

На правах рукописи

ГРАЧЕВ Александр Михайлович

**РАСПОЗНАВАНИЕ ЗВУЧАЩЕЙ РУССКОЙ РЕЧИ
В ТЕОРЕТИЧЕСКОМ И ЭКСПЕРИМЕНТАЛЬНОМ
ОСВЕЩЕНИИ: СЕМЕЙНЫЕ, ВОЗРАСТНЫЕ
И ГЕНДЕРНЫЕ АСПЕКТЫ ЛИНГВИСТИЧЕСКОЙ
ИДЕНТИФИКАЦИИ ЛИЧНОСТИ**

Специальность 10.02.01 – Русский язык

Автореферат

диссертации на соискание ученой степени

кандидата филологических наук

Нижний Новгород — 2013

Работа выполнена на кафедре современного русского языка и общего языкознания ФГБОУ ВПО «Нижегородский государственный университет им.Н.И. Лобачевского»

Научный руководитель: доктор филологических наук, профессор
Радбиль Тимур Беньюминович

Официальные оппоненты: **Хромов Сергей Сергеевич**,
доктор филологических наук, профессор,
ФГБОУ ВПО «Московский государственный университет экономики, статистики и информатики», заведующий кафедрой лингвистики и межкультурной коммуникации;

Маркелов Валерий Сергеевич,
кандидат филологических наук, доцент,
ФГАОУ ВПО «Казанский (Приволжский) федеральный университет», доцент кафедры русского языка

Ведущая организация: ФГБОУ ВПО «Пятигорский государственный лингвистический университет»

Защита состоится 12 декабря 2013 года в _____ ч. на заседании диссертационного совета Д 212.166.02 на базе ФГБОУ ВПО «Нижегородский государственный университет им. Н.И. Лобачевского» по адресу: 603000, г. Нижний Новгород, ул. Большая Покровская, 37, ауд. 312.

С диссертацией можно ознакомиться в фундаментальной библиотеке ФГБОУ ВПО «Нижегородский государственный университет им. Н.И. Лобачевского» по адресу: 603950, г. Нижний Новгород, пр. Гагарина, 23.

Автореферат разослан

2013 года

Ученый секретарь
диссертационного совета

Юхнова Ирина Сергеевна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Реферируемое диссертационное исследование посвящено описанию лингвистической идентификации личности по родственному, гендерному и возрастному признаку в экспериментальных моделях автоматического распознавания звучащей русской речи.

Мы живем в информационную эпоху, когда проблема работы с информацией (обработка, получение, хранение, преобразование и пр.) является крайне востребованной в науке и в обществе. В настоящее время для решения разнообразных практических задач широко востребованы технологии по автоматическому распознаванию речи. В отечественной и зарубежной науке сегодня научные исследования, стимулируемые практическими интересами, сконцентрированы на поиске оптимального решения проблемы автоматического распознавания и синтеза речи путем попыток внедрения новых подходов к речевым технологиям.

Актуальность настоящего исследования состоит в том, что оно находится в русле новейших передовых разработок в области автоматического синтеза и анализа русской речи на основе оригинальных отечественных технологий. Очевидна и общественная значимость указанной проблемы. В настоящее время в прикладных целях актуальным становится использование речевых технологий и прежде всего таких, как распознавание, анализ речи, автоматизированная постановка произношения и проч.

Эти технологии можно активно использовать при построении справочных систем, в ходе предварительного обучения систем распознавания и синтеза речи, в научных целях для создания различных речевых корпусов и для практических нужд. Из практических аспектов следует выделить прежде всего использование указанных технологий в практике судебной фоноскопической экспертизы для решения задач по лингвистической идентификации личности.

Совершенно очевидно, что в дальнейшем сфера применимости технологий по распознаванию русской речи будет только расширяться. Ведь с по-

мощью этих технологий можно решать разнообразные идентификационные задачи по выявлению социальных, национальных, профессиональных, гендерных и возрастных особенностей человека в правоведческих, педагогических или медицинских целях.

Объектом исследования являются современные технологии по автоматическому распознаванию русской устной речи, построенные на основе статистических методов (скрытых Марковских моделей).

Непосредственным **предметом** исследования является экспериментальное моделирование выявления родственных, гендерных и возрастных особенностей личности на основе информационной теории восприятия речи (ИТВР).

Гипотеза исследования:

— Экспериментальные исследования русской устной речи дикторов на основе информационной теории восприятия речи, базирующиеся на статистическом анализе минимума рассогласования фонем, позволяют с достаточной степенью надежности выявлять наличие родственных связей, гендерные и возрастные различия дикторов.

Объект, предмет и гипотеза исследования определяют его **цель** — осуществить теоретическое и экспериментальное исследование проблемы лингвистической идентификации личности по родственным, гендерным и возрастным признакам на базе анализа русской устной речи.

Поставленная цель предполагает решение следующих исследовательских **задач**:

— описать историю изучения речевых технологий в экспериментальной фонетике;

— раскрыть проблему автоматического распознавания речи в историко-научном и теоретическом освещении, охарактеризовав лингвистические и статистические подходы к автоматическому распознаванию речи;

— дать теоретическое описание акустических и артикуляционных свойств звуков речи в русской фонетической системе;

— раскрыть теоретические аспекты фонологического восприятия речи как основы теории и практики автоматического распознавания речи;

— обосновать понятие фонетического речевого модуля (речевой сенсорной модальности) как способа обработки информации, содержащейся в звуках речи на базе теории фонемы;

— раскрыть социальные, возрастные и гендерные аспекты распознавания звучащей речи для лингвистической идентификации личности;

— охарактеризовать информационную теорию восприятия речи как технологическую платформу для экспериментальных исследований по автоматическому распознаванию речи;

— подготовить и осуществить три эксперимента по выявлению родственных, гендерных и возрастных особенностей личности посредством оригинальной технологии автоматического распознавания русской устной речи;

— обработать и проанализировать основные результаты экспериментов с точки зрения их соответствия заявленной гипотезе исследования.

