискретного вейвлет преобразования с дробным показателем сжатия в системе поиска ключевых слов// 16 ВНТК студентов молодых ученых и специалистов «Новые информационные технологии в научных исследованиях и образовании». Материалы конференции. Рязань 2011. – С. 244-245.

11 Баландин И.В. Разработка алгоритма поиска ключевых слов на основе рекуррентной нейронной сети// 16 ВНТК студентов молодых ученых и специалистов «Новые информационные технологии в научных исследованиях и образовании». Материалы конференции. Рязань 2011. – С. 245-246.

БАЛАНДИН Иван Васильевич

**Алгоритмы поиска ключевых слов в радиотехнических
устройствах обработки речевой информации, устойчивые к
воздействию мешающих факторов**

Автореферат диссертации на соискание
ученой степени кандидата технических наук

Подписано в печать 27.09.2013

Формат бумаги 60x84 1/16.

Условных печатных листов 1,1.

Тираж 100 экз. Заказ № 24.

Участок оперативной полиграфии ГНУ ВНИИМС Россельхозакадемии,
390025, Рязань, ул. Щорса, д. 38/11.