На правах рукописи

ХУСАИНОВ Айдар Фаилович

ТЕХНОЛОГИЯ АВТОМАТИЗАЦИИ СОЗДАНИЯ И ОЦЕНКИ КАЧЕСТВА ПРОГРАММНЫХ СРЕДСТВ АНАЛИЗА РЕЧИ С УЧЕТОМ ОСОБЕННОСТЕЙ МАЛОРЕСУРСНЫХ ЯЗЫКОВ

Специальность:
05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ
диссертации на соискание ученой степени кандидата технических наук

Уфа – 2014

Работа выполнена на кафедре информационных систем
Казанского (Приволжского) федерального университета

Научный руководитель:	доктор технических наук, профессор **Сулейманов Джавдет Шевкетович**
Официальные оппоненты:	доктор технических наук, профессор **Ронжин Андрей Леонидович** ФГБУН «Санкт-Петербургский институт информатики и автоматизации Российской академии наук», заместитель директора по научной работе
	доктор технических наук, доцент **Мещеряков Роман Валерьевич** ФГБОУ ВПО «Томский государственный университет систем управления и радиоэлектроники», профессор кафедры комплексной информационной безопасности электронно-вычислительных систем
Ведущая организация:	ФГАУВО «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

Защита диссертации состоится «26» сентября 2014 г. в 12^{00} часов на заседании диссертационного совета Д-212.288.07 на базе ФГБОУ ВПО «Уфимский государственный авиационный технический университет» по адресу:
450000, г. Уфа, ул. К. Маркса, 12.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВПО «Уфимский государственный авиационный технический университет» и на сайте *http://www.ugatu.ac.ru/*.

Автореферат разослан «__» _____ 20__ года.

Ученый секретарь
диссертационного совета,
д.т.н., доцент И. Л. Виноградова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Развитие и широкое внедрение информационных технологий делает актуальной задачу развития более совершенных видов человеко-машинных интерфейсов. Одним из подходов к решению данной задачи является использование речи в качестве канала взаимодействия человека с компьютером. Для практической реализации данного подхода необходимо наличие средств автоматического анализа речи, задача создания которых лежит в области речевых технологий. В целом, в области речевых технологий можно выделить следующие основные направления: автоматическое распознавание речи, идентификация и верификация языка, идентификация и верификация диктора, распознавание эмоций диктора, синтез речи, распознавание тематики разговора.

В настоящее время разработано множество моделей и алгоритмов анализа речи, создано и успешно используется множество коммерческих систем, однако, несмотря на это, существуют задачи, которые не решены до конца, например, задача распознавания слитной и спонтанной речи. Кроме того, степень развития речевых технологий сильно отличается между различными языками. Так, высокое качество работы речевых систем для английского, испанского, французского, китайского и некоторых других языков сочетается со слабым развитием или даже их полным отсутствием для многих других языков. На примере России можно говорить о развитии программных средств распознавания речи, примерно сопоставимых по качеству работы с мировыми аналогами, только для русского языка. Однако в то же время по данным переписи 2010 года в России насчитывается 38 языков, на каждом из которых разговаривает более 100 тысяч человек, и 7 языков, помимо русского, на которых говорят более миллиона человек.

Таким образом, в настоящий момент в мире выделяется класс малоресурсных языков[1], для которых не создано средств автоматического распознавания речи, что препятствует их использованию в современных информационных системах и способствует их вытеснению ведущими мировыми языками.

Факт слабого развития речевых технологий для малоресурсных языков как в России, так и в мире, может быть объясним целым рядом причин. Во-первых, данная ситуация объясняется научной сложностью стоящих перед исследователями задач. Во-вторых, высокими финансовыми затратами на подготовку необходимых программных инструментов, речевых и текстовых корпусов. Однако важным также является тот факт, что существующие на данный момент способы моделирования и создания комплексов распознавания речи, чаще всего, стремятся к решению узкого спектра задач, не учитывая при этом все особенности разработки в контексте работы с малоресурсными языками. Это приводит к тому, что разрабатывать и оценивать качество работы программных средств анализа большинства малоресурсных языков приходится с нуля, используя лишь базовый набор имеющегося инструментария, сталкиваясь и решая схожие для многих других языков проблемы.

---

[1] Малоресурсный язык (термин предложен S. Krauwer, V. Berment) – язык, развитие информационных технологий для которого является недостаточным

Таким образом, можно говорить об актуальности проблемы создания технологии, которая бы позволила реализовать программный комплекс автоматизации создания и оценки качества программных средств распознавания речи для малоресурсных языков.

Гипотеза, проверяемая в данной работе, состоит в том, что использование технологии и реализующего её программного комплекса автоматизации, учитывающих специфику обработки малоресурсных языков, существенно сократит время создания средств распознавания речи для множества малоресурсных языков, сопоставимых по качеству работы с существующими мировыми аналогами.

Степень разработанности темы исследования. Термин «малоресурсные языки» для обозначения класса языков с недостаточным уровнем развития информационных технологий был введен в работах S. Krauwer и V. Berment. Ими также были предложены методики экспертной оценки степени развития данных языков.

Разработками в области речевых технологий в целом, и в контексте малоресурсных языков занимаются такие ведущие научные центры, как Университет Карнеги Меллон, Университет Кембриджа, Массачусетского технологического университета, компании Nuance, IBM, Google. Одними из основоположников подходов, применяемых в области автоматического анализа речи, являются Т. К. Винцюк, предложивший метод динамического программирования, а также А.А. Марков, разработавший теорию стохастических процессов.

