

На правах рукописи



**Сагацян Максим Владимирович**

**РАЗРАБОТКА И ИССЛЕДОВАНИЕ КОЛЛЕКТИВНЫХ  
НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ ДИКТОРОНЕЗАВИСИМОГО  
РАСПОЗНАВАНИЯ РЕЧЕВЫХ СИГНАЛОВ**

Специальность: 05.12.04

Радиотехника, в том числе системы и устройства телевидения

**АВТОРЕФЕРАТ**

диссертации на соискание учёной степени  
кандидата технических наук

Владимир – 2015

Работа выполнена на кафедре динамики электронных систем  
ФГБОУ ВПО «Ярославский государственный университет им. П.Г. Демидова»

Научный руководитель: **Брюханов Юрий Александрович**  
доктор технических наук, профессор,  
заведующий кафедрой динамики электронных  
систем ФГБОУ ВПО «Ярославский  
государственный университет им. П.Г.  
Демидова», г. Ярославль.

Официальные оппоненты: **Андреев Владимир Григорьевич**  
доктор технических наук, доцент, профессор  
кафедры радиотехнических систем ФГБОУ  
ВПО «Рязанский государственный  
радиотехнический университет», г. Рязань.

**Новоселов Сергей Александрович**  
кандидат технических наук, старший научный  
сотрудник, ООО «ЦРТ-инновации»,  
г. Санкт-Петербург

Ведущая организация: ОАО «Ярославский радиозавод», г. Ярославль

Защита диссертации состоится «6» октября 2015 г. в 16.30 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 «15» июля 2015 г.

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

Ученый секретарь диссертационного  
совета д.т.н., профессор



А.Г. Самойлов

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы и состояние вопроса.** В настоящее время вопросы проектирования и создания системы распознавания речевых сигналов, устойчивых к шумам, с низкой частотой появления ошибок, являются актуальной проблемой. Коммерческие программы управления радиотехническими устройствами посредством речевых сигналов появились в начале девяностых годов прошлого века. Они востребованы людьми с ограниченными возможностями, которым из-за травмы руки сложно набирать большое количество текста. Также данные технологии востребованы людьми, у которых по какой либо причине заняты руки. Например, пожарному при чрезвычайной ситуации легче с помощью голоса воспользоваться радиотехническим устройством, чем с помощью рук. Данные программы основаны на обработке сигналов, то есть переводят голос пользователя в текст, таким образом снимая нагрузку с его руки.

Применение технологий распознавания речевых сигналов актуально в области управления радиотехническими устройствами, такими как, например: радиоприемником, рацией, телевизионным устройством, мобильным телефоном, сканером магнитно-резонансной томографии, рентгеновским сканером и др.

Анализируя работы ученых по созданию и исследованию систем дикторнезависимого распознаванию речевых сигналов, можно отметить, что на настоящий момент достигнута вероятность дикторнезависимого распознавания речевых сигналов для малого словаря 93 % и для большого словаря 90,41 %. Также стоит отметить, что данные результаты получены не для русскоязычных речевых сигналов. Следовательно, создание и исследование систем дикторнезависимого распознавания русскоязычных речевых сигналов является весьма актуальной задачей.

Наибольший интерес состоит в создании алгоритма автоматического дикторнезависимого распознавания речевых сигналов ориентированного на большой словарь и дающего высокую точность распознавания даже в присутствии различных шумов. Для отечественного рынка (рынка Российской Федерации) также востребована возможность работы таких систем с русской речью. Такую задачу можно решить, создав интеллектуальный человекоподобный алгоритм аналогичный акустической системе человека. На сегодняшний день наиболее приближены к данной системе искусственные нейронные сети (ИНС). Проблема создания систем автоматического распознавания речи на основе ИНС изучается с 70-х годов, но из-за низких вычислительных мощностей, больших успехов не достигала. С увеличением вычислительных мощностей ЭВМ возникает среда, в которой можно создавать и тестировать алгоритмы с большой вычислительной сложностью. На сегодняшний день таких мощностей становится достаточно,

чтобы с высокой точностью решить поставленную задачу.

Задача распознавания речевых сигналов является частью задачи распознавания слитной речи. Анализируя работы ученых L. Breiman, Lawrence R. Rabiner, Y.T. Chen, S. Furui, W. Sigin и спрос современного рынка, можно установить, что для управления радиотехническими устройствами при помощи речевых сигналов система автоматического распознавания речи должна отвечать следующим требованиям:

- возможность работы в режиме реального времени;
- достаточное качество распознавания;
- дикторнезависимость;
- возможность работы с русской речью;
- устойчивость к внешним шумам.

Последнее требование связано с тем, что для повышения надежности распознавания речевых сигналов требуется построить систему, не зависящую от внешних шумов для применимости алгоритма в различных условиях.

Существующие методы распознавания речевых сигналов не отвечают абсолютно всем заявленным требованиям. Данное обстоятельство определяет актуальность исследований в данном направлении.

**Направление диссертационной работы соответствует области исследований:**

1. Разработка методов и устройств передачи, приема, обработки, отображения и хранения информации. То есть в диссертационной работе исследуется разработка методов приема, обработки, отображения и хранения информации дикторнезависимого распознавания русскоязычных речевых сигналов в радиотехнических устройствах.

2. Разработка перспективных информационных технологий, в том числе цифровых в радиотехнических устройствах. То есть с помощью систем обработки сигналов, выполняющих дикторнезависимое распознавание русскоязычных речевых сигналов, можно повысить эффективность радиотехнических устройств, таких как, например, радиоприемника, рации, мобильного телефона, телевизионного устройства, сканера магнитно-резонансной томографии, рентгеновского сканера и др.