Материалом исследования являются обработанные компьютером и оцифрованные записи русской устной речи 30 дикторов разного возраста и гендерной принадлежности, 12 из которых находятся в перекрестных родственных связях. Для записи устной речи был выбран художественный текст — русский перевод романа Стендаля «Красное и чёрное», начало которого каждый диктор читал в течение пяти минут.

Методологической основой данного исследования являются теоретические идеи акустической и перцептивной фонетики, общей фонологии и теории фонологического восприятия речи; технологической платформой исследования выступает новейшая информационная теория восприятия речи.

Методы исследования определяются общими методологическими принципами, согласно которым автоматическое распознавание речи должно опираться на лингвистические данные и моделировать восприятие речи в естественных условиях. В числе методов, использованных в данной работе,

следует назвать метод традиционного лингвистического описания, методы теории информации и теории вероятностей, а также метод эксперимента.

Научная новизна исследования состоит в том, что в работе впервые осуществлено экспериментальное исследование по выявлению родственных, гендерных и возрастных особенностей личности посредством применения современных новейших технологий автоматического распознавания русской устной речи.

Теоретическая значимость исследования состоит в теоретическом обосновании принципов экспериментального моделирования лингвистической идентификации личности по родственным, гендерным и возрастным признакам путем анализа устной речи.

Практическая значимость исследования заключается в том, что его результаты можно использовать в практике проведения судебной фоноскопической экспертизы. Используемую нами методику можно также применять для установления национальной принадлежности говорящего по фонограммам его речи, представляющей индивидуальный «речевой портрет», в котором отражается соответствие или несоответствие произношения человека системе норм неродного языка, а также для идентификации личности по социальному, семейному, возрастному и гендерному признакам, что представляет большую ценность при получении розыскной информации для оперативных служб, следствия и суда.

Отдельные результаты исследования могут быть использованы при построении справочных систем, в ходе предварительного обучения систем распознавания и синтеза речи, в научных целях для создания различных речевых корпусов и для практических нужд, например, в практике вузовского преподавания на кафедре судебной экспертизы юридических вузов.

На защиту выносятся следующие положения:

1. Современные модели автоматического распознавания речи должны основываться на принципах естественного языкового фонологического членения и восприятия речи.

2. Экспериментальные исследования по автоматическому распознаванию русской устной речи дают максимально надежные результаты при использовании новейшей информационной теории восприятия речи.

3. Эксперименты по автоматическому распознаванию русской устной речи на базе информационной теории восприятия речи по критерию минимума информационного рассогласования фонем позволяют осуществить адекватную лингвистическую идентификацию личности посредством выявления родственных, гендерных и возрастных особенностей.

Апробация результатов исследования. Основные положения и результаты исследования были представлены в докладах на международных и межвузовских научных, научно-методических и научно-практических конференциях: «Студенческий гений» (Нижний Новгород, 2010, 2012), «Человек и общество в противоречиях и согласии» (Нижний Новгород, 2010, 2012), «Материалы международной научно-практической конференции, Алматы, 17-18 мая 2012» (Алматы, 2012), «Международная конференция по когнитивной науке» (Калининград, 2012), «Инновации в государстве и праве России» (Нижний Новгород 2013), «Актуальные проблемы права и экономики» (Брест, 2013), «Русский язык как государственный язык Российской Федерации в условиях полиэтнического и поликультурного региона» (Саранск, 2013), «Инновации в государстве и праве России» (Нижний Новгород, 2013), «Историческая русистика и славянское языкознание в начале XXI века: проблемы и перспективы» (Нижний Новгород, 2013). Работа обсуждалась на заседании кафедры современного русского языка и общего языкознания ННГУ им. Н.И. Лобачевского.

Структура работы. Диссертация состоит из введения, двух глав, заключения, библиографического списка и приложения. Библиографический список включает в себя 238 наименований (из них — 10 на английском языке). В Приложении приводятся основные данные по экспериментам, представленные в форме таблиц.

Объем диссертационного исследования, не считая приложения, составляет 168 с. Общий объем исследования — 190 с.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во Введении обосновывается актуальность работы, ее научная новизна, теоретическая и практическая значимость, определяются цель и задачи работы, рабочая гипотеза, объект, предмет и материал исследования, формулируются положения, выносимые на защиту.

Глава I «Распознавание речи как теоретическая проблема современной экспериментальной фонетики» содержит обоснование исходных теоретических положений работы.

В разделе 1.1. «Проблема распознавания речи в историко-научном и теоретическом освещении» рассматривается история изучения речевых технологий в экспериментальной фонетике и делается вывод, что проблема автоматического распознавания речи сегодня является одной из актуальных задач современных исследовательских поисков в сфере создания речевых технологий и имеет целый спектр областей практического применения в самых разных отраслях современной жизни.

В настоящее время наибольшую популярность получили статистические методы распознавания речи, в основе которых лежат скрытые Марковские модели. В ближайшее десятилетие задача распознавания и понимания естественной речи любого говорящего будет занимать центральное место в речевых технологиях.

В разделе 1.2. «Акустические и артикуляционные свойства звуков речи в фонетической системе современного русского языка» обосновывается важность учета акустических и фонологических особенностей естественной человеческой речи в теории и практике автоматического распознавания речи. Раскрывается значимое для данного исследования теоретическое положение о том, что восприятие речи основано не на фонетическом, а на фонематическом членении речи, которое принимает во внимание сигнификативное членение звучащей речи на смыслоразличительные элементы. В этом

плане важно определить принципы классификации наблюдаемых в потоке речи единиц, которые позволили бы отделить значимые звуковые различия от незначимых.

В разделе 1.3. «Фонологическое восприятие речи как основа теории и практики автоматического распознавания речи» рассматриваются основные положения современной теории фонем. В разделе обосновывается, что восприятие речи отличается от восприятия других форм аудиальной стимуляции и потому требует своего собственного, специального механизма обработки информации. Нервная система человека имеет специальный «процессор» — речевую сенсорную модальность — фонетический, или речевой, модуль, предназначенный исключительно для обработки информации, содержащейся в звуках, идентифицированных как речь. Таким образом, речевой слух фонологичен по своей природе. Поэтому в основе экспериментального моделирования автоматического распознавания речи должна лежать теория фонологического восприятия речи, чтобы в результате можно было построить адекватные модели распознавания речи, имеющие определенные аналогии с естественным восприятием речи человеком.