Лидерами в области распознавания и синтеза речи в России являются такие научные центры, как Санкт-Петербургский институт информатики и автоматизации РАН, Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, компания Центр речевых технологий.

Объект исследования. Объектом исследования является процесс создания и оценки качества программных средств автоматического анализа речи.

Предмет исследования. Предметом исследования в диссертационной работе является разработка технологии автоматизации создания и оценки качества программных средств анализа речи с учетом особенностей малоресурсных языков.

Цель работы и задачи исследования. Основной целью диссертационной работы является разработка технологии автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков, которая позволила бы повысить скорость создания данных средств при условии сохранения качества их работы на уровне соответствующих мировых аналогов.

Для достижения поставленной цели в ходе диссертационной работы сформулированы и решены следующие задачи:

1. Разработка технологии построения программных средств распознавания речи, позволяющей повысить скорость создания и качество работы данных средств для множества малоресурсных языков.
2. Разработка модели комплекса автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков.

3. Программная реализация комплекса автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков, включающего средства решения вспомогательных задач, таких как проектирование и запись текстовых и речевых корпусов, вычисление параметров речи.

4. Создание программных средств распознавания фонем и слитной речи на татарском языке на базе разработанного комплекса автоматизации.

5. Исследование эффективности разработанного комплекса автоматизации создания и оценки качества программных средств распознавания речи для малоресурсных языков.

Научная новизна.

1. Разработана технология автоматизации создания и оценки качества программных средств анализа речи малоресурсных языков, отличающаяся применением моделей, учитывающих специфику обработки малоресурсных языков и обеспечивающих совместную работу экспертов в области языка, анализа речи, программистов и других специалистов при многоэтапной процедуре проектирования и верификации прикладных систем распознавания речи.

2. Разработана модель комплекса автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков, отличающаяся от существующих аналогов охватом всех основных подзадач области распознавания речи, а также возможностью их настройки для работы с конкретным малоресурсным языком, что позволяет существенно ускорить процесс создания программных средств анализа речи для малоресурсных языков.

3. Разработан программный комплекс автоматизации создания и оценки качества программных средств анализа речи малоресурсных языков и инструментальные средства выполнения алгоритмов автоматического анализа речи малоресурсных языков, отличающиеся использованием созданной технологии, обеспечивающей существенное ускорение процесса создания программного обеспечения анализа речи малоресурсных языков при сохранении качества и скорости его работы на уровне мировых аналогов.

4. Впервые созданы программные средства распознавания фонем и слитной речи на татарском языке на базе разработанного программного комплекса средств, позволяющие использовать их для обеспечения речевого интерфейса взаимодействия человека с компьютером.

Теоретическая и практическая значимость работы. Разработанные модели и программные реализации направлены на решение проблем в области речевых технологий, возникающих при построении и оценке качества программных средств распознавания речи для малоресурсных языков. Предложенная модель позволяет использовать выявленные особенности процессов создания и оценки качества программных средств распознавания речи. Например, учитывая междисциплинарный характер области речевых технологий, предоставляется возможность одновременной работы специалистам по фонетике, лингвистике, программистам с возможностью предоставления настраиваемого для каждого из специалистов доступа к функционалу. Реализация в рамках комплекса модели системы распознавания речи для малоресурсных языков позволяет автоматизировать про-

цессы решения стандартных задач распознавания речи и, таким образом, заметно ускорить процесс создания систем для множества малоресурсных языков.

Методология и методы исследования. Для решения поставленных задач в работе используются методы статистического анализа, теории вероятности, математической статистики, математического моделирования в лингвистике. Программная реализация основана на объектно-ориентированном подходе.

Положения, выносимые на защиту:

1. Технология построения и оценки качества программных средств распознавания речи для малоресурсных языков, основанная на использовании моделей, учитывающих специфику обработки данного класса языков и позволяющих одновременно осуществлять проектирование и верификацию прикладных систем распознавания речи специалистам из разных областей знаний.

2. Модель комплекса автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков, основанная на учете особенностей решения всех основных подзадач области распознавания речи в контексте работы с малоресурсными языками и позволяющая существенно ускорить процесс создания программных средств анализа речи для малоресурсных языков.

3. Программная реализация комплекса автоматизации создания и оценки качества программных средств анализа речи, а также инструментальных средств решения вспомогательных задач автоматического распознавания речи, основанных на использовании разработанной технологии и позволяющих существенно ускорить процесс создания программного обеспечения анализа речи малоресурсных языков при сохранении качества и скорости его работы на уровне мировых аналогов.

4. Программные средства распознавания фонем и слитной речи на татарском языке, созданные на базе разработанного комплекса автоматизации и позволяющие использовать их для обеспечения речевого интерфейса взаимодействия человека с компьютером.

Степень достоверности и апробация результатов. Разработанный программный комплекс был использован в рамках проекта по созданию онлайн-школы обучения татарскому языку «Ана Теле»; проект осуществляется совместно с Министерством образования и науки Республики Татарстан и компанией «English First». Результаты работы внедрены в учебный процесс кафедры математической лингвистики и информационных систем в филологии Института филологии и межкультурной коммуникации Казанского федерального университета.

