

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.93

Уласевич
Сергей Александрович

Модели, методы и программное средство распознавания и идентификации
речи

АВТОРЕФЕРАТ

на соискание академической степени
магистра технических наук

по специальности 1-40 80 05 – Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

Научный руководитель
Серебряная Л.В.
к.т.н., доцент

Минск 2015

Библиотека БГУИР

КРАТКОЕ ВВЕДЕНИЕ

На сегодняшний день, под понятием «распознавание речи» скрывается целая сфера научной и инженерной деятельности. Каждая задача распознавания речи сводится к тому, чтобы выделить, классифицировать и соответствующим образом отреагировать на человеческую речь из входного звукового потока. Работы по распознаванию речи не только не потеряли актуальности, но и развиваются широким фронтом, находя для себя множество областей для практического применения. Так, можно выделить четыре сравнительно изолированных направления в области развития речевых технологий: распознавание речи, определение индивидуальности говорящего, синтез речи и компрессия речи.

Предполагалось, что когда компьютер научится понимать человеческую речь, мы быстро сможем создать искусственный интеллект. Но точность систем распознавания речи достигла своего пика в 1999 году и с тех пор застыла на месте. Академические тесты 2014 года констатируют факт: системы общего профиля так и не преодолели уровень 80%, тогда как у человека этот показатель составляет 96-98%. На данный момент существует ряд нерешенных проблем, такие как невозможность подавления внешних шумов, проблемы с точностью распознавания и низкий уровень безопасности в верификации диктора.

В результате проведения анализа существующих моделей, алгоритмов и программных средств (ПС) распознавания речи в работе были выделены вышеуказанные проблемы организации ПС распознавания и идентификации речи. Исходя из них, разработаны модели и алгоритмы организации систем распознавания речи, предлагающие решения данных проблем. Разработанные модели и алгоритмы реализованы в ПС.

Учитывая сформированные общие проблемы, сформулирована система критериев оценки эффективности ПС распознавания. Сформулированная система критериев использована для анализа результатов экспериментов, проводимых с разработанным и исследованными ПС, с целью проверки улучшения значений критериев оценки эффективности у разработанного ПС по сравнению с исследованными ПС распознавания речи. Так как, в результате проведения экспериментов достигнуто улучшение значений критериев оценки эффективности у реализованного ПС, то сделан вывод о том, что разработанные модели и алгоритмы организации ПС распознавания и идентификации речи предлагают более эффективные решения вышеуказанных общих проблем области распознавания речи.

В качестве исходных данных для проведения исследования использовались научная литература по рассматриваемой теме, ПО для создания проектируемого ПС, также для проведения экспериментов с исследуемыми и разработанным ПС распознавания и идентификации речи.

Таким образом, тема работы обоснована и актуальна в научном (создание моделей и алгоритмов организации систем распознавания и идентификации речи, построение методологии оценки работы эффективности ПС распознавания идентификации речи) и практическом плане (эксперименты,

подтверждающие улучшение значений критериев методологии оценки эффективности у ПС, разработанного на основе созданных моделей и алгоритмов по сравнению с исследованными ПС распознавания и идентификации речи).

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является разработка алгоритмов и программного обеспечения для решения задач по распознаванию и идентификации речи и проведения практической проверки наличия преимуществ разработанных моделей и алгоритмов в рамках созданного на их основе программного средства (ПС) распознавания и идентификации речи над уже существующими моделями и алгоритмами.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Исследовать существующие модели, алгоритма и программные средства распознавания и идентификации речи с целью определения проблем распознавания и идентификации речи.
2. Разработать модели и методы распознавания и идентификации речи.
3. Разработать программное средство на основе созданных моделей и методов распознавания и идентификации речи.
4. Построить методологию оценки работы программного средства распознавания и идентификации речи, опираясь на выявленные проблемы.
5. Провести экспериментальные исследования разработанного программного средства.

Объектом исследования являются область распознавания и идентификации речи.

Предметом исследования является моделирование и алгоритмизация распознавания и идентификации речи.

Основной гипотезой, положенной в основу диссертационной работы, является возможность улучшения эффективности распознавания и идентификации речи у разработанного метода по сравнению с существующими методами на словаре малого объема.

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Работа выполнялась в соответствии с научно-техническим заданием и планом работ кафедры «Программное обеспечение информационных технологий» по теме «Разработать модели, методы, алгоритмы для оценки параметров, повышения надежности и качества функционирования аппаратно-программных средств систем и сетей сложной конфигурации и внедрить в со-

временные обучающие комплексы» (ГБ № 11-2004, № ГР 20111065, научный руководитель НИР – В. В. Бахтизин).