Глава II «Экспериментальные модели по распознаванию звучащей русской речи: социальные, семейные, возрастные и гендерные аспекты лингвистической идентификации личности» содержит теоретическое обоснование и практическое описание хода и результатов экспериментов по автоматическому распознаванию звучащей русской речи на основе информационной теории восприятия речи.

В разделе 2.1. «Социальные, возрастные и гендерные аспекты распознавания звучащей речи и проблема лингвистической идентификации личности» показано, что национальные, социокультурные, социопсихологические, профессиональные, межличностные, семейные, возрастные и гендерные особенности употребления языка проявляются на всех уровнях, в том числе и на уровне устной речи. Выявлено, что существенные отличия на уровне темпа, тембра, произношения отдельных фонем и их групп, а также

просодических и интонационных характеристик звучащей речи присущи людям разного социального положения, возраста, пола и пр.

Все это приводит к обоснованию возможности исследовать родственные, возрастные и гендерные особенности фонации с точки зрения лингвистической идентификации личности экспериментальным путем, а именно посредством применения современных технологий автоматического распознавания речи. Теоретической платформой для экспериментального моделирования автоматического распознавания речи является информационная теория восприятия речи (ИТВР).

В разделе 2.2. «Информационная теория восприятия речи в экспериментах по автоматическому распознаванию речи» описывается теоретический аппарат информационной теории восприятия речи (ИТВР).

Основы информационной теории восприятия речи (ИТВР) были заложены в работах В.В. Савченко [Савченко 2007]. Согласно этой теории, построенной в русле скрытых Марковских моделей, проблема вариативности устной речи решается на основе теоретико-информационного определения самого понятия «фонема».

Несмотря на существующие различия в реализациях некоторой r -ой фонемы все они воспринимаются человеком как нечто общее, иначе речь утратила бы свою информативность. Можно поэтому утверждать, что одноименные реализации $\mathbf{x}_{r,j}$, $j = \overline{1, J_r}$, $J_r \gg 1$, в сознании человека группируются в соответствующие классы или речевые образы фонем $X_r = \{\mathbf{x}_{r,j}\}$, $r = \overline{1, R}$, вокруг некоторого центра — эталонной метки данного образа. В информационной теории восприятия речи указанные эталоны определяются в строгом, теоретико-информационном смысле [Савченко Акатьев, Губочкин 2008: 9]: речевая метка $\mathbf{x}_r^* \subset X_r$ образует *информационный центр* — эталон r -го речевого образа, если в пределах множества X_r она характеризуется минимальной суммой информационных рассогласова-

ний (ИР) по Кульбаку-Лейблеру относительно всех других его меток-реализаций $\mathbf{x}_{r,j}$, $j = \overline{1, J_r}$.

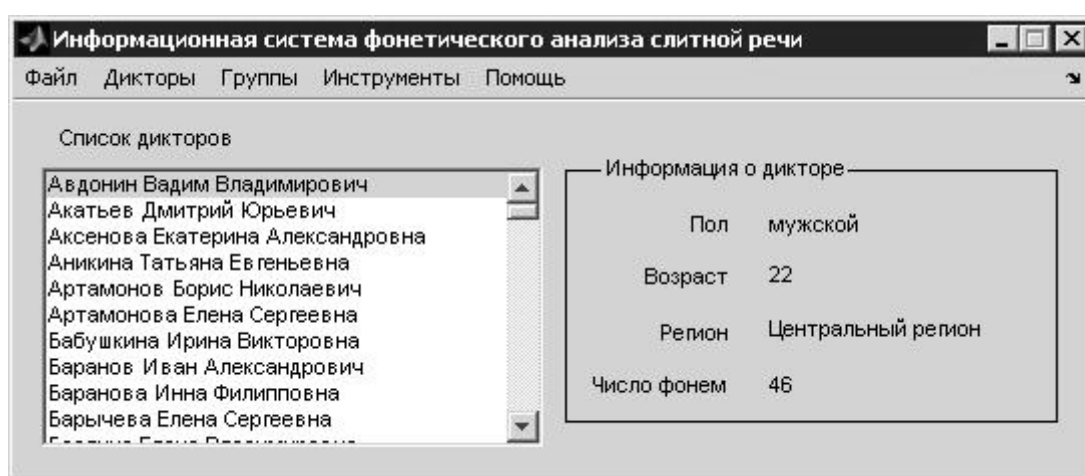
Нетрудно увидеть, что именно в понятии информационного центра (ИЦ) r -го множества реализаций X_r дается наиболее информативное описание свойств соответствующей фонемы. Одновременно становится очевидным и механизм формирования самого этого множества. Анализируемый (входной) речевой сигнал $X(t)$ в дискретном времени $t = 0, 1, \dots$ сначала разбивается на ряд последовательных сегментов данных $\mathbf{x}(t)$ длиной в одну элементарную речевую единицу (ЭРЕ): примерно 10 – 15 *мс*. После этого каждый такой парциальный сигнал рассматривается в пределах конечного списка фонем $\{X_r\}$ и отождествляется с той X_ν из них, которой отвечает минимум информационного рассогласования (МИР) между вектором $\mathbf{x}(t)$ и соответствующим эталоном \mathbf{x}_ν^* , $\nu \leq R$. Это известная формулировка критерия минимального информационного рассогласования в задачах автоматического распознавания речи (АРР) [Савченко, Акатьев 2007: 3].

Теория использует кластерную модель элементарных речевых единиц. Кластерная модель речевых единиц охватывает фонетический (звуковой) строй национального языка. Ее применение в задачах распознавания и анализа речи позволило существенно сократить вычислительную сложность используемых для этого алгоритмов и повысить точность их работы. Кроме того, кластерная модель позволяет существенно снизить требования к объему обучающей выборки в задаче АРР. Так, при использовании кластерной модели для обучения достаточно всего нескольких (в идеальном случае одной) эталонных реализаций каждого слова. Это позволяет не только существенно снизить требования к объему обучающей выборки, но и сократить временные затраты на обучение системы АРР.