**Целью работы** является разработка и исследование результативного алгоритма дикторнезависимого распознавания речевых сигналов для управления радиотехническими системами на базе математического аппарата искусственных нейронных сетей с устойчивостью к внешним шумам.

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

1. Анализ существующих моделей, методов и алгоритмов распознавания речевых сигналов с целью выявления степени их соответствия современным требованиям и выбора прототипов для собственных исследований и создания модифицированного алгоритма.

2. Разработка моделей и алгоритмов распознавания речи, обеспечивающих достижение следующих показателей распознавания речевых сигналов:

– скорость работы, достаточная для использования в режиме реального времени;

– высокая вероятность дикторонезависимого распознавания (для малого словаря не менее 93 % и для большого словаря не менее 90,41 %);

– возможность работы с русской речью;

– устойчивость к шумам без большой потери вероятности распознавания.

3. Программная реализация в среде MatLAB предлагаемых алгоритмов и проведение экспериментальных исследований, подтверждающих их результативность.

#### **Методы исследования**

При решении поставленных задач использованы методы теории вероятностей, теории случайных процессов, математического анализа и аналитической геометрии, цифровой обработки сигналов, дискретного преобразования Фурье, теории нейронных сетей и теории программирования.

**Объектом исследования** являются системы автоматического дикторонезависимого распознавания речевых сигналов.

**Предметом исследования** являются модели и алгоритмы распознавания речевых сигналов на основе искусственных нейронных сетей.

#### **Научная новизна**

Впервые получены следующие научные результаты:

1. Разработан нейросетевой алгоритм bagging-коллектива на основе перцептронов Розенблатта с обучением масштабируемых сопряженных градиентов (Scaled Conjugate Gradient Backpropagation, SCG) позволяющий решать задачу дикторонезависимого распознавания русскоязычных речевых сигналов для малого словаря с вероятностью распознавания 97,1 %, что на 4,1 процентных пункта выше существующих результатов.

2. Предложена модификация коллективного нейросетевого алгоритма, позволяющая результативно решать задачу дикторонезависимого распознавания русскоязычных речевых сигналов.

3. Разработан модифицированный коллективный нейросетевой алгоритм на основе перцептронов Розенблатта с обучением SCG, позволяющий решать задачу дикторонезависимого распознавания русскоязычных речевых сигналов для

большого словаря с вероятностью распознавания 95,7 %, что на 5,29 процентных пункта выше существующих результатов.

4. Разработан коллективный и модифицированный коллективный нейросетевые алгоритмы с блоками шумоподавления дикторонезависимого распознавания русскоязычных речевых сигналов, работающие в условиях шумов.

#### **Практическая значимость**

1. Предложенная модификация коллективного нейросетевого алгоритма расширяет возможности нейросетевых алгоритмов дикторонезависимого распознавания русскоязычных речевых сигналов.

2. Вероятность распознавания речевых сигналов для разработанного нейросетевого алгоритма bagging-коллектива на основе персептронов Розенблатта с обучением SCG с блоком шумоподавления для малого словаря в интервале от 5 до 20 дБ равняется 93,5 % при использовании алгоритма шумоподавления Скалара на основе винеровской фильтрации.

3. Вероятность распознавания речевых сигналов для разработанного модифицированного нейросетевого алгоритма bagging-коллектива на основе персептронов Розенблатта с обучением SCG с блоком шумоподавления для большого словаря в интервале от 15 до 20 дБ равняется 93,6 % при использовании алгоритма шумоподавления на основе бинарных масок, использующего критерий статистического детектирования на основе апостериорного отношения сигнал/шум.

4. Разработана программа «NN-SCG speech recognition» (свидетельство о государственной регистрации программы для ЭВМ № 2015616920), с помощью которой проведен анализ различных алгоритмов нейросетевого дикторонезависимого распознавания русскоязычных речевых сигналов.

Результаты работы внедрены в соответствующие разработки ООО «ПАНТЕОН» (г. Ярославль) и ООО «А-Вижн» (г. Ярославль). Все результаты внедрения подтверждены соответствующими актами.

**Достоверность материалов диссертационной работы** подтверждена согласованностью результатов математического моделирования разработанных алгоритмов и экспериментальной проверки в условиях полунатурного моделирования на реальных речевых сигналах, апробацией в печати и на научно-практических конференциях различного уровня.

**Апробация работы.** Результаты работы докладывались и обсуждались на следующих конференциях:

– 14-й и 15-й Международной конференции «Цифровая обработка сигналов и её применение», Москва, 2012-2013;

– Международной конференции «Системы синхронизации, формирования и

обработки сигналов в инфокоммуникациях», Ярославль, 2013;

– 11-й и 12-й Международной научно-технической конференции «Опτικο-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации», Курск, 2013, 2015;

– Международной конференции «Перспективные технологии в средствах передачи информации», Владимир, 2013;

– Международной конференции студентов и аспирантов «Путь в науку», Ярославль, 2014-2015;

– 53-й Международной научной студенческой конференции МНСК-2015, Новосибирск, 2015;

– XIII Всероссийской научной конференции «Нейрокомпьютеры и их применение», Москва, 2015.