Основные результаты диссертационного исследования представлялись на Международных конференциях: «Речь и Компьютер» SPECOM (Казань 2011; Пльзень, Чехия, 2013), «Open Semantic Technologies for Intelligent Systems» OSTIS (Белоруссия, 2013), «Computer Science and Information Technologies» CSIT (Австрия, Венгрия, Словакия, 2013), «Computer processing of Turkic languages» (Казахстан, 2013).

Публикации. Основные положения и результаты диссертационной работы опубликованы в 10 публикациях, включающих 2 статьи в научных журналах из перечня ВАК («Доклады Томского государственного университета систем управ-

ления и радиоэлектроники», «Программные продукты и системы»), 1 статью в журнале, цитируемом SCOPUS («Speech and Computer, Lecture Notes in Computer Science», издательство Springer), 1 свидетельство о государственной регистрации программы для ЭВМ.

Структура и объем работы. Диссертационная работа включает введение, три главы, заключение, список литературы. Материал диссертации изложен на 162 страницах текста, включающих в себя 42 рисунка и 26 таблиц. Количество библиографических ссылок – 104.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении содержится обоснование актуальности, сформулированы цели диссертационной работы и решаемые задачи, отмечена научная новизна, практическая значимость результатов, приведены основные положения диссертационной работы, выносимые на защиту, а также сведения о реализации, апробации и внедрении результатов работы.

В первой главе диссертации представлен обзор как текущего состояния исследований в области речевых технологий в целом, так и существующих технологий и программных средств, направленных на решение задач распознавания малоресурсных языков. Приведено описание понятия малоресурсных языков, отмечены основные отличия от миноритарных и вымирающих языков, отмечен способ количественной оценки степени информационного развития языков мира.

Отмечается, что область анализа речи содержит множество задач, начиная от распознавания фонем до анализа тематики высказывания, и, несмотря на значительные успехи в области разработки необходимых алгоритмов и методов, большинство из этих задач не могут считаться до конца решенными ни для одного из мировых языков. Кроме того, отмечается, что уровень разработанности задач анализа речи для класса малоресурсных языков существенно отстает от систем для ведущих мировых языков, что накладывает ограничения по использованию данных языков в инфокоммуникационной среде.

Проведен анализ базовых подходов к автоматическому распознаванию речи. Используемые на сегодняшний день подходы к автоматическому распознаванию речи делятся на 4 основные группы: методы динамического программирования, подходы на основе скрытых Марковских моделей (HMM) или нейросетей, а также множество прочих подходов. Подход на основе аппарата скрытых Марковских моделей, используемый более чем в 80 % современных систем распознавания речи, может быть использован при разработке инструментальных средств автоматического создания средств анализа речи благодаря использованию развитого математического аппарата, а также наличию способов нивелирования основных недостатков данной модели.

Активные исследования в области разработки новых и адаптации существующих программных средств распознавания речи для работы с малоресурсными языками, осуществляемые с 90х годов 20 века, позволили создать инструментальные средства, способные автоматизировать решение отдельных задач, которые возникают при создании систем распознавания речи. Данные средства ак-

тивно используются в действующих системах при решении задач анализа речи, так как позволяют с высоким качеством решать небольшие подзадачи. Однако функциональности данных программных средств недостаточно для осуществления автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков. Кроме того, их использование требует наличия экспертов как в области информационных технологий для обеспечения взаимодействия с программными интерфейсами, так и в области анализа речи – для настройки параметров работы систем. В контексте работы с малоресурсными языками данные требования по наличию экспертов являются сдерживающим фактором создания программных средств автоматического распознавания речи.

Делается вывод о необходимости создания технологии и инструментальных средств автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков, которые позволили бы повысить скорость создания данных средств при условии сохранения качества их работы на уровне соответствующих мировых аналогов.

Вторая глава посвящена описанию подхода к построению комплекса автоматизации создания и оценки качества программных средств распознавания речи для малоресурсных языков. Выделяются и описываются ключевые особенности процесса создания и оценки качества работы программных средств распознавания речи для малоресурсных языков, на базе выявленных особенностей формулируются требования к разрабатываемому комплексу автоматизации. Приводится описание подхода к распознаванию речи на основе HMM, на его базе создаётся модель программных средств распознавания речи для малоресурсных языков.

Предлагаемый комплекс автоматизации создания и оценки качества программных средств распознавания речи для малоресурсных языков базируется на технологии, основанной на выявленных особенностях для малоресурсных языков и способных повысить скорость разработки, а также на модели, обеспечивающей решение непосредственно задачи распознавания речи.

При построении модели данного комплекса предлагается исходить из следующих целевых характеристик: скорость, качество и простота создания базовых систем распознавания речи для малоресурсных языков; возможность оценки качества и скорости работы созданных программных средств распознавания речи.

На основе проведенного анализа выделен список основных особенностей процесса автоматического анализа малоресурсных языков, а также сформированы требования к комплексу автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков, позволяющие реализовать поддержку выявленных особенностей (таблица 1).