Эксперименты, описанные в реферируемом исследовании, проводились с помощью специальной экспериментальной установки — информационной

системы фонетического анализа речи (ИСФАР), запатентованной в Роспатенте по заявке 2008614233 от 15.09.2008 [Савченко, Акатьев, Губочкин 2008]. Данная программа (ИСФАР) содержит **адаптивный алгоритм**, впервые опубликованный в работе проф. В.В. Савченко «Автоматическое распознавание речи методом дерева на основе информационного ($R+1$)-элемента» [Савченко 2006: 13–22].

Для реализации предложенного алгоритма была разработана информационная система фонетического анализа слитной речи — ИСФАР. Ее главное окно показано на следующем рисунке.



В окне отображается список дикторов, внесенных в базу данных (БД) и главное меню программы. При выборе любого диктора из списка в правой части окна показывается краткая информация о нем.

Процесс создания фонетической базы данных (ФБД) на базе данной информационной системы (ИС) выполняется в несколько этапов. На первом этапе формируется группа дикторов, и каждый из них проговаривает в среднем темпе лингвистически сбалансированный текст или отрывок из художественного произведения. Объем текста составляет минимум (1 ... 1,5) тыс. печатных знаков. Каждая такая запись с помощью звукового редактора сохраняется в виде соответствующего звукового файла.

На втором этапе в ИС ФБД производится обработка полученных данных по адаптивному алгоритму. В результате формируется множество пер-

сональных фонетических баз данных (ФБД) $\{X_r\}$, учитывающих особенности разных дикторов. Это первый результат автоматической обработки речевых сигналов.

На третьем, заключительном, этапе обработки речевых сигналов для исследуемой группы дикторов формируется отдельная подгруппа по признаку подобия, или похожести, их произношения. В результате отбирается множество соответствующих персональных ФБД. В пределах полученного множества осуществляется объединение выбранных нескольких списков фонем в один общий список $\{X_r\}$ — согласно тому же понятию информационного центра-эталона речевого образа и при учете того же ограничения, что и на этапе формирования персональных ФБД.

Концепция экспериментального исследования. В реферируемом диссертационном исследовании описывается **три эксперимента**, которые представляют собой начало задуманной нами большой работы по созданию системы автоматического распознавания звучащей речи для идентификации личности по различным социальным и / или психологическим признакам.

Проведенные эксперименты призваны продемонстрировать возможности предлагаемой работы и наметить ее основные векторы. В основе указанных экспериментов — моделирование систем автоматического распознавания звучащей речи с использованием критерия минимума информационного рассогласования по трем признакам: родственному, гендерному и возрастному, — которые, на наш взгляд, наиболее адекватно могут быть идентифицированы.

Основные данные по экспериментам приводятся в **Приложении** к настоящей работе.

Программно-аппаратные средства для проведения экспериментов. Для проведения экспериментов использовались следующие программно-аппаратные средства: внешний микрофон AKG C 3000 B, ламповый предусилитель Tube MP PS USB, ПК на базе процессора Intel Core 2 Quad, ОС

Windows XP SP2. В память ПК записывались звуковые файлы в формате *.wav, которые затем обрабатывались с помощью ИС ФАР.

В разделе 2.3. «Экспериментальные модели автоматического распознавания звучащей русской речи по критерию минимума информационного рассогласования» описан ход и результаты трех проведенных экспериментов.

Эксперимент № 1. Распознавание речи по родственному признаку.

В первом эксперименте при помощи программы ИСФАР дан анализ сходства устной речи лиц, состоящих в родственных отношениях. Главной задачей явилась идентификация речи родственников. В эксперименте участвовали 16 дикторов, часть из которых являются родственниками. В группе существуют 12 родственных связей.

Для записи устной речи был выбран художественный текст – роман Стендаля «Красное и чёрное», начало которого каждый диктор читал в течение пяти минут. Темп чтения всех участников эксперимента – средний. Объём текста – около 1050 слов, включая служебные. Выбор текста был обусловлен дифференцированным набором фонем. За это время предполагалось получить максимально возможный объём аллофонов. Далее, после записи речи, был произведён подбор параметров программы ИСФАР для получения оптимально точных результатов исследований.

При записях были установлены восемь наиболее информативных фонем: [а], [и],[м], [м`],[н], [н`],[о], [у]. Выбор данных фонем определился благодаря минимальным значениям возможности сходства коэффициентов информационного рассогласования (сходства фонем друг с другом). Анализируемые фонемы также показывают наибольшие различия по величине информационного рассогласования у разных дикторов.

На первом этапе исследовалась речь одного диктора. С помощью кластерной модели информационных речевых единиц была создана фонетическая база данных (ФБД) диктора $\{X_r\}$ на множестве из 24 основных фонем

русского языка. В состав вошли следующие фонемы: [а], [в], [в'], [ж], [з], [з'], [и], [л], [л'], [м], [м'], [н], [н'], [о], [с], [с'], [у], [ф], [ф'], [х], [х'], [ш], [ы], [э].

Критерием информативности фонем был выбран минимум информационного рассогласования в пределах каждого отдельного фонетического кластера. В результате получили 8 информативных фонем: [а], [и], [м], [м'], [н], [н'], [о], [у].

На втором этапе те же исследования были проведены по двум другим дикторам. Причём один из них был выбран из числа родственников первого диктора. В дополнение к предыдущему критерию МИР был добавлен критерий вероятности ошибки II – го рода. Полученные результаты подтвердили как информативность выбранных фонем, так и высокую надёжность их распознавания. Наиболее информативными в области родства и возраста оказались фонемы [а], [н], [н'] и [у].

Данный выбор 4-х фонем был сделан как наиболее оптимально показывающий сходство и различие в речи дикторов, имеющих родственные и неродственные отношения. Это продемонстрировано в **таблице 1** (показано информационное рассогласование между фонемами, выявленное у следующей пары дикторов — Грачева А.М. и Малышева Г.В., не имеющих родственных связей) и **таблице 2** (показано информационное рассогласование, выявленное между фонемами отца и сына — Грачева А.М. и Грачева М.А.).