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя Л. В. Серебряной, заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Часть положений диссертационной работы докладывались и обсуждались на 11-й Международной молодёжной научно-технической конференции «Современные проблемы радиотехники и телекоммуникаций РТ-2015» (Севастополь, Российская Федерация, 2015).

Опубликованность результатов диссертации

По теме диссертации опубликована одна работа в сборниках трудов и материалов международных конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора и приложений. В первой главе представлен анализ предметной области, выявлены основные существующие проблемы в рамках тематики исследования, показаны направления их решения. Вторая глава посвящена разработке моделей и алгоритмов распознавания и идентификации речи. В третьей главе рассмотрена методология оценивания работы ПС распознавания и идентификации речи и построена система критериев оценки работоспособности данного ПС, разработано ПС на основе созданных моделей и алгоритмов распознавания и идентификации речи; проведены эксперименты по проверке улучшения значений критериев оценки работы у разработанного ПС по сравнению с исследованными ПС. Общий объем работы составляет 62 страниц, из которых основного текста – 52 страниц, 10 рисунков на 10 страницах, 3 таблиц на 2 страницах, список использованных источников из 40 наименований на 4 страницах и 1 приложения на 3 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В первой главе проведен анализ литературы по теме распознавания и идентификации речи. В рамках данного анализа, рассмотрена предметная область распознавания и идентификации речи. Сформулированы основные понятия и определения в данной области, в частности определения понятий: «распознавание речи», «синтез речи», «компьютерное клонирование голоса и дикции», «понимание (интерпретация) речи», «распознавание голоса», «анализ и распознавание речи».

Выполнен анализ основных моделей распознавания и идентификации речи: объектно-ориентированная модель, агентно-ориентированная модель и сервис-ориентированная модель. В ходе проведенного анализа выявлены основные достоинства и недостатки рассматриваемых моделей.

Проанализированы методы распознавания и идентификации речи, именно: метод скрытых марковских моделей и метод динамической трансформации временной шкалы.

Выполнен анализ двух ПС распознавания и идентификации речи CMU Sphinx и Julius. Выбор данных средств обусловлен возможностью анализа моделей и методов за счет открытого программного кода и возможности установки данных ПС для экспериментов по сравнению характеристик с ПС, разрабатываемым в третьей главе.

CMU Sphinx – это дикторонезависимый распознаватель непрерывной речи, который использует скрытую марковскую модель и n-граммную статистическую языковую модель.

Julius – это высокопроизводительный распознаватель непрерывной речи с большим словарем, декодер программного обеспечения для исследования в области связанной речи и разработки.

Вторая глава посвящена разработке моделей и методов распознавания и идентификации речи.

На основе веб-модели и модели клиент-сервер разработана многослойная модель распознавания и идентификации речи, описывающая распределение функциональных возможностей по слоям. Разработанная модель состоит из четырех основных слоев: слой веб-приложения, слой интерфейса программирования веб-приложения, слой логики предметной области, слой доступа к данным, слой обеспечения надежности и слой поддержки безопасности. Данная модель представлена на рисунке 1.

Разработана спецификация требований для разработки ПС распознавания и идентификации речи, дополненная диаграммой прецедентов.