**Публикации.** По теме диссертации опубликовано 17 научных работ, из них 3 статьи в журналах, рекомендованных ВАК для публикации результатов кандидатских и докторских диссертаций, 14 докладов на научных конференциях. Получено свидетельство о регистрации программы для ЭВМ.

**Личный вклад автора.** Выносимые на защиту положения предложены и реализованы автором самостоятельно в ходе выполнения научно-исследовательских работ на кафедре динамики электронных систем Ярославского государственного университета им. П.Г. Демидова.

**Структура и объем работы.** Диссертация состоит из введения, четырех глав, заключения, списка литературы и четырех приложений. Содержание работы изложено на 135 страницах. Список литературы включает 104 наименований. В работе представлено 28 рисунков и 9 таблиц.

**Основные научные положения и результаты, выносимые на защиту:**

1. Алгоритм bagging-коллектива на основе перцептронов Розенблатта с обучением SCG для решения задачи дикторонезависимого распознавания русскоязычных речевых сигналов.

2. Модификация коллективного нейросетевого алгоритма, позволяющая решать задачу дикторонезависимого распознавания русскоязычных речевых сигналов для большего размера словаря.

3. Результаты исследования работы коллективных и модифицированных коллективных нейросетевых алгоритмов с блоком шумоподавления для решения задачи дикторонезависимого распознавания русскоязычных речевых сигналов в условиях шумов.

## **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** освещаются актуальные проблемы машинного распознавания речи. Ставятся задачи, которые необходимо решить в ходе выполнения работы.

Показана необходимость использования мел-частотных кепстральных коэффициентов речевых сигналов для задачи дикторонезависимого распознавания речи, так как данное представление приближенно к восприятию речевых сигналов слуховым анализатором мозга человека. Использование кепстральных коэффициентов позволяет уменьшить размерность исходного пространства признаков речевого сигнала, что сказывается на быстродействии исследуемых алгоритмов.

Проведена классификация систем распознавания речи. Можно выделить ряд признаков классификации систем распознавания речи: по типу речи, по зависимости от диктора, по степени детализации эталонов, по количеству распознаваемых слов, по механизму функционирования и по типу грамматики. По типу речи различают системы распознавания речевых сигналов, слитной речи и ключевых слов. В работе рассматривается система распознавания русскоязычных речевых сигналов. По зависимости от диктора различают дикторозависимые и дикторонезависимые системы, а так же системы с автоматической подстройкой. В работе исследуется система дикторонезависимого распознавания русскоязычных речевых сигналов. По степени детализации эталонов различают системы, использующие в качестве эталонов целые слова и части (монофоны, трифоны, слоги, и т. д.) слов. В работе рассматривается система распознавания целых слов. По количеству распознаваемых слов (или объему словаря) можно выделить две категории: системы с малыми (обычно, до 100 слов) и большими словарями. В работе анализируется система дикторонезависимого распознавания русскоязычных речевых сигналов, ориентированная на словарь объемом 102 слова.

Задача распознавания речевых сигналов характеризуется высокой вычислительной сложностью и большим объемом данных для обучения (например, классический речевой корпус для обучения распознаванию англоязычной речи ТИМТ содержит более 500 Мб речевого материала). Для решения такой задачи предлагается использовать подход, основанный на формировании коллектива нейросетевого распознавания на основе метода bagging, потому что:

– обучение отдельных нейронных сетей осуществляется независимо, что позволяет ускорить формирование коллектива за счет распараллеливания процессов обучения отдельных нейронных сетей;



– данный подход позволяет повысить качество обучения и в последующем – распознавания за счет коллективного голосования.

Коллективный нейросетевой алгоритм bagging строится по принципам равноправного голосования нейронных сетей, входящих в данный алгоритм.

Задача дикторнезависимого распознавания речевых сигналов может быть реализована в незашумленных средах. Присутствие в речевом тракте фонового шума может значительно ухудшить качество распознавания речевых сигналов. Для решения данной проблемы рассмотрено три алгоритма шумоподавления:

– алгоритм на основе бинарных масок, использующий критерий статистического детектирования на основе апостериорного отношения сигнал/шум (Ideal Binary Mask – A Posteriori Signal-to-Noise Ratio, далее IBM-PostSNR);

– алгоритм на основе бинарных масок, использующий критерий статистического детектирования на основе априорного отношения сигнал/шум, для оценки которого используется алгоритм TSNR (Ideal Binary Mask – Two-Step Noise Reduction, далее IBM-TSNR);

– алгоритм шумоподавления Скалара на основе винеровской фильтрации (Wiener – A Priori Signal-to-Noise Ratio, далее Wiener-PriorSNR).

**Во второй главе** разработан и исследован нейросетевой алгоритм дикторнезависимого распознавания речевых сигналов.

Для повышения точности распознавания слов предложено объединить отдельные нейросетевые распознаватели в единую систему по принципам коллективного равноправного голосования метода bagging.

Вероятность распознавания речевых сигналов алгоритмом bagging  $P_{bag}$  можно описать следующим образом:

$$P_{bag} = \frac{\sum_{z=1}^H P_{bag}^z}{H}, \quad z = 1, \dots, H, \quad P_{bag}^z = \frac{\sum_{q=1}^L P_q^z}{L}, \quad q = 1, \dots, L,$$

где  $P_q^z$  – вероятность распознавания речевого сигнала  $z$  нейронной сетью  $q$ ,  $L$  – количество нейросетевых распознавателей, входящих в алгоритм bagging,  $H$  – количество речевых сигналов в речевом корпусе.