Таблица 1.

| | а | и | м | м` | н | н` | о | у |
|----|---------|----------|---------|---------|---------|---------|---------|---------|
| а | 4.8728 | 233.7232 | 46.7675 | 11.0005 | 28.6932 | 7.8708 | 33.5972 | 5.3595 |
| и | 153.498 | 28.6827 | 7.2076 | 11.0497 | 6.8171 | 20.2882 | 18.4084 | 6.8164 |
| м | 2.9461 | 54.3739 | 10.3362 | 5.3238 | 5.9499 | 3.263 | 28.2344 | 1.9648 |
| м` | 3.2085 | 48.7537 | 10.9291 | 6.1959 | 9.7207 | 15.1302 | 22.0061 | 1.2037 |
| н | 6.3025 | 104.693 | 15.8431 | 10.4231 | 4.5547 | 1.4449 | 43.7617 | 3.282 |
| н` | 2.6246 | 74.2511 | 17.4144 | 10.9142 | 11.7406 | 9.2191 | 32.2828 | 0.84011 |
| о | 8.6044 | 146.0278 | 37.1909 | 11.3025 | 29.3577 | 10.9684 | 76.0955 | 2.4491 |
| у | 6.9215 | 117.4814 | 23.5162 | 9.6556 | 21.3129 | 30.8435 | 50.3124 | 1.5723 |

Таблица 2.

| | а | и | м | м` | н | н` | о | у |
|----|----------|----------|---------|---------|---------|---------|----------|---------|
| а | 2.1771 | 71.1562 | 8.1055 | 4.504 | 9.698 | 4.2375 | 38.2171 | 13.2061 |
| и | 391.1301 | 3.5972 | 10.1957 | 27.2158 | 8.0757 | 4.3436 | 76.1632 | 56.1623 |
| м | 7.0916 | 360.009 | 76.8557 | 16.3412 | 29.4546 | 4.0031 | 169.8201 | 5.313 |
| м` | 11.9431 | 164.2198 | 32.0464 | 5.5166 | 16.9526 | 14.239 | 83.6102 | 2.6986 |
| н | 4.736 | 219.1389 | 8.247 | 0.7267 | 1.0746 | 3.7865 | 34.7027 | 8.549 |
| н` | 178.7055 | 4.9111 | 31.338 | 46.0114 | 17.0789 | 4.3851 | 15.1353 | 6.5251 |
| о | 29.4342 | 103.8844 | 20.0278 | 37.4076 | 13.221 | 46.778 | 30.3668 | 2.3004 |
| у | 40.8708 | 33.1313 | 5.271 | 8.3505 | 6.4851 | 13.0722 | 12.8538 | 0.21007 |

Значение информационного рассогласования между фонемами [а], [н], [н`] и [у] у родственников гораздо меньше, чем данное значение у дикторов, не имеющих родственной связи. Данные двух таблиц показывают, что среднее рассогласование по выделенным фонемам у неродственников равно 5.054725, а между родственниками — 1.9617175. Таким образом, на данном примере убедительно доказано, что, используя 4 информативные фонемы ([а], [н], [н`] и [у]), можно определять родство дикторов.

Был сделан предварительный вывод о том, что самыми информативными в русском языке являются фонемы с минимальными вариациями своих контролируемых в пределах соответствующего кластера, с минимальными значениями ошибки II-го рода – перепутыванием. При этом в случае родственных отношений дикторов интегральное расстояние между списками информативных фонем минимизируется. Этот принцип может служить признаком установления родственных отношений в задачах голосовой верификации.

В экспериментальной группе существуют 12 родственных связей. После записи речи, был произведён подбор параметров программы ИСФАР для получения оптимально точных результатов исследований. Затем была проведена сегментация фонем дикторов.

Благодаря расчётам при помощи указанной программы было получено 240 таблиц рассогласований речевых единиц.

Данное количество получилось благодаря тому, что была взята величина информационного рассогласования дикторов каждого с каждым, то есть $16^2 - 16 = 240$. Далее производилась обработка результатов.

В процессе анализа таблиц информационного рассогласования были проведены исследования по выявлению родственных связей по следующей формуле $z_n = (x_1 + x_2 + \dots + x_y)$ медиана (9) где:

x – минимальное значение информационного рассогласования отдельной речевой единицы одного диктора между речевой единицей другого диктора.

Y – количество минимальных речевых единиц в записи диктора

z_n - значение общего среднемедианного рассогласования между отдельными дикторами (к примеру между дикторами №1 и №2)

n - тестируемая пара дикторов

K – общая сумма z по отдельным парам дикторов.

В итоге была получена база информационного рассогласования по дикторам, которая была оформлена в виде таблицы минимальных рассогласований (значения рассогласований были округлены до сотых единиц)

