

УДК 004.52

А.Ф. Хусаинов

Программный комплекс для анализа речи (на примере распознавания фонем татарского языка)

Предложен и реализован в виде программного комплекса подход к проектированию, созданию и управлению системами анализа речи. В его состав входят как реализованные средства решения стандартных задач (например, построения систем транскрибирования, проектирования и записи речевых корпусов), так и возможности изменения и добавления новых компонент. Эффективность применения предложенного комплекса при построении систем анализа речи продемонстрирована на примере создания системы автоматического распознавания фонем татарского языка.

Ключевые слова: анализ речи, средство проектирования, распознавание фонем, татарский язык.

В настоящее время в мире насчитывается около пяти тысяч языков, из которых около ста активно используются более чем 80% населения планеты. В то же время происходит активное развитие систем речевого интерфейса взаимодействия человека с компьютерами. Однако степень развития данных технологий сильно отличается между различными языками, и высокое качество работы речевых систем, например, для английского и немецкого языков, сочетается с полным отсутствием данных технологий для многих языков регионального масштаба. Данная ситуация объясняется как научной сложностью стоящих задач, так и высокими финансовыми затратами на подготовку необходимых программных инструментов и речевых корпусов [1]. В настоящий момент существуют проекты, направленные на преодоление данной ситуации, однако чаще всего они решают узкие задачи и не предоставляют достаточной свободы при создании средств анализа речи [2].

В данной работе описывается комплекс программных средств, позволяющий осуществлять проектирование систем анализа речи для различных языков. Для этого могут быть использованы уже имеющиеся в комплексе инструменты (выбор инвентаря фонем языка, описание акустических правил языка, средства проектирования и записи корпусов). Кроме того, существует возможность добавления новых и расширения функциональности уже имеющихся модулей комплекса. В рамках предложенной системы проектирования разработана и реализована система автоматического распознавания фонем татарского языка. На её примере показана эффективность предложенных программных средств при разработке систем анализа речи для новых языков.

Комплекс проектирования систем анализа речи. Предлагаемый комплекс проектирования систем анализа речи обладает целым рядом свойств, наличие которых продиктовано особенностями процессов создания и функционирования данных систем [3]:

1. Модульная архитектура комплекса, наличие механизма расширения функциональности модулей.

Область речевых технологий представляет собой совокупность множества различных направлений анализа речи, включая такие направления, как синтез речи, распознавание команд слитной речи, идентификация диктора, языка, эмоций, качества произношения [4]. Данные направления используют различные алгоритмы и подходы, однако с точки зрения алгоритмов вычисления характеристик речи, а также способов построения статистических моделей звуков или языков они имеют много общего, что делает оправданным выделение обособленных частей алгоритма в отдельные модули и их совместное использование в различных системах.

Кроме того, следует отметить высокую взаимосвязанность различных систем анализа речи между собой [5, 6]. Так, к примеру, система распознавания фонем языка может использовать информацию от модуля идентификации языка для настройки акустических моделей на нужный язык; в то же время модуль идентификации языка в качестве исходной информации для работы может принимать данные от модуля распознавания фонем. Наличие данных взаимосвязей делает необходимой возможность обмена информацией между целыми модулями систем анализа речи, что позволит повысить качество их работы.

Для реализации модульной архитектуры в комплексе проектирования были реализованы и используются следующие сущности:

а. Модуль – базовая сущность комплекса, решающая строго определенную, обособленную часть задачи, например вычисление вектора признаков речи или создание речевого корпуса. Модули реализуют модель «черного ящика»; взаимодействие модулей друг с другом происходит посредством передачи значений входных и выходных параметров. При этом параметры могут иметь как строго заданные значения (доступные типы данных: число, строка, дата, объект), так и являться указателями на значение параметра другого модуля, что позволяет реализовать возможность обмена необходимой информацией между модулями. С точки зрения программной архитектуры каждый модуль представляет собой реализацию базового интерфейса, посредством которого происходит управление модулем со стороны программного комплекса. Описание функций данного интерфейса представлено в табл. 1.

Таблица 1

Интерфейс управления модулями системы

Название функции	Параметры	Назначение
GetParameters	Входные: – Выходные: список параметров модуля с их значениями по умолчанию	Комплекс вызывает данную функцию для получения списка входных и выходных параметров конкретного модуля
Initialize	Входные: список входных параметров с их значениями Выходные: успешность инициализации модуля	Данная команда сообщает модулю о необходимости настройки параметров работы в соответствии со значениями переданных параметров. Выходное значение говорит о готовности модуля к запуску
Start	Входные: родительская оконная форма, в которой необходимо разместить модуль Выходные: –	Вызов данной команды инициирует запуск модуля в указанном родительском окне
Execute	Входные: – Выходные: –	Аналог функции Start для запуска модуля без необходимости отображения пользовательского интерфейса
Save	Входные: – Выходные: список выходных параметров с их значениями	Производит сохранение значений выходных параметров
CloseForm	Входные: – Выходные: –	Прекращает работу модуля

б. Проект – сущность, объединяющая модули для решения конкретной задачи. Визуальное отображение проекта позволяет легче отслеживать зависимости между модулями, а также производить их настройки в рамках решения данной задачи. Кроме того, проект содержит необходимую информацию о версиях и о правильной последовательности запуска модулей.