Таблица 1 – Взаимосвязь между особенностями анализа речи для малоресурсных языков и требованиями к комплексу автоматизации

Особенности создания систем анализа речи для малоресурсных языков	Требования к комплексу автоматизации создания и оценки качества систем анализа речи для малоресурсных языков
Большое количество задач в области речевых техно-	Поддержка параллельной работы нескольких пользователей. Возможность настройки прав до-

Особенности создания систем анализа речи для малоресурсных языков	Требования к комплексу автоматизации создания и оценки качества систем анализа речи для малоресурсных языков
логий, мультидисциплинарный характер задач	ступа пользователей. Универсальность используемых компонентов систем (возможность их доработки и использования в контексте различных задач). Расширяемость системы.
Взаимосвязанность частей программных средств анализа речи	Наличие механизма обмена информацией между составными частями систем, а также между целыми системами анализа речи
Использование в высоконагруженных приложениях	Наличие средств тестирования работоспособности отдельных компонент и всей системы. Наличие средств оценки качества и скорости работы систем
Проблемно-ориентированность программных средств анализа речи	Наличие объекта, отвечающего за решение конкретной стоящей задачи анализа речи. Возможность создания базовых реализаций систем решения задач анализа речи, а также адаптации и настройки систем для их функционирования в конкретных условиях использования.
Наличие базовых подзадач при создании программных средств анализа речи	Наличие инструментов для решения стандартных задач (например, вычисления векторов признаков речи). Возможность использования существующих программных продуктов (HTK, Julius).

Выявленные требования к комплексу автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков, с одной стороны, а также цели по обеспечению скорости создания и возможности оценки качества работы программных средств анализа речи, с другой стороны, делают необходимым использование различных принципов создания программного обеспечения при проектировании комплекса и создаваемых с его помощью систем.

Так, согласно принципа открытой архитектуры, регламентируются и стандартизируются только описание принципа действия системы и её конфигурация. Система при этом может быть собрана из отдельных составных элементов, разработанных и изготовленных в независимом друг от друга порядке. Использование особенностей данного принципа в контексте проектирования предлагаемого комплекса автоматизации позволит реализовать сформулированные ранее требования по универсальности используемых компонентов систем анализа, возможности их адаптации и повторного использования при решении новых задач анализа речи. Особенности открытой архитектуры в рамках комплекса реализуются следующим образом: создание программных средств анализа речи предлагается осуществлять на основе модульного подхода, где каждый модуль реализует специфическую функциональность, но при этом для всех модулей характерна единая структура.

Объектный подход проектирования информационных систем основывается на понятии объекта как замкнутой независимой сущности, взаимодействующей с внешним миром через строго определенный интерфейс в виде перечня сообще-

ний, которые объект может принимать. Выделение модуля, обладающего характерной структурой, можно также считать реализацией объектного подхода к проектированию информационных систем. Объект модуля также реализует принцип наследования: базовыми для любого модуля комплекса будут являться интерфейсы, которые позволяют комплексу осуществлять с ним следующие стандартные операции, вне зависимости от особенностей реализации каждого конкретного модуля: запускать выполнение заложенных в модуль команд, обеспечивать обмен информацией между модулями, предоставлять возможность настройки прав доступа для различных групп пользователей, осуществлять обмен информацией между модулями с помощью значений входящих и выходящих параметров.

Помимо определения класса модуля, важным также является выделение класса проектов - объектов, объединяющих в себе совокупность модулей. Проекты должны инкапсулировать функциональность и настройки входящих в его состав модулей, а также обеспечивать на их основе выполнение поставленных задач анализа речи. В соответствии со стоящими требованиями по поддержке параллельной работы нескольких пользователей и обеспечению возможности настройки их прав доступа, оправдано выделение отдельного класса объектов пользователей, которые бы хранили информацию, позволяющую проводить аутентификацию и предоставлять права доступа к различной функциональности комплекса.

Необходимость обеспечения высокой скорости и удобства разработки программных средств анализа речи совпадает с целями концепции проектирования программных продуктов быстрой разработки приложений (Rapid Application Development, RAD), наиболее существенными из которых являются: высокая скорость разработки, низкая стоимость, высокое качество. Соответственно, оправдано использование принципов RAD: работа ведется группами; разработка базируется на моделях (моделирование позволяет выполнить его декомпозицию на составные части); используется итерационное прототипирование.

В соответствии с перечисленными принципами, а также выделенными ранее объектами модулей, проектов и пользователей, процесс разработки программных средств анализа речи для малоресурсного языка можно представить следующим образом: создаётся проект для решения конкретной задачи, происходит добавление в него предустановленных в комплекс модулей для решения стандартных подзадач, создание новых модулей для учета особенностей языка и условий использования конечной системы, устанавливаются взаимосвязи между модулями, назначаются права доступа, после чего работа может осуществляться отдельно над каждым модулем различными группами исследователей. Изменение версии реализации любого модуля при этом не нарушает целостность всего проекта; каждая итерация изменения функциональности модуля приводит к расширению функциональности всей системы.

Во второй главе представлен обзор архитектуры модели автоматического распознавания речи. Создание и внедрение в комплекс модели автоматического распознавания речи позволит обеспечить скорость создания и качество работы базовых систем распознавания речи малоресурсных языков. Предложенная модель основывается на подходе к распознаванию речи с использованием HMM.