| дик-тоы | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1 | 0 | 0.73 | 0.92 | 0.74 | 2.07 | 1.75 | 1.5 | 1.29 | 1.34 | 1.65 | 2.08 | 1.46 | 1.93 | 1.5 | 1.67 | 1.7 |
| 2 | 0.96 | 0 | 0.65 | 0.57 | 1.44 | 1.24 | 1.19 | 0.91 | 1.16 | 1.13 | 2.58 | 1.12 | 1.1 | 1.26 | 1.18 | 1.29 |
| 3 | 0.83 | 0.59 | 0 | 0.61 | 1.78 | 1.64 | 1.05 | 1.22 | 1.27 | 1.6 | 3.13 | 1.57 | 1.66 | 1.82 | 1.09 | 2.12 |
| 4 | 0.8 | 0.53 | 0.63 | 0 | 1.89 | 1.47 | 1.04 | 1.03 | 1.24 | 1.4 | 2.63 | 1.27 | 1.34 | 1.26 | 0.96 | 1.79 |
| 5 | 1.76 | 1.53 | 1.96 | 1.59 | 0 | 0.73 | 1.61 | 1.03 | 2.01 | 1.37 | 1.67 | 1.07 | 1.63 | 4.07 | 2.19 | 1.4 |
| 6 | 1.65 | 1.02 | 1.72 | 1.56 | 0.67 | 0 | 1.5 | 1 | 1.07 | 1.65 | 1.13 | 1.04 | 1.45 | 3.9 | 1.81 | 1.24 |
| 7 | 1.36 | 1.11 | 1.02 | 0.85 | 1.45 | 1.07 | 0 | 0.54 | 1.09 | 1.06 | 2.8 | 1.33 | 1.25 | 1.27 | 0.86 | 2 |
| 8 | 1.08 | 0.88 | 1.31 | 0.8 | 0.98 | 0.87 | 0.59 | 0 | 1.04 | 0.78 | 2.28 | 1.23 | 1.16 | 1.69 | 0.93 | 1.51 |
| 9 | 1.16 | 0.97 | 1.09 | 0.98 | 1.74 | 0.97 | 1.13 | 0.9 | 0 | 1.46 | 1.34 | 1.01 | 1.56 | 3.53 | 1.36 | 2.1 |
| 10 | 1.64 | 1.14 | 1.8 | 1.5 | 1.26 | 1.31 | 1.18 | 0.86 | 1.54 | 0 | 3.45 | 1.41 | 0.9 | 1.59 | 1.16 | 1.43 |
| 11 | 2.14 | 4.51 | 2.25 | 1.54 | 1.4 | 1.02 | 2.75 | 1.94 | 1.45 | 3.05 | 0 | 0.7 | 2.41 | 8.17 | 3.48 | 1.79 |

| | | | | | | | | | | | | | | | | |
|----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 12 | 1.38 | 1 | 1.59 | 1.37 | 1 | 0.98 | 1.48 | 1.07 | 1.16 | 1.42 | 0.74 | 0 | 1.67 | 4.41 | 1.96 | 1.42 |
| 13 | 1.59 | 1.36 | 1.98 | 1.5 | 1.52 | 1.28 | 1.02 | 1.17 | 1.93 | 0.87 | 4.08 | 1.71 | 0 | 2.39 | 0.64 | 1.24 |
| 14 | 1.48 | 1.36 | 1.77 | 1.02 | 5.54 | 4.88 | 2.07 | 2.88 | 5.81 | 1.49 | 7.47 | 4.59 | 2.52 | 0 | 0.84 | 2.75 |
| 15 | 1.57 | 0.88 | 1.15 | 0.91 | 1.55 | 1.6 | 1.01 | 0.91 | 1.71 | 0.7 | 4.24 | 1.81 | 0.72 | 1 | 0 | 1.39 |
| 16 | 2.06 | 1.34 | 2.38 | 2.05 | 1.46 | 1.42 | 2.55 | 1.82 | 3.28 | 1.48 | 2.5 | 1.48 | 1.32 | 4.3 | 4.89 | 0 |

Для удобства и увеличения скорости обработки экспериментальных данных было создано вспомогательное программное обеспечение.

Целью программы является автоматизация работы по анализу большого массива данных, полученных при помощи ИСФАР (информационной системы фонетического анализа речи). Главной задачей, которой решается данным программным обеспечением, является формирование статистических показателей на основе матриц.

В качестве языка разработки используется C# 3.0 под платформой «Microsoft .NetFramework 3.5». Для взаимодействия с документами формата «MicrosoftExcel» используется библиотека «NativeExcel», позволяющая значительно ускорить доступ к данным документа по сравнению с технологией «OLE». Для реализации алгоритмов расчета статистических показателей использовалась стандартная библиотека математических методов.

Порядок работ по анализу исходной информации при помощи ПО:

1. При помощи ИСФАР подготавливается набор матриц.
2. Матрицы с названиями собираются в единой директории.
3. Запускается программное обеспечение, где в качестве параметра «Папка» необходимо указать путь к директории с подготовленными данными.
4. Программное обеспечение полностью производит в автоматическом режиме анализ директории и расчет статистических показателей.
5. Результат работы программы служит отчет .

Так, данное программное обеспечение позволяет осуществить анализ большого набора исходных данных за сравнительно небольшое время.

Значение $K < 1.65$ было определено нами как граница родства. В эксперименте к данному значению относятся те пары дикторов, которые не имеют фактического родства, за исключением пары Соснин В. А. - Грачева А.В. (отец-дочь). Всего из 12 родственных связей было распознано 11, что составило 91, 66% . Двадцать два выделенных значений в **таблице** — значение общего среднемедианного рассогласования между родственниками. Среднее значение рассогласования между родственниками определялось по

формуле:
$$\frac{z_{1-2} + z_{2-1} + z_{1-3} + \dots + z_n}{b}$$
 (10), где b количество взятых вели-

чин рассогласований между родственниками, z_n — значение общего среднемедианного рассогласования между отдельными дикторами (к примеру, между дикторами №1 и №2)

Среднее значение рассогласования между родственниками составило 0.69 и таким же образом было найдено среднее значение между дикторами, не имеющими родственных связей, которое составило 1.74.

В связи с поставленной целью было найдено среднее отношение в речи между родственниками.

В проведённом эксперименте принимали участие две группы родственников: с одной стороны, родители и дети, с другой стороны, родные братья. Для сравнения с устной речью лиц, имеющих родственные связи, была использована группа № 5 (дикторы, не имеющие между собой родственных связей).

Первая группа — рассогласования родителей с детьми. Сюда вошли дикторы: Кажаяев В.А., Кажаяев С.В., Кажаяев А.В., Кажаяев Н.В., Грачев М.А., Грачев А.М., Грачева Е.А., Грачева А.В., Соснина М.С.

Вторая группа — рассогласования между братьями. Сюда вошли дикторы: Кажаяев С.В., Кажаяев А.В., Кажаяев Н.В., Минаев М.И., Минаев С.И., Соснин В.А., Соснин И.А.

Аналогично предыдущему эксперименту (№1) было найдено среднее рассогласование между двумя группами родственников (братьями и родите-

лями с детьми). С помощью формулы № 10 были получены следующие данные:

- среднее рассогласование между неродственниками — **1.74**;
- среднее рассогласование между родственниками — **0.69**;
- среднее рассогласование между родственниками и неродственниками — **1.65**.