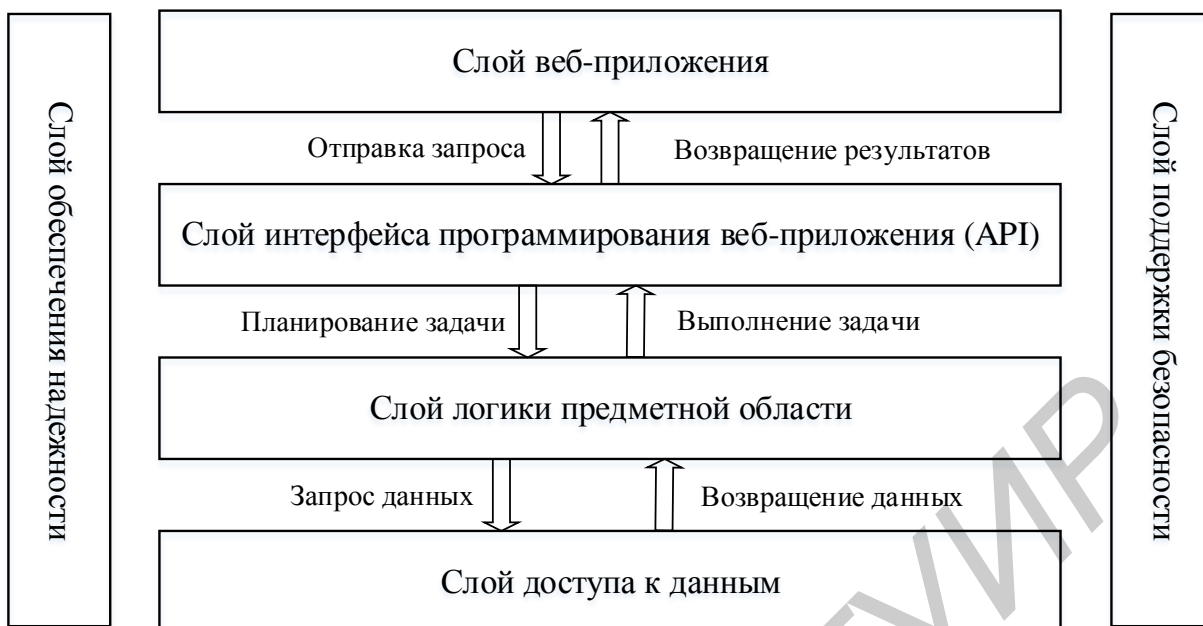


Рисунок 1 - Многослойная модель организации распознавания речи

Разработана структура ПС распознавания и идентификации речи, состоящая из таких блоков: необработанная речь, анализ сигнала, речевой кадр акустический анализ, акустические модели, множество кадров, корректировка времени, сегментация и последовательность слов. Данная структура представлена на рисунке 2.

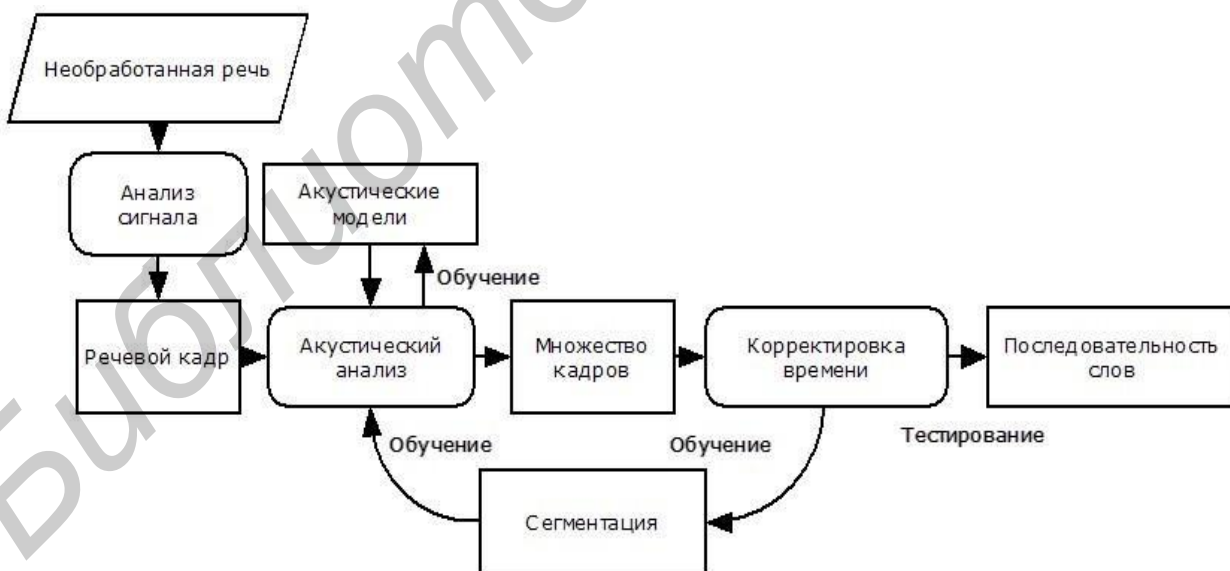


Рисунок 2 – Структура ПС распознавания речи

Предложены методы фильтрации шумов и распознавания речи. Определены достоинства и недостатки данных методов. Достоинства метода фильтрации шума: наименьшая потеря полезного сигнала, возможность эффективного устранения избыточности в частотной области, эффективность

фильтрации шумов с большой длительностью. Метод распознавания и идентификации речи имеет три особых преимущества:

1. Матрица трансформации представляется с помощью разреженных матриц, что приводит к уменьшению средней пространственной сложности по сравнению с другими методами.

2. Разреженный метод динамической трансформации временной шкалы всегда выдает оптимальный путь трансформации.

3. Так как метод выдает оптимальное выравнивание, то он может быть легко использован в сочетании с другими методами.

Определен главный недостаток данного метода - $O(n^2v)$ сложность не удовлетворяет большим словарям.