С формальной точки зрения процесс распознавания начинается с преобразования входящего аудиосигнала в последовательность векторов признаков $O=(o_1,o_2,...,o_T)$. Далее, задача системы сводится к поиску среди всех возможных последовательностей слов $\mathbf{w}=(w_1,w_2,...,w_L)$ такой, которая бы максимально соответствовала данному вектору признаков, то есть поиску такого вектора , который вычислялся бы следующим образом:

$$\hat{\mathbf{w}}=\underset{\mathbf{w}}{\operatorname{argmax}}\ \mathrm{P}(\mathbf{w}\,|\,O). \qquad (1)$$

Существуют подходы, моделирующие процесс распознавания в данном представлении, однако более успешным является подход, основанный на генеративной модели, получаемой после применения Байесовского правила:

$$\hat{\mathbf{w}}=\underset{\mathbf{w}}{\operatorname{argmax}}\ \mathrm{P}(\mathbf{w}\,|\,O)=\underset{\mathbf{w}}{\operatorname{argmax}}\ \frac{\mathrm{P}(O\,|\,\mathbf{w})\mathrm{P}(\mathbf{w})}{\mathrm{P}(O)}=\underset{\mathbf{w}}{\operatorname{argmax}}\ \mathrm{P}(O\,|\,\mathbf{w})\mathrm{P}(\mathbf{w}). \qquad (2)$$

Полученное равенство отражает необходимость в основных моделях систем распознавания речи: вероятность произнесения того или иного слова при заданной последовательности звуков $\mathrm{P}(O\,|\,\mathbf{w})$ определяется акустическими моделями и моделью произношения, а вероятность произнесения слова $\mathrm{P}(\mathbf{w})$ определяется языковой моделью. Схематично процесс работы программных средств распознавания на основе описанного выше алгоритма представлена на рисунке 1.

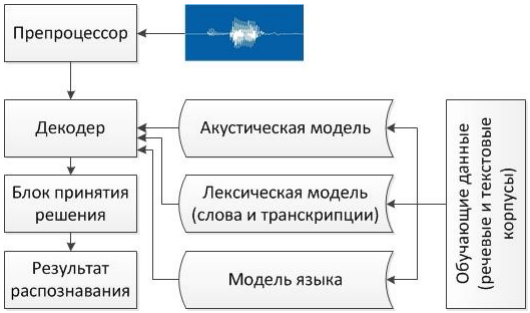

Рисунок 1 – Схема работы программных средств распознавания речи

Первым шагом работы системы распознавания является параметрическое преобразование исходного речевого сигнала. В качестве параметров речевого сигнала используется вектор коэффициентов MFCC, основанный на кепстральном анализе. Процедура его вычисления схематично представлена на рисунке 2.

Рисунок 2 – Процесс вычисления параметра MFCC

На основе вычисленных кепстральных коэффициентов происходит обучение акустических моделей фонем языка. Каждая модель представляет собой отдельную HMM $\lambda=(N,M,A,B,\pi)$, где N – число состояний в модели; M – число Гауссиан, описывающих каждое состояние модели; A – матрица вероятностей пе-

реходов между состояниями модели; B – задает распределение вероятностей появления векторов признаков в каждом из состояний модели; π - распределение вероятностей начальных состояний модели.

Для определения параметров HMM используется алгоритм Баума-Велша с алгоритмом прямого-обратного хода. Для начала работы алгоритма необходимо распределить вектора признаков речи o_t по состояниям модели в пропорциях, равных вероятностям нахождения модели в данном состоянии в момент наблюдения вектора признаков. Обозначим через $L_j(t)$ вероятность нахождения в состоянии j в момент времени t. Тогда оценки математического ожидания и матрицы ковариаций для состояния j с учетом необходимой нормализации могут быть получены на основе следующих выражений:

$$\hat{\mu}_j = \frac{\sum_{t=1}^{T} L_j(t) o_t}{\sum_{t=1}^{T} L_j(t)}, \tag{3}$$

$$\hat{\Sigma}_j = \frac{\sum_{t=1}^{T} L_j(t)(o_t - \mu_j)(o_t - \mu_j)^{\mathrm{T}}}{\sum_{t=1}^{T} L_j(t)}. \tag{4}$$

Вычисление значений $L_j(t)$ осуществляется с помощью алгоритма прямого-обратного хода: согласно формуле 6 вычисляются значения прямых вероятностей $\alpha_j(t)$, согласно формуле 8 – значения обратных вероятностей $\beta_j(t)$.

$$\alpha_j(t) = \mathrm{P}(o_1, o_2, ..., o_t, x(t) = j \mid M), \tag{5}$$

$$\alpha_j(t) = \left[\sum_{i=2}^{N-1} \alpha_i(t-1) a_{ij} \right] b_j(o_t), \tag{6}$$

$$\beta_j(t) = \mathrm{P}(o_t, o_{t+1}, ..., o_T, x(t) = j \mid M), \tag{7}$$

$$\beta_i(t) = \sum_{j=2}^{N-1} a_{ij} b_j(o_{t+1}) \beta_j(t+1). \tag{8}$$

Искомое значение $L_j(t)$ может быть вычислено следующим образом:

$$L_j(t) = \mathrm{P}(x(t) = j \mid O, M) = \frac{P(O, x(t) = j \mid M)}{\mathrm{P}(O \mid M)} = \frac{\alpha_j(t)\beta_j(t)}{\alpha_N(T)}. \tag{9}$$

Процедуру обучения моделей HMM можно представить в виде итеративного процесса, на каждой итерации которого осуществляется расчет прямых $\alpha_j(t)$ и обратных $\beta_j(t)$ вероятностей для всех состояний модели j и моментов времени t. На основе вероятностей вычисляются значения $L_j(t)$, с помощью которых происходит обновление параметров Гауссиан, описывающих состояния HMM, согласно формулам 3 и 4. Процедура продолжается до тех пор, пока не будут с достаточной для каждой конкретной задачи точностью установлены параметры моделей.