При нахождении среднего значения рассогласований получены следующие данные: в первой группе среднее рассогласование — 0.747, во второй группе — 0.642.

Тем самым можно сделать важный вывод: подтверждается генетическая составляющая того, что устная речь братьев более сходна между собой, чем речь детей и родителей.

При дальнейшем изучении рассогласования фонем по определению родственных связей был выработан критерий минимума информационного рассогласования. По той же схеме проводились эксперимент № 2 и № 3.

Эксперимент № 2. Распознавание речи по гендерному признаку.

Цель эксперимента — выявление гендерной принадлежности по ключевым фонемам: [а], [и],[м], [м`],[н], [н`],[о], [у].

Данный эксперимент аналогичен предыдущему — установление родства по указанным выше информативным фонемам. Особенностью данного эксперимента стало использование дикторов обоих полов и по степени родства. В общей базе оказались две родственные пары: отец — сын, сестра — сестра. Также для достоверности эксперимента были добавлены два диктора разного пола, находящиеся в неродственных отношениях (всего 6 чел.).

Для проведения эксперимента у каждого диктора были сделаны аудио записи восьми фонем: [а], [и], [м], [м`], [н], [н`], [о], [у]. В ходе эксперимента каждый диктор проговаривал отдельные фонемы по сто раз с целью получения максимального количества вариаций отдельной фонемы.

Далее полученные фонемы заносились в программу ИСФАР, где создавалась отдельная фонетическая группа по каждому диктору. В работе с

различными реализациями информативных фонем были выделены эталонные реализации, исходя из частоты нахождения их в отдельной фонетической базе каждой фонемы. Таким образом, каждый из шести дикторов получил собственную базу из восьми эталонных фонем.

В программе ИСФАР была взята величина информационного рассогласования каждого диктора. Благодаря этому были получены таблицы с рассогласованиями фонем.

Так же как и в предыдущих исследованиях, из восьми фонем были выделены четыре, в том числе фонемы [а], [н], [н`], [у]. В дальнейшем опыты показали необходимость использования дополнительной фонемы [м`]. Затем было определено среднее арифметическое значение по пяти выделенным нами фонемам у каждой пары дикторов.

Среднее рассогласование по выбранным нами фонемам между дикторами высчитывалось по формуле (11), где z_n – коэффициент величины информационного рассогласования между одинаковыми фонемами разных дикторов, n – количество фонем, использованных в эксперименте.

$$\frac{z_1 + z_2 + z_3 + \dots + z_n}{n} \quad (11)$$

В итоге данный анализ позволил создать общую таблицу рассогласований фонем.

| Дикторы; | Сергеев М.В. | Сергеев В.А. | Большаков Г.Н. | Сомова Ж.А. | Сомова К.А. | Шубникова Е.В. |
|----------------|--------------|--------------|----------------|-------------|-------------|----------------|
| Сергеев М.В. | 0 | 1.83 | 6.54 | 12.52 | 14.04 | 13.06 |
| Сергеев В.А. | 1.64 | 0 | 9.1 | 14.33 | 12.75 | 15.46 |
| Большаков Г.Н. | 8.03 | 8.71 | 0 | 11.32 | 15.69 | 14.01 |
| Сомова Ж.А. | 13.02 | 10.51 | 12.17 | 0 | 2.55 | 6.11 |
| Сомова К.А. | 14.88 | 14.08 | 10.67 | 2.89 | 0 | 8.94 |
| Шубникова Е.В. | 10.79 | 9.9 | 13.07 | 7.14 | 7.56 | 0 |

В данной таблице по минимальным значениям (выделенным красным цветом) показаны распознанные пары родственников (Сергеевы и Сомовы). По меньшим различиям рассогласования (выделенным зелёным цветом) видно сходство в мужской и женской группах. Благодаря данной таблице можно

провести несложные подсчёты. Из различных групп дикторов было взято среднее арифметическое значение.

В результате подсчётов было выявлено, что среднее рассогласования мужских фонем отличается от женских, а именно: отношение в мужской группе равно 8.1, в женской группе равно 7.43, между группами мужчин и женщин — 12.9.

Эксперимент № 3. Распознавание речи по возрастному признаку.

Цель эксперимента — выявление возрастной принадлежности по ключевым фонемам. Был проведён эксперимент, аналогичный предыдущему, — выявление возрастного признака по конкретным информативным фонемам: [а], [и], [м], [м`], [н], [н`], [о], [у].

Его особенностью стало использование дикторов разного возраста. В общей базе оказались три группы дикторов.

В первую группу вошли лица в возрасте 55-58 лет. Во вторую группу вошли дикторы в возрасте 40-41 года. К третьей группе относятся дикторы в возрасте 25- 27 лет.

Для проведения эксперимента у каждого диктора были сделаны аудио записи восьми фонем: [а], [и], [м], [м`], [н], [н`], [о], [у]. В ходе записи каждый диктор проговаривал отдельные фонемы по сто раз. Далее полученные результаты заносились в программу ИСФАР, где создавалась отдельная фонетическая база по каждому диктору.

В работе с различными реализациями различных фонем были выделены эталонные реализации, исходя из частоты нахождения их в отдельной фонетической базе. Таким образом, каждый из девяти дикторов получил собственную базу из восьми эталонных фонем.

В программе ИСФАР была взята величина информационного рассогласования каждого диктора. Благодаря этому были составлены таблицы с рассогласованиями фонем. Так же, как и в предыдущих исследованиях, из восьми фонем были выделены пять, в том числе [а], [н], [н`], [у], [м]. Затем было взято среднее арифметическое значение у каждой пары дикторов. Среднее

рассогласование по выбранным нами фонемам между дикторами также вычислялось по формуле (11).