2. Возможность настройки прав доступа для различных пользователей.

Создание систем анализа речи представляет собой комплексную междисциплинарную задачу, в которой необходимы знания по акустике, фонетике, лингвистике, математической статистике и моделированию, программированию. Поэтому большинство речевых систем разрабатывается коллективами ученых из разных областей знаний. Этим продиктована необходимость внедрения системы разделения и контроля прав доступа к различной функциональности комплекса для разных пользователей. К примеру, в проекте по распознаванию фонем языка администратор проекта может предоставить лингвисту доступ на изменение параметров исключительно модуля по формированию текстового корпуса, фонетисту – модуля определения правил транскрибирования и т.д. Таким образом, гарантируется сохранение целостности и работоспособности системы за счет точного распределения прав в команде исследователей.

3. Возможность сравнения различных конфигураций работы системы.

Возможность предоставления прав доступа к отдельным модулям для специалистов в узких областях знаний при этом не ограничивает их способность контролировать работу всей системы целиком. Так, отдельно могут быть прописаны права на запуск каждого модуля (метод Execute описанного выше интерфейса) без возможности его редактирования. Таким образом, например,

специалист, занимающийся редактированием алгоритма вычисления вектора признаков, может контролировать качество работы всей системы распознавания в зависимости от вносимых им изменений в конкретный модуль.

4. Предоставление реализаций базовых алгоритмов решения стандартных задач.

Одной из сложностей, возникающих в процессе создания систем анализа речи для множества новых языков, является необходимость разработки и реализации большого количества вспомогательного инструментария и базовых алгоритмов решения стандартных задач. В предлагаемом программном комплексе на данный момент реализованы и доступны для использования следующие инструменты:

- а. Создание фонемного алфавита для конкретного языка.
- б. Описание акустических правил языка.
- в. Контроль процесса разработки текстового корпуса.
- г. Проектирование и запись речевого корпуса.

Общая структура комплекса представлена на рис. 1.

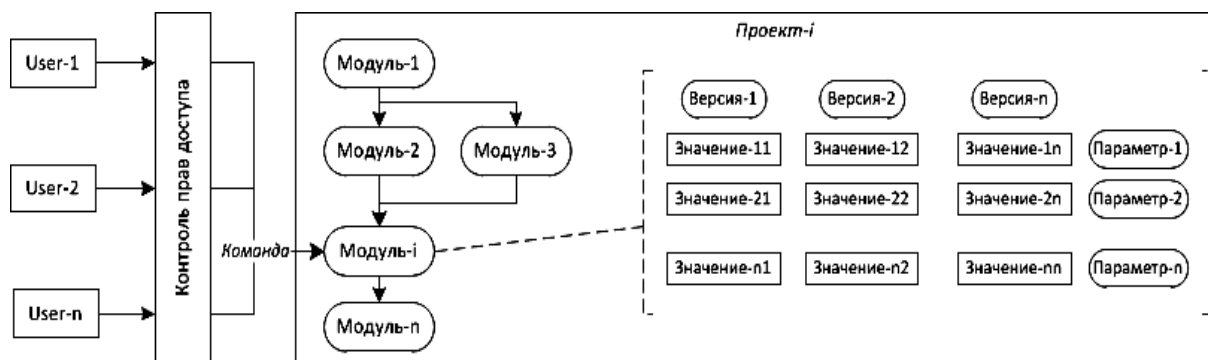


Рис. 1. Общая структура комплекса

Система распознавания фонем татарского языка

С помощью созданного комплекса, а также реализованного на его основе инструментария была разработана система распознавания фонем татарского языка. Татарским языком в мире владеет около 8 миллионов человек, в то же время для него не существует разработанных средств анализа речи. Отсутствуют также необходимые для развития данных средств речевые и текстовые корпуса, что позволит в полной мере использовать разработанный для этих целей инструментарий.

Визуальное представление структуры проекта изображено на рис. 2.