Для перехода от фонетического уровня представления сигнала к представлению в виде отдельных слов используется лексическая модель языка. Данная модель содержит список слов, распознавание которых должно поддерживаться

системой, а также транскрипции данных слов на основе выбранных базовых акустических единиц. В случае небольшого количества слов (малого словаря распознавания) фонетические транскрипции могут быть составлены экспертным путём, однако для словаря в несколько тысяч слов необходимо разработка так называемых автоматических графем-фонемных систем преобразования, которые позволяют строить фонетические транскрипции для произвольного текста.

Языковая модель призвана описывать существующие в языке закономерности и на их основе уметь оценивать вероятности произнесения конкретных последовательностей слов. В случае моделирования закономерностей языка набором грамматических правил, описывающих структуру возможных в контексте данной предметной области фраз, языковая модель получает способность оценивать возможность произнесения определенной последовательности слов.

Этап распознавания является заключительным и основывается на информации о построенных моделях. Происходит построение полной HMM для распознавания: модели слов собираются из моделей отдельных фонем согласно транскрипциям с добавлением состояния «конец слова», вероятности переходов между словами определяются на основе языковой модели.

Пример того, как может выглядеть полная HMM для задачи распознавания слов «да» и «нет» представлен на рисунке 3. В данном примере предполагается возможность многократного произнесения слов; наличие паузы между словами является необязательным. Модель каждой фонемы представляет собой HMM, состоящую из трёх состояний, что отображено на рисунке 3 для фонемы «е». Таким образом, полная HMM представляется одновременно на трёх различных уровнях: уровне слов, фонем и состояний моделей.

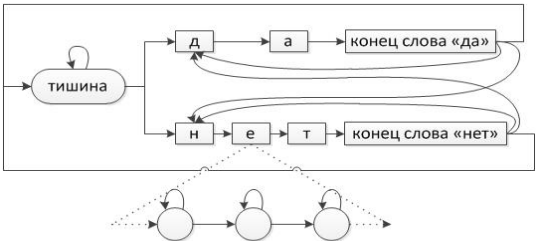

Рисунок 3 – Представление HMM для слов «да» и «нет»

Поиск наиболее вероятной последовательности состояний скрытой Марковской модели осуществляется на основе алгоритма Витерби, схожего с вычислением прямых вероятностей метода прямого-обратного хода. Отличительной особенностью является замена операции суммирования в формуле 6 на операцию поиска максимума:

$$\phi_j(t) = \max_i \{\phi_i(t-1)a_{ij}\}b_j(o_t). \quad (10)$$

В качестве начальных значений используются следующие:

$$\begin{aligned} &\phi_1(1) = 1, \\ &\phi_j(1) = a_{1j}b_j(o_1),\ 1 < j < N \end{aligned}. \quad (11)$$

Результатом работы алгоритма является значение вероятности наблюдения последовательности векторов признаков O в рамках данной модели M – $\hat{\mathrm{P}}(O \mid M)$:

$$\hat{\mathrm{P}}(O \mid M) = \phi_N(T) = \max_i \{\phi_i(T)a_{iN}\}. \quad (12)$$

Для того чтобы избежать потери точности вычислений, связанной с многократным перемножением значений вероятностей, расчеты производятся с логарифмами значений, следовательно, уравнение 10 записывается в следующем виде:

$$\psi_j(t) = \max_i\{\psi_i(t-1) + \log(a_{ij})\} + \log(b_j(o_t)). \quad (13)$$

В контексте распознавания речи используется модифицированная версия описанного алгоритма под названием «token passing», суть которого заключается в использовании специальных маркеров. Каждый маркер представляет собой пару ($\psi_j(t)$, link), хранящую последовательность состояний модели и значение логарифма вероятности данной последовательности. В начальный момент времени маркеры помещаются во все возможные начальные состояния HMM. Далее, в каждый момент времени все маркеры переходят в следующие состояния согласно определённым в HMM связям, обновляются значения текущего пути, пройденного каждый маркером, а значение логарифма увеличивается на величину вероятности осуществленного перехода и согласно вероятностям нового состояния:

$$\psi_j(t) = \psi_i(t-1) + \log(a_{ij}) + \log(b_j(o_t)). \quad (14)$$

В момент обработки последнего входящего вектора признаков выбирается маркер, которому соответствует наибольшая вероятность, набор состояний модели, хранящийся в его истории, считается результатом распознавания.

Заключительным этапом работы алгоритма распознавания служит оценка качества работы системы на основе следующих характеристик:

1. Корректность распознавания – Corr (correctness):

(15)

где N – число элементов в правильной транскрипции, D – количество пропусков, S – количество замен.

2. Точность распознавания – Acc (accuracy):

$$\mathbf{Acc} = \frac{N - D - S - I}{N}, \quad (16)$$

где I - количество включений лишних элементов.

3. Скорость распознавания – RTF (real time factor):

$$\mathbf{RTF} = \frac{P}{T}, \quad (17)$$

где P –время обработки речевого фрагмента, T – продолжительность фрагмента.

В третьей главе приводится описание программной реализации комплекса автоматизации создания и оценки качества программных средств распознавания речи для малоресурсных языков. Для апробации созданного программного комплекса, построенного на основе разработанных моделей, а также для оценки преимуществ его использования, на основе комплекса были реализованы программные средства распознавания фонем и слитной речи для татарского языка.