В третьей главе осуществлено разработка ПС и проведены экспериментальные исследования. Были выбраны наиболее подходящие инструменты разработки, технологии и языки программирования. Разработана диаграмма классов, представленная на рисунке 3.

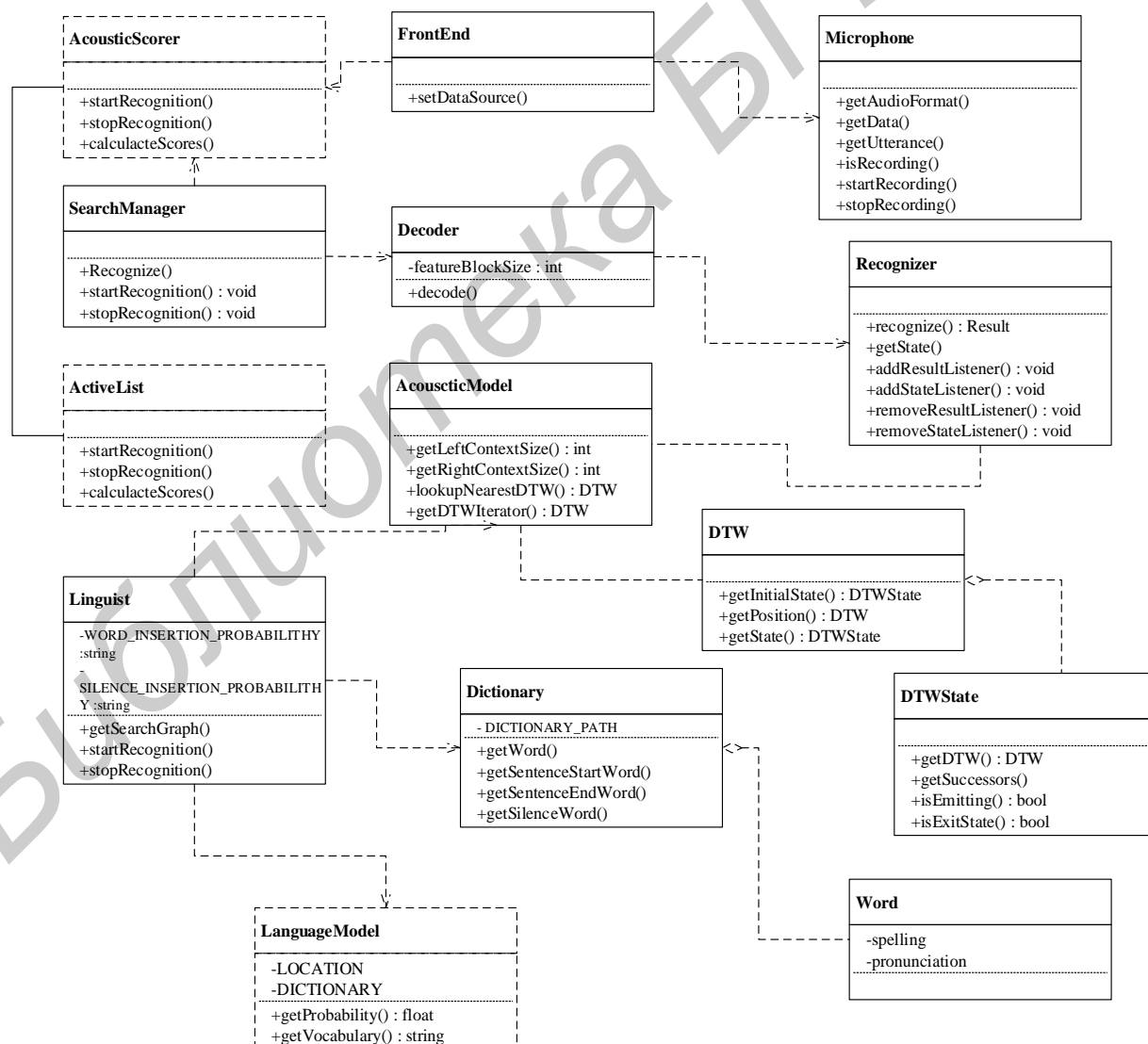


Рисунок 3 – Диаграмма классов

Методология оценивания работы программного средства распознавания и идентификации речи. В ходе, которые были сформулированы основные критерии эффективности работы, а именно: критерий WRR процент правильно распознанных слов и критерий SF скорость распознавания и идентификации речи.