В качестве алгоритма обучения сетей выбран алгоритм масштабируемых сопряженных градиентов (Scaled Conjugate Gradient Backpropagation, SCG), т. к. он стабильный и достаточно быстрый.

В стандартной форме алгоритма сопряженных градиентов требуется использование линейного поиска, что из-за его характера «проб и ошибок» может занять много времени. В модифицированной (данной) версии алгоритма сопряженных градиентов линейный поиск отсутствует. Он заменен одномерной

формой Левенберга-Марквардта. Основанием для использования именно этого метода было желание обойти сложности, вызываемые неположительностью матрицы Гессе. Формула обновления коэффициентов данного метода:

$$w_{k+1} = w_k + \alpha_k p_k,$$

$$\alpha_k = \frac{\mu_k}{\delta_k}, \quad \mu_k = -p_k^T E'_k, \quad \delta_k = p_k^T s_k,$$

$$s_k = E''(w_k) p_k = \frac{E'(w_k + \sigma_k p_k) - E'(w_k)}{\sigma_k} + \lambda_k p_k,$$

где  $k = 1, 2, \dots, N$ ,  $\alpha_k$  – размер шага,  $p_k$  – сопряженный вектор,  $E_k$  – вектор ошибки,  $\delta_k$  – матрица Гессе,  $\sigma_k = \frac{\sigma}{|p_k|}$ ,  $\lambda_k$  – параметр масштабирования матрицы Гессе.

Для увеличения технических возможностей распознавания речевых сигналов коллективного нейросетевого алгоритма предложено модифицировать bagging-алгоритм. Данное улучшение алгоритма позволяет увеличить размер словаря без потери качества дикторонезависимого распознавания речевых сигналов и расширить сферу применения такого распознавания.

Для построения модифицированного bagging-алгоритма предполагается использовать в качестве основного элемента нейросетевой блок коллективного голосования. Один нейросетевой блок способен обучиться и распознать речевые сигналы без потери качества распознавания на словаре ограниченной размерности. В данном алгоритме используется  $L$  нейросетевых блоков.

Количество нейросетевых блоков пропорционально размеру словаря для того, чтобы качество распознавания речевых сигналов оставалось высоким. То есть количество нейросетевых блоков  $L$  равняется требуемому размеру словаря  $\Omega$ , деленному на размер словаря одного нейросетевого блока  $U$ :

$$L = \frac{\Omega}{U}.$$

Принцип работы данного алгоритма заключается в следующем. Допустим, требуется построить систему распознавания речевых сигналов размерности словаря  $\Omega$ . Так как один нейросетевой блок не способен обучиться без потери качества распознавания на словаре размерности более определенного размера  $U$  для данного нейросетевого блока, то предполагается разбить словарь на  $L$  словарей с размерностью не более определенного размера словаря для одного нейросетевого блока. Таким образом, предполагается обучить каждый нейросетевой блок на словаре, отличном от обучаемых словарей других нейросетевых блоков. И каждый нейросетевой блок обучается на словаре с размерностью, не превышающей определенного размера словаря для данного

нейросетевого блока. В качестве параметров речевого сигнала, по которым проводится распознавание, выбран логарифм энергии сигнала по  $J$  мел-частотным кепстральным коэффициентам. Преобразование входного речевого сигнала в массив MFCC-коэффициентов производится также как и в базовом нейросетевом алгоритме распознавания.

Для исследований использовались собственные речевые базы «КРИПТОН-01» и «КРИПТОН-02». Речевая база «КРИПТОН-01» содержит 10 классов речевых сигналов, а речевая база «КРИПТОН-02» – 102 класса. Данные базы записаны 20 дикторами и разделены на два непересекающихся множества: учебное и тестовое. В качестве записывающих дикторов взяты люди разного пола и возраста.

Проведены исследования по определению параметров для нейросетевого bagging-коллектива дикторонезависимого распознавания русскоязычных речевых сигналов:

– Определено значение размера bagging-коллектива, после которого вероятность распознавания с ростом размера bagging-коллектива растет незначительно (рис. 1). Данный параметр равен 10 нейросетевым распознавателям в коллективном нейросетевом алгоритме, при которых вероятность дикторонезависимого распознавания 10 речевых сигналов равняется 76,1 % распознавания для 4 обучающих дикторов. В качестве нейросетевого распознавателя использовался многослойный перцептрон Розенблатта с обучением SCG.

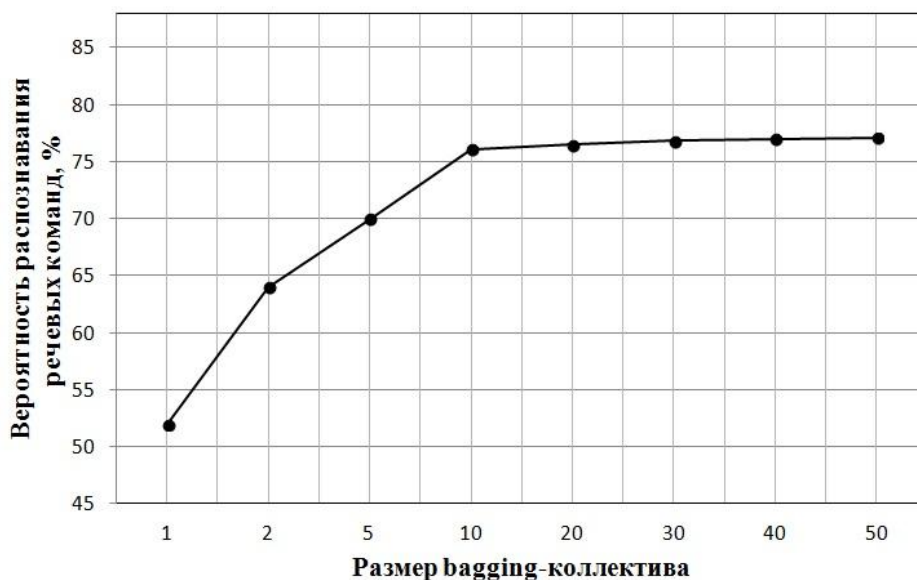


Рис. 1. Зависимость вероятности распознавания речевых сигналов от размера bagging-коллектива нейросетевых распознавателей