Модель структуры комплекса автоматизации создания и оценки качества программных средств распознавания речи для малоресурсных языков представлена на рисунке 4.

Приводится описание программных средств распознавания речи для малоресурсных языков, созданных на базе предложенного комплекса. Программные средства распознавания представлены в виде проекта из 6 модулей, каждый из которых может быть доработан и повторно использован для распознавания конкретного языка. В качестве основы для распознавания был использован описанный в главе 2 подход на основе аппарата HMM. Основными элементами выбранного подхода являются три типа моделей: модель произношения, акустические модели и языковая модель.

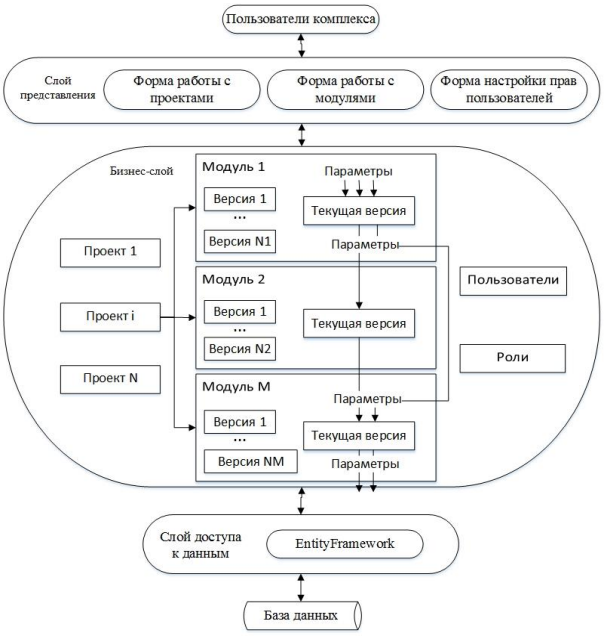

Рисунок 4 – Структура комплекса автоматизации

Структура программных средств распознавания речи для малоресурсных языков представлена на рисунке 5.

Рисунок 5 – Структура программных средств распознавания слитной речи

Модули в проекте распознавания речи решают следующие задачи:

1. Модуль «Акустические особенности языка» – модуль, позволяющий задать алфавит языка, фонем, правила фонетического транскрибирования текстов;
2. Модуль «Анализ текстов» – модуль фонетического транскрибирования;
3. Модуль «Запись» – модуль записи речевого корпуса;
4. Модуль «Акустические модели» – модуль создания и обучения моделей HMM для каждой фонемы языка;
5. Модуль «Грамматика» – модуль создания языковой модели на основе грамматических правил, описывающих допустимые фразы;
6. Модуль «Распознавание речи» – модуль, осуществляющий распознавание речевых фрагментов и позволяющий оценить качество работы системы.

С целью анализа эффективности комплекса создания и оценки качества программных средств распознавания речи приводятся результаты его апробации в контексте создания средств распознавания фонем и слитной татарской речи.

Решение задачи распознавания татарских фонем осуществлялось на основе комплекса автоматизации и созданного набора модулей. В комплексе был создан проект «Распознавание фонем татарского языка». В него добавлены модули «Акустические особенности языка», «Анализ текстов», «Запись», «Акустические модели» и «Распознавание фонем». В результате работы с модулем «Акустические особенности языка» был задан алфавит из 39 символов, выделены 57 фонем татарского языка, сформирован список из 37 правил фонетической транскрипции. Был сформирован список слов для озвучивания, подготовлены их фонетические транскрипции, а также произведена запись речевого корпуса одного диктора длительностью 5 часов. Созданные на основе корпуса акустические модели позволили достичь на тестовой части корпуса 61%-го качества распознавания фонем.

В данном исследовании была также решена задача создания базового алгоритма дикторонезависмого распознавания слитной татарской речи со словарём среднего размера. Для решения данной задачи в рамках комплекса был создан проект «Распознавание слитной речи» и решены следующие подзадачи:

1. Создание татарского текстового корпуса (776 предложений, 6913 слов);
2. Проектирование и запись корпуса татарской речи (251 диктор, 8 часов);
3. Обучение акустических моделей фонем татарского языка;
4. Построение грамматических правил предметной области (1135 слов);
5. Распознавание слитной речи на татарском языке.

Для распознавания была использована тестовая часть созданного корпуса. Был загружен список произнесённых фраз, что позволило оценить качество распознавания. Значения рассчитанных коэффициентов Corr, Acc и RTF представлены в таблице 2. Полученные значения качества распознавания 77 % и 67 % соответствуют результатам, продемонстрированным системами распознавания данного типа для других языков; значение скорости работы, равное 0,35 RT, говорит о возможности программных средств работать в режиме реального времени.

Таблица 2 – Тестирование программных средств распознавания слитной речи

Параметр	Значение
Скорость распознавания, RTF	0,35 RT
Процент корректно распознанных слов, Corr	76,69 % (2583 слов из 3368)
Точность распознавания, Acc	67,31 % (2267 слов из 3368)
Процент ошибок типа «Замена»	21,82 % (735 слов из 3368)
Процент ошибок типа «Пропуск»	1,48 % (50 слов из 3368)
Количество ошибок типа «Вставка»	316 слов

Комплекс автоматизации в процессе распознавания занимает 63 МБ оперативной памяти и 13 % CPU, объем БД составляет 40 МБ. Результаты были получены при работе комплекса автоматизации на компьютере со следующими характеристиками: процессор Intel Core i7-2630QM, 2 GHz, ОЗУ 6GB, Windows 7 x64.