Итогом стало получение общей таблицы рассогласований дикторов

| Дикторы; Возраст дикторов | Никитин В.А. 58 | Сергеев Н.В. 55 | Беляев И.М. 56 | Седов Д.А. 40 | Николаев Д.С. 41 | Пименов А.С. 25 | Анисимов И.Д. 25 | Чернышев А.Н. 26 | Груздев А.М. 27 |
|---------------------------------|-----------------------|-----------------------|----------------------|---------------------|------------------------|-----------------------|------------------------|------------------------|-----------------------|
| Никитин В.А. 58 | 0 | 4.57 | 5.5 | 9.35 | 12.77 | 12.22 | 9.97 | 8.67 | 10.37 |
| Сергеев Н.В. 55 | 6 | 0 | 7.32 | 10.77 | 14.08 | 13.02 | 11.4 | 10.45 | 12.63 |
| Беляев И.М.56 | 6.68 | 4.56 | 0 | 8.01 | 10.45 | 8.34 | 9.54 | 12.98 | 11.48 |
| Седов Д.А.40 | 8.22 | 7.23 | 8.69 | 0 | 6.12 | 8.95 | 10.56 | 9.55 | 11.43 |
| Николаев Д.С. 43 | 8.32 | 7.87 | 11.34 | 4.45 | 0 | 9.01 | 11.01 | 10.13 | 10.74 |
| Пименов А.С. 25 | 11.44 | 9.34 | 12.65 | 13.95 | 10.48 | 0 | 7.91 | 5.24 | 7.85 |
| Анисимов И.Д.25 | 13.82 | 11.67 | 10.58 | 13.61 | 9.92 | 6.87 | 0 | 5.04 | 4.96 |
| Чернышев А.Н.26 | 12.91 | 10.25 | 12.49 | 12.71 | 8.47 | 5.66 | 6.76 | 0 | 5.54 |
| Груздев А.М. 27 | 11.65 | 14.38 | 9.92 | 11.93 | 8.1 | 4.31 | 7.87 | 3.99 | 0 |

Определив в каждой группе дикторов среднее значение, мы получили следующие данные:

- в первой группе среднее значение по коэффициентам информационного рассогласования по выделенным фонемам — 7.22;
- во второй группе среднее значение по коэффициентам информационного рассогласования по выделенным фонемам — 5.28;
- в третьей группе среднее значение по коэффициентам информационного рассогласования по выделенным фонемам — 6.

Среднее значение по коэффициентам информационного рассогласования по выделенным фонемам между дикторами разного возраста — 10.88.

Следовательно, доказано, что с помощью программы ИСФАР и методу информационных фонем можно различать дикторов по возрасту.

В Заключение приводятся основные выводы и раскрываются дальнейшие перспективы исследования.

Результаты исследования могут иметь большую практическую значимость в практике проведения судебной фоноскопической экспертизы. Отдельные результаты исследования могут быть использованы при построении справочных систем, в ходе предварительного обучения систем распознавания и синтеза речи, в научных целях для создания различных речевых корпусов и для практических нужд, например, в практике вузовского преподавания на кафедре судебной экспертизы юридических вузов.

Перспективы исследования связаны с расширением экспериментальной базы исследования за счет разработки методики распознавания национальных, психоэмоциональных, криминогенных и других особенностей личности посредством разработанной технологии автоматического распознавания устной речи.

Основные положения диссертации отражены в следующих публикациях:

А. Публикации в научных изданиях, рекомендованных ВАК:

1. Грачев А.М. К вопросу об идентификации речи дикторов по родственному признаку // Вестник Нижегородского университета. Математика и информатика. – 2012. – №5 (2). — С.316-320.
2. Грачев А.М. Фонетический речевой модуль как способ обработки звуковой информации // Вестник Нижегородского университета. – 2013. – № 4 (1). – С.367-372.
3. Грачев А.М. Адаптивные алгоритмы распознавания речевого сигнала // Вестник Нижегородского лингвистического университета им. Н.А. Добролюбова. – 2013. – №22. – С.32-39.
4. Грачев А.М. Изучение русских фонем и произношения для идентификации и диагностики личности: к постановке проблемы // Вестник Пятигорского государственного лингвистического университета. – 2013. – №2. – С. 12-17.
5. Грачев А.М. Особенности произношения в русских социальных вариантах языка: диахронический аспект // Вестник Вятского государственного гуманитарного университета. – 2013. – № 2. – С.32-35.

Б. Публикации в других научных изданиях:

6. Грачев А.М. Основные параметры идентификации речи диктора // Человек и общество в противоречиях и согласии: Сб. науч. тр. по материалам международной научно-практической конференции. — Нижний Новгород: Гладкова О.В., 2010. — С.24-27.

7. Грачев А.М. Речевая идентификация личности с помощью компьютерно-математических программ // Вестник Кокшетауского государственного университета им. Ш. Уалиханова. – 2011. – № 3-4. Т.2. – С.64-67.
8. Грачев А.М. Современное состояние фоноскопии и её использование в практике проведения судебных экспертиз // Студенческий гений-2012: сборник статей по материалам X Международной практической конференции студентов, аспирантов и молодых учёных (24 мая 2012 г.): в 2 ч. — Нижний Новгород: Гладкова О. В., 2012. — С.149-152.
9. Грачев А.М. Определение сходства и различия устной речи дикторов при исследовании внутреннего рассогласования фонем в русском языке // Материалы международной научно-практической конференции «Полилингвизм: проблемы и перспективы», Алматы, 17-18 мая 2012 г. — Алматы: КазАТК, 2012. — Т.1. — С. 71-74.
10. Грачев А.М. Актуальные проблемы рассогласования речи дикторов в научных трудах исследователей // Человек и общество в противоречии и согласии: сборник науч. трудов по материалам Международной научно-практической конференции: в 2-х ч. Ч.1. — Нижний Новгород: Гладкова О.В., 2012. —С.35-39.
11. Грачев А.М. Идентификация речи дикторов по родственному признаку // 5-я Международная конференция по когнитивной науке. — Калининград: Калининградский гос. ун-т им. И. Канта, 2012. — С.326-327.
12. Грачев М.А. К вопросу о теоретических аспектах фонемы в общей и русской фонологии // Русский язык как государственный язык Российской Федерации в условиях полиэтнического и поликультурного региона. Материалы XI Межрегиональной конференции по вопросам функционирования русского языка как государственного языка Российской Федерации 22 мая 2013 года. — Саранск: Мордовский гос. ун-т, 2013. — С. 246-252.