– Определено значение количества обучающих дикторов, после которого вероятность распознавания с ростом количества дикторов растет незначительно (рис. 2). Для базового нейросетевого алгоритма данный параметр определился при

11 обучающих дикторах и равен 83 % вероятности распознавания 10 сигналов. Для коллективного нейросетевого алгоритма данный параметр определен при 10 обучающих дикторах и равен 97,1 % вероятности распознавания 10 сигналов. В качестве нейросетевого распознавателя также использовался многослойный перцептрон Розенблатта с обучением SCG.

– Определено значение количества слоев в нейросетевом распознавателе, после которого вероятность распознавания с ростом количества слоев растет незначительно (рис. 3). Исследованы 2 нейросетевых алгоритма bagging-коллектива из 10 перцептронов Розенблатта с обучением SCG и 10 сетей Эльмана с обучением GDХ. Для исследованных алгоритмов данный параметр определен при 12 слоях нейросетевых распознавателей и равен для алгоритмов bagging-коллектива из 10 перцептронов Розенблатта с обучением SCG – 97 %, а для 10 сетей Эльмана с обучением GDХ – 90,7 %.



Рис. 2. Зависимость вероятности распознавания речевых сигналов от количества обучающих дикторов

– Определено значение размера словаря, после которого вероятность распознавания с ростом количества речевых сигналов быстро уменьшается (рис. 4). Исследованы 2 нейросетевых алгоритма bagging-коллектива из 10 перцептронов Розенблатта с обучением SCG и 10 сетей Эльмана с обучением GDХ. Для bagging-коллектива из 10 перцептронов Розенблатта с обучением SCG данный параметр определен на 12 речевых сигналах и равен 96,1 % распознавания. Для bagging-коллектива из 10 сетей Эльмана с обучением GDХ данный параметр определен на 7 речевых сигналах и равен 93 % распознавания.

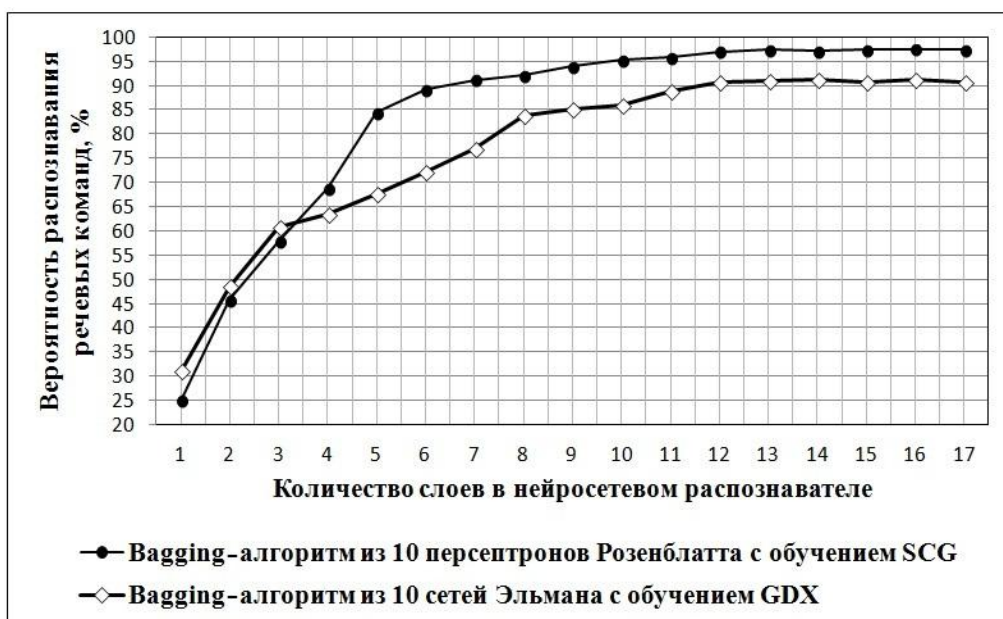


Рис. 3. Зависимость вероятности распознавания речевых сигналов от количества слоев нейросетевых распознавателей

В результате проведения серии экспериментов по исследованию работы модифицированного алгоритма нейросетевого распознавания исследованы модифицированные алгоритмы на основе двух разновидностей нейронных сетей: 10 перцептронов Розенблатта с обучением SCG и 10 сетей Эльмана с обучением GDX. Для вероятности попадания всех значений точности распознавания из генеральной выборки 0,95 оценен доверительный интервал определенной вероятности распознавания речевых сигналов.