Оценка эффективности использования комплекса автоматизации была проведена на основе экспертной оценки трудозатрат на создание базовых средств распознавания речи. Отмечено существенное снижение объема работ экспертов в области речевых технологий и программистов; наличие инструментария позволяет специалистам в области языка выделять существенные особенности языка.

В заключении диссертации изложены основные результаты работы.

ОСНОВНЫЕ ВЫВОДЫ И РЕЗУЛЬТАТЫ РАБОТЫ

1. Разработана технология построения программных средств распознавания речи, основанная на использовании моделей, учитывающих специфику обработки малоресурсных языков и обеспечивающих возможность одновременной работы экспертов в области языка, анализа речи, программистов и других специалистов в процессе проектирования и верификации прикладных систем распознавания речи.

2. Разработана модель комплекса автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков, отличающаяся охватом всех основных подзадач области распознавания речи, а также возможностью их настройки для работы с конкретным малоресурсным языком, что позволяет существенно ускорить процесс создания программных средств анализа речи для малоресурсных языков.

3. Разработан программный комплекс автоматизации создания и оценки качества программных средств анализа речи для малоресурсных языков и инструментальные средства выполнения алгоритмов автоматического анализа речи малоресурсных языков, основанные на использовании разработанной технологии, обеспечивающей существенное ускорение процесса создания программного обеспечения анализа речи малоресурсных языков при сохранении качества и скорости его работы на уровне мировых аналогов.

4. Впервые созданы программные средства распознавания фонем и слитной речи на татарском языке на базе разработанного программного комплекса автоматизации, позволяющие использовать их для обеспечения речевого интерфейса взаимодействия человека с компьютером. Установлено, что созданные программные средства анализа речи корректно распознают 76,7 % слов, точность распознаваний составляет 67,3 %, скорость обработки – 0,35 RT.

5. Проведено исследование эффективности разработанного комплекса средств автоматизации создания и оценки качества программных средств распознавания речи для малоресурсных языков, продемонстрировавшее снижение объема работы экспертов в области речевых технологий и программистов, а также сокращение общего времени создания базовых систем анализа речи малоресурсных языков на 52,5 % при сохранении качества и скорости их работы на уровне мировых аналогов.

Перспективы дальнейшей разработки темы. В рамках дальнейших исследований планируется развитие модели и программной реализации комплекса за счет включения в неё дополнительных подсистем анализа речи для малоресурсных языков.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

В рецензируемых журналах списка ВАК

1. Language Identification System for the Tatar Language / A. F. Khusainov, D. Sh. Suleymanov // Speech and Computer, Lecture Notes in Computer Science. 2013. Volume 8113. P. 203–210.
2. Программный комплекс для анализа речи (на примере распознавания фонем татарского языка) / А. Ф. Хусаинов // Доклады Томского государственного университета систем управления и радиоэлектроники. 2013. № 3 (29). С. 129–133.
3. Система автоматического распознавания речи на татарском языке / А. Ф. Хусаинов, Д. Ш. Сулейманов // Программные продукты и системы. 2013. № 4. С. 301–304.

Свидетельства об официальной регистрации программ для ЭВМ

4. Свидетельство о государственной регистрации программы для ЭВМ № 2014610952. Платформа создания и исполнения систем анализа речи / А. Ф. Хусаинов. Зарег. 21.01.2014. М.: Роспатент, 2014.

В других изданиях

5. Аналитический обзор подходов распознавания речи / А. Ф. Хусаинов // Материалы 14-й междунар. конф. SPECOM-2011. (Казань, 27-30 сентября, 2011). М.: Типография «ПАЛАДИН», 2011. С. 151–158.
6. Создание системы автоматического распознавания изолированных команд на татарском языке в инфокоммуникационной среде / А. Ф. Хусаинов // Корпусы национальных языков: модели и технологии: материалы XII Казанской школы-семинара TEL’2012. (Казань, 25-28 января, 2012). Казань: Фэн, 2012. С. 131–141.
7. Прототип платформы анализа речи на татарском языке / А. Ф. Хусаинов, Д. Ш. Сулейманов // Открытые семантические технологии проектирования интеллектуальных систем: материалы 3-й международной научно-технической конференции (OSTIS'2013). (Минск, 21-23 февраля, 2013). Минск: БГУИР, 2013. С. 361–368.
8. Speech analysis platform for Tatar language / A. Khusainov // Proc. of the 15th International Workshop on Computer Science and Information Technologies (CSIT'2013). (Vienna, Budapest, Bratislava, September 15-21, 2013). Ufa: USATU, 2013. P. 75–80.
9. Система автоматического распознавания фонем татарского языка / А. Ф. Хусаинов // Компьютерная обработка тюркских языков: труды первой международной конференции. (Астана, 3-4 октября, 2013). Астана: ЕНУ им. Л. Н. Гумилева, 2013. С. 211–217.
10. Система распознавания речи на татарском языке / А. Ф. Хусаинов // Языковая семантика - модели и технологии: материалы конференции с международным участием по когнитивной лингвистике (TEL-2014). (Казань, 20-22 февраля, 2014). Казань: Фэн, 2014. С. 20–24.

Диссертант А. Ф. Хусаинов