Было проведено два эксперимента на двух словарях разного объема (малого и большого). В ходе первого эксперимента разработанное ПС показала улучшение критериев WRR и SF относительно рассматриваемых ПС в первой главе. В ходе второго эксперимента критерий WRR был хуже у разработанного ПС относительно рассмотренных в первой главе ПС распознавания и идентификации речи. Из этого следует, что разработанное ПС показывает улучшение показателей на словарях малого объема.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Проведен анализ основных моделей организации ПС распознавания и идентификации речи: объектно-ориентированная модель, веб-модель, модель «клиент-сервер», агентно - ориентированная модель, сервис - ориентированная модель.

2. Показано, что веб-модель непригодна для использования в условиях, когда требуется независимость от наличия постоянного соединения с интернетом и независимость от решений сторонних производителей (хостинг, предоставление вычислительных мощностей). Модель «клиент-сервер» обладает следующими недостатками: сложность обеспечения масштабируемости, зависимость производительности выполнения распознавания и идентификации речи от доступности и надежности сервера и повышенные требования к нему. Объектно-ориентированная модель характеризуется такими недостатками, как снижение производительности из-за использования промежуточных объектов-посредников и сериализации, десериализации параметров методов; необходимость в механизмах регистрации источников удаленных объектов на промежуточных серверах или на клиентах. Для агентно - ориентированной модели определены недостатки: сложность обеспечения безопасного взаимодействия между агентом(-ами) и средой функционирования, необходимость поддержки агентом протоколов коммуникации и обмена полезной информацией в мультиагентных системах и потребление ресурсов системы агентами, чьи цели противоречат общесистемным целям. Сервис - ориентированная модель непригодна, если требуется высокая начальная скорость и невысокая конечная стоимость разработки распознавания и идентификации речи на основе данной модели.

3. Проанализированы методы распознавания и идентификации речи: метод скрытых марковских моделей и метод динамической трансформации временной шкалы. Установлены недостатки метода скрытых марковских моделей – большая вычислительная сложность и необходимость подробного дополнительного моделирования посторонней речи. Недостаток метода динамической

трансформации временной шкалы - $O(n^2v)$ сложность не удовлетворяет большим словарям, которые увеличивают успешность процесса распознавания. Кроме того, трудно вычислить два элемента в двух разных последовательностях, если принять во внимание, что существует множество каналов с различными характеристиками.

4. Проведен анализ двух образцов открытых программных средств распознавания и идентификации речи: CMU Sphinx и Julius. В CMU Sphinx выявлены следующие недостатки - трудный процесс обучения акустической модели. В ПС Julius на данный момент отсутствует достаточное количество подготовленных акустических моделей для разных языков, что делает ее использование для различных языков недостаточно эффективным

5. В результате проведенного анализа существующих моделей, алгоритмов и программных средств (ПС) распознавания и идентификации речи выявлены общие проблемы – проблемы обеспечения производительности на малых словарях и трудности в обучении системы.

6. Разработана многослойная модель организации распознавания и идентификации речи. Разработанная модель состоит из четырех основных слоев: слой веб-приложения, слой интерфейса программирования веб-приложения, слой логики предметной области, слой доступа к данным, слой обеспечения надежности и слой поддержки безопасности.

7. Разработана спецификация программного средства распознавания и идентификации речи, а также его архитектура.

8. Предложены методы фильтрации шумов и распознавания речи.

9. Выбрана технология и среда разработки ПС.

10. На основе созданных моделей и методов реализовано ПС распознавания и идентификации речи.

11. Построена методология оценивания результатов работы ПС распознавания и идентификации речи.

12. Проведены эксперименты, в ходе которых были выявлены достоинства и недостатки разработанного ПС.

13. Экспериментально подтверждена гипотеза исследований о том, что будет наблюдаться улучшения эффективности идентификации и распознавания речи у разработанного метода по сравнению с существующими методами на словаре малого объема.

Рекомендации по практическому использованию результатов

Полученные результаты формируют теоретическую и практическую базу для разработки ПС распознавания и идентификации речи. В дополнение к этому, разработанные модели и методы распознавания и идентификации речи, могут быть использованы для модернизации и дальнейшего развития существующих ПС.

Разработанные модели и методы распознавания и идентификации речи могут применяться для создания ПС, работающих с малыми словарями данных.

Результаты работы представляют собой экономическую ценность – разработанное ПС может быть использовано для решения задачи «умного дома»

Результаты работы могут использоваться при подготовке персонала для разработки и обслуживания ПС распознавания и идентификации речи, решающего задачи проектирования и модернизации данных ПС.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1-А. Уласевич, С.А. Методы распознавания речи / С.А. Уласевич // Современные проблемы радиотехники и телекоммуникаций «РТ-2015»: материалы 11-й междунар. молодежной науч.-техн. Конф.– Севастополь, 2015 – с. 170.

Библиотека БГУИР