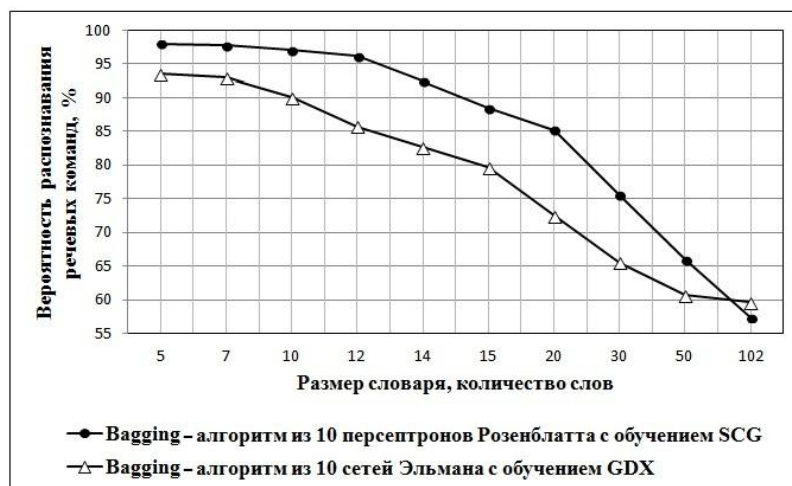


Рис. 4. Зависимость вероятности распознавания речевых сигналов от размера

Модифицированный bagging-коллектив на основе 10 сетей Эльмана распознал 102 речевых сигналов с вероятностью распознавания  $91,5\% \pm 5,5$  п.п.

при времени обучения данного алгоритма 3030 секунд и времени тестирования 380 секунд. Модифицированный bagging-коллектив на основе 10 перцептронов Розенблатта распознал 102 речевых сигналов с вероятностью распознавания  $95,7 \% \pm 3,2$  п.п. при времени обучения данного алгоритма 2688 секунд и времени тестирования 381 секунда. Данные результаты показали конкурентоспособность модифицированного коллективного нейросетевого алгоритма для распознавания речевых сигналов словаря из 102 речевых сигналов.

**В третьей главе** проведен анализ работы нейросетевых алгоритмов обучения в задаче дикторонезависимого распознавания русскоязычных речевых сигналов. Рассмотрено три коллективных нейросетевых алгоритма, основанных на разных алгоритмах обучения: bagging-коллектив 12-слойных перцептронов на основе обучения Левенберга-Марквардта; bagging-коллектив 12-слойных сетей Эльмана на основе обучения GDX и bagging-коллектив 12-слойных перцептронов на основе обучения SCG.

Для оценки качества обучения нейронных сетей выбран стандартный критерий ошибок – средняя сумма квадратов ошибки (СКО) обучения:

$$СКО = \frac{1}{N} \sum_{i=1}^N (e_i)^2 = \frac{1}{N} \sum_{i=1}^N (d_i - y_i)^2,$$

где  $N$  – объем выборки (число примеров в обучающем множестве),  $e_i$  – ошибка сети,  $d_i$  – желаемая величина выхода,  $y_i$  – реально полученные на сети значения для каждого примера  $i$ .

При сравнении работы алгоритмов обучения нейронных сетей по критерию СКО при целевом СКО=0,01 лучшие результаты, равные 0,017 при 41 цикле обучения и 27 секундах, затраченных на данное обучение распознавать 10 речевых сигналов, показал 12-слойный перцептрон на основе обучения SCG; средние результаты, равные 0,052 при 143 циклах обучения и 31 секундах, затраченных на обучение, показала 12-слойная сеть Эльмана; худшие результаты, равные 0,073 при 46 циклах обучения и 1784 секундах, затраченных на данное обучение, показал 12-слойный перцептрон на основе обучения Левенберга-Марквардта.

Для вероятности попадания всех значений точности распознавания из генеральной выборки 0,95 оценен доверительный интервал определенной вероятности распознавания речевых сигналов. При сравнении работы исследуемых коллективных нейросетевых алгоритмов по критерию вероятности распознавания при тестировании алгоритмов обучения лучшие результаты, равные  $97,1 \% \pm 2,8$  п.п. вероятности распознавания 10 речевых сигналов при затраченном времени обучения 262 секунды, показал bagging-коллектив из 10 многослойных перцептронов на основе обучения SCG; средние результаты, равные  $90,5 \% \pm 4,1$  п.п. вероятности распознавания 10 речевых сигналов при

затраченном времени обучения 300 секунд, показал bagging-коллектив из 10 многослойных сетей Эльмана на основе обучения GDX; плохие результаты, равные  $84 \% \pm 5,3$  п.п. вероятности распознавания 10 речевых сигналов при затраченном времени обучения 18 000 секунд, показал bagging-коллектив из 10 многослойных перцептронов на основе обучения Левенберга-Марквардта.

**В четвертой главе** проведен анализ работы нейросетевых алгоритмов дикторонезависимого распознавания речевых сигналов в условиях шумов.

Проведен анализ работы нейросетевых алгоритмов дикторонезависимого распознавания речевых сигналов при наличии шумов. Исследованы коллективный и модифицированный коллективный нейросетевые алгоритмы распознавания речевых сигналов с блоками предобработки. Представлено три алгоритма шумоподавления: IBM-PostSNR; IBM-TSNR и Wiener-PriorSNR. Обучение исследуемых нейросетевых алгоритмов производилось на незашумленной речевой базе. Тестирование производилось на речевых базах с различной зашумленностью (-15, -10, -5, 0, 5, 10, 15, 20 дБ). В качестве шума выбран аддитивный белый гауссовский шум. Каждый нейросетевой блок модифицированного bagging-коллектива состоит из 10 многослойных перцептронов на основе обучения SCG.

Для алгоритма bagging-коллектива из 10 многослойных перцептронов на основе обучения SCG с блоками предобработки результаты для трех алгоритмов шумоподавления в целом оказались близкими. При количественной оценке шумоподавления данных алгоритмов средняя вероятность распознавания речевых сигналов на интервале от -15 дБ до 20 дБ оказалась лучшей у алгоритма Скалара – 80,4 %. При показателях зашумления от 5 дБ до 20 дБ алгоритмы шумоподавления дают высокие показатели вероятности распознавания речевых сигналов, такие как 93,5 %, 91,7 %, 91,6 % вероятности распознавания соответственно для алгоритмов шумоподавления Wiener-PriorSNR, IBM-TSNR и IBM-PostSNR.

Для алгоритма модифицированного bagging-коллектива, состоящего из 11 нейросетевых блоков и блоков предобработки, результаты для двух исследуемых алгоритмов шумоподавления (IBM-TSNR, Wiener-PriorSNR) в целом оказались близкими, а для алгоритма IBM-PostSNR они несколько ниже. При количественной оценке шумоподавления данных алгоритмов средняя вероятность распознавания речевых сигналов на интервале от -10 дБ до 20 дБ оказалась лучшей у алгоритмов IBM-TSNR и Wiener-PriorSNR. При показателях зашумления от 5 дБ до 20 дБ использование исследуемых алгоритмов шумоподавления дает удовлетворительные показатели вероятности распознавания речевых сигналов. Высокие показатели распознавания, равные соответственно 93,55 % и 92,65 % дают алгоритмы шумоподавления IBM-TSNR и Wiener-PriorSNR в условиях слабой зашумленности от 15 до 20 дБ, что также имеет практическую значимость.

Из полученных результатов следует, что коллективный и модифицированный коллективный нейросетевые алгоритмы в условиях шумов и без блока шумоподавления распознают речевые сигналы малорезультативно.

Дополнение коллективного и модифицированного коллективного нейросетевых алгоритмов блоками шумоподавления существенно расширяет возможности применения данных нейросетевых алгоритмов для решения задачи дикторонезависимого распознавания речевых сигналов.

**В заключении** подводятся итоги выполненной работы.

**В приложениях** приведены: информация об используемых речевых базах, копии актов о внедрении результатов работы, копия свидетельства о государственной регистрации программы для ЭВМ.

## **ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ**

1. Разработан алгоритм bagging-коллектива на основе перцептронов Розенблатта с обучением SCG и его модификация для решения задачи дикторонезависимого распознавания русскоязычных речевых сигналов.

2. Созданы речевые базы «КРИПТОН-01» с размерностью 10 сигналов и «КРИПТОН-02» с размерностью 102 сигнала для тестирования нейросетевых алгоритмов.

3. Проведен анализ параметров bagging-коллектива многослойных перцептронов Розенблатта с обучением SCG, в результате чего определено, что рационально выбирать размер bagging-коллектива, равным 10; использовать 10 обучающих дикторов; устанавливать по 12 слоев в каждом нейросетевом распознавателе и использовать размер словаря не больше 10. При данных параметрах получена вероятность дикторонезависимого распознавания русскоязычных речевых сигналов 97,1 %, что на 4,1 процентных пункта выше существующих результатов. Учитывая доверительный интервал полученного значения  $\pm 2,8$  процентных пункта, следует, что с вероятностью 0,95 точность распознавания речевых сигналов также лучше существующих результатов.

4. Разработана научно-исследовательская программа, с помощью которой произведен анализ алгоритмов дикторонезависимого распознавания русскоязычных речевых сигналов путем математического моделирования данных алгоритмов, обученных и протестированных на речевых базах «КРИПТОН-01» и «КРИПТОН-02». Авторские права защищены свидетельством о государственной регистрации программы для ЭВМ.

5. Проведены исследования по определению параметров для нейросетевого bagging-коллектива дикторонезависимого распознавания русскоязычных речевых сигналов.

6. Проведены исследования модифицированных алгоритмов на основе



двух разновидностях нейронных сетей: 10 перцептронов Розенблатта с обучением SCG и 10 сетей Эльмана с обучением GDХ. Модифицированный bagging-коллектив на основе 10 сетей Эльмана распознал 102 речевых сигнала с вероятностью распознавания 91,5 % при времени обучения данного алгоритма 3030 секунд и времени тестирования 380 секунд. Модифицированный bagging-коллектив на основе 10 перцептронов Розенблатта распознал 102 речевых сигнала с вероятностью распознавания 95,7 % при времени обучения данного алгоритма 2688 секунд и времени тестирования 381 секунд, что на 5,29 процентных пункта выше существующих результатов. Учитывая доверительный интервал  $\pm 3,2$  процентных пункта для полученного значения вероятности распознавания модифицированного bagging-коллектив на основе перцептронов Розенблатта с обучением SCG, можно утверждать, что с вероятностью 0,95 точность распознавания речевых сигналов также лучше существующих результатов.

7. Представлены коллективный и модифицированный коллективный нейросетевые алгоритмы с блоком шумоподавления для решения задачи дикторонезависимого распознавания русскоязычных речевых сигналов в условиях шумов.

8. Проведены исследования работы коллективного и модифицированного коллективного нейросетевых алгоритмов с блоками шумоподавления дикторонезависимого распознавания речевых сигналов в условиях шумов. В результате этого показаны высокие результаты распознавания в условиях воздействия слабых шумов.

## **ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

### **Статьи в журналах из перечня ВАК**

1. Сагациян, М.В. Разработка и исследование нейросетевого алгоритма дикторонезависимого распознавания речевых команд / М.В. Сагациян, А.В. Куликов, Г.С. Тупицин // Вестник Поволжского государственного технологического университета. Сер.: Радиотехнические и инфокоммуникационные системы. – Йошкар-Ола: 2014. – № 1(20). – С. 62-68.
2. Сагациян, М.В. Анализ эффективности нейросетевых алгоритмов в задаче дикторонезависимого распознавания речевых команд / М.В. Сагациян, Г.С. Тупицин // Информационные системы и технологии.- Орел: 2015. – № 3. – С. 19-26.
3. Сагациян, М.В. Повышение эффективности коллективного нейросетевого алгоритма на основе обучения SCG в задаче дикторонезависимого распознавания речевых команд в условиях шумов / М.В. Сагациян, Г.С. Тупицин, С.А. Кравцов, А.Л. Приоров // Информационные системы и технологии. – Орел: 2015. – № 4. - С. 39-46.

## Материалы российских и международных конференций

4. Сагациян, М.В. Зависимость точности дикторонезависимого распознавания речевых команд нейросетевыми алгоритмом от количества обучающих дикторов / М.В. Сагациян, Г.С. Тупицин // Докл. 11-й междунар. научно-технической конф. «Опτικο-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации». – Курск: 2013. – С. 189-191.
5. Сагациян, М.В. Коллективное нейросетевое распознавание речи с алгоритмом обучения масштабируемых сопряженных градиентов / М.В. Сагациян // XIII Всероссийская научная конференция «Нейрокомпьютеры и их применение». – М.: ГБОУ ВПО МГППУ, 2015. – С. 45.
6. Сагациян, М.В. Метод обучения и тестирования нейронных сетей для выполнения задачи дикторонезависимого распознавания речевых команд / М.В. Сагациян // Докл. 66-й Всероссийской научно-технической конф. студентов, магистрантов и аспирантов с международным участием. – Ярославль: Издательство ЯГТУ, 2013. – С. 119-121.
7. Сагациян, М.В. Нейросетевое распознавание речевых команд в условиях шумов / М.В. Сагациян // Международная молодежная научно-практическая конференция «Путь в науку», секция «цифровая обработка сигналов и изображений». – Ярославль: 23-30 апреля 2015.
8. Сагациян, М.В. Обучение нейронной сети алгоритмом SCG в задаче дикторонезависимого распознавания речи / М.В. Сагациян, Г.С. Тупицин // Докл. 12-й междунар. научно-технической конф. «Опτικο-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации». – Курск: 2015.
9. Сагациян, М.В. Повышение эффективности коллективного нейросетевого алгоритма в задаче дикторонезависимого распознавания речевых команд в условиях шумов с помощью бинарных масок / М.В. Сагациян, С.А. Кравцов // Докл. 53-й Международной научной студенческой конференции МНСК-2015, секция «Радиотехника и связь». – Новосибирск: 2015. – С. 39.
10. Сагациян, М.В. Разработка и исследование нейросетевого алгоритма дикторонезависимого распознавания слов в устной речи / М.В. Сагациян, С.А. Кравцов, Г.С. Тупицин // Докл. 15-й междунар. конф. «Цифровая обработка сигналов и её применение». – М.: 2013. – Т. 1. – С. 252-255.
11. Тупицин, Г.С. Использование бинарных масок для повышения качества идентификации диктора / Г.С. Тупицин, М.В. Сагациян // Международная конференция студентов и аспирантов «Путь в науку». – Ярославль: 2014.
12. Тупицин, Г.С. Повышение качества закрытой текстонезависимой идентификации диктора с помощью бинарных масок / Г.С. Тупицин, М.В.

Сагациян // Международная молодежная научно-практическая конференция «Путь в науку», секция «Цифровая обработка сигналов и изображений». – Ярославль: 23-30 апреля 2015.

13. Тупицин, Г.С. Повышение качества закрытой текстонезависимой идентификации диктора в условиях шумов с помощью бинарных масок / Г.С. Тупицин, М.В. Сагациян // Докл. 12-й междунар. научно-технической конф. «Оптико-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации». – Курск: 2015.

14. Тупицин, Г.С. Повышение качества идентификации диктора в условиях шумов с помощью бинарных масок / Г.С. Тупицин, А.В. Куликов, М.В. Сагациян // Доклады международной конференции «Перспективные технологии в средствах передачи информации». – Владимир: 2013.

15. Тупицин, Г.С. Использование априорного отношения сигнал/шум для построения бинарных масок в задаче идентификации диктора / Г.С. Тупицин, А.В. Куликов, М.В. Сагациян // Докл. междунар. конф. «Системы синхронизации, формирования и обработки сигналов в инфокоммуникациях». – Ярославль: 2013. - Т. 1. – С. 168-170.

#### **Свидетельство о государственной регистрации программы для ЭВМ**

16. Сагациян, М.В. NN-SCG speech recognition – научно-исследовательская программа по изучению алгоритмов нейросетевого дикторонезависимого распознавания речевых команд / М.В. Сагациян, Г.С. Тупицин // Свидетельство о государственной регистрации программы для ЭВМ № 2015616920 от 30 апреля 2015 г.

Подписано в печать «9» июля 2015 г.  
Формат 60x84 1/16. Тираж 100 экз.

Ярославский государственный университет  
150000, Ярославль, ул. Советская, 14