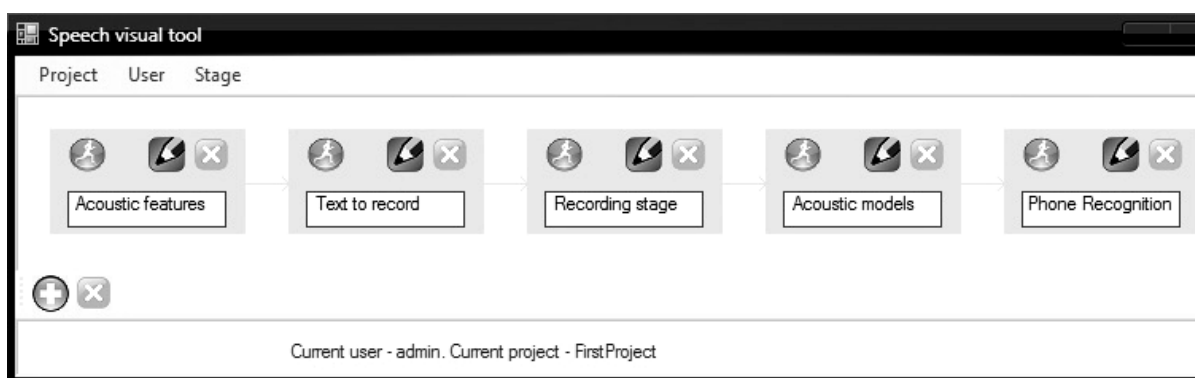


Рис. 2. Структура проекта распознавания фонем татарского языка

Описание характеристик и результатов работы всех модулей созданного проекта представлено в табл. 2. Для расчета параметров речевого сигнала, акустического моделирования, а также при реализации алгоритма распознавания фонем было задействовано средство НТК toolkit [7].

Анализ результатов работы проекта, а именно файлов с распознанными последовательностями фонем для 1157 файлов из тестового корпуса татарской речи, показал качество распознавания фонем, равное 61,25%.

Описание модулей проекта по распознаванию фонем татарского языка

Название модуля	Параметры	Описание результатов работы
Acoustic features	Входные: рабочая директория. Выходные: пути к файлам алфавита языка, алфавита фонем, правил транскрибирования и исполняемому файлу транскрибирования	С помощью данного инструмента был сформирован алфавит из 57 фонем татарского языка. Кроме того, были определены и описаны правила составления фонетических транскрипций [8], общее число правил равняется 35
Preparing text corpus to record	Входные: все выходные параметры модуля «Acoustic features». Выходные: пути к файлам с подготовленными для записи предложениями и их транскрипциями	Инструмент позволяет создавать наборы предложений для записи с постоянным контролем частотности входящих в их состав фонем и количества фонемных 1-, 2- и 3-gram
Recording stage	Входные: все выходные параметры модуля «Preparing text corpus to record». Выходные: пути к файлам списка дикторов, пронумерованных предложений с транскрипциями, распределения предложений по дикторам и путь к записанным аудиофайлам корпуса	Данный модуль предоставляет 3 основные возможности: 1. Создавать и редактировать список дикторов с указанием подробной информации. 2. Настраивать параметры распределения предложений по дикторам. 3. Производить запись аудиофайлов согласно заданным параметрам. Было записано 10788 файлов общей продолжительностью 5 ч
Acoustic models	Входные: все выходные параметры модуля «Recording stage». Выходные: пути к файлам конфигурации вычисляемых параметров, а также описаниям построенных акустических моделей (файлы hmmdefs, macros)	С помощью данного модуля была произведена настройка необходимых параметров (максимальное число гауссовских распределений в смеси – 29) и проведено обучение акустических моделей фонем
Phoneme recognition	Входные: все выходные параметры модуля «Acoustic modules», а также путь к аудиофайлам для распознавания. Выходные: файл с результатами распознавания	Было передано в качестве входных параметров и распознано 1157 файлов из тестовой части записанного корпуса

Заключение. Описание особенностей процесса построения и функционирования систем анализа речи позволило в конечном итоге разработать и построить программный комплекс проектирования подобных систем. Комплекс учитывает необходимость одновременной работы над проектом множества специалистов из разных областей, предоставляя возможности по настройке прав доступа для каждого из участников проекта. Основная сущность комплекса – модуль – обладает способностью обмениваться данными с другими модулями проекта, что предоставляет возможность, во-первых, создать единые точки доступа к базовой функциональности (например, вычислению вектора признаков) и избегать благодаря этому дублирования схожей функциональности в других проектах, и во-вторых, повысить качество работы отдельных подсистем анализа речи за счет использования информации, полученной от других подсистем (например, использование в работе распознавателя речи информации о языке диктора, полученной от подсистемы идентификации диктора). Кроме того, для упрощения процесса создания систем анализа речи для множества языков, для которых такие системы ещё не созданы, в комплексе был реализован набор инструментария и базовых алгоритмов, что заметно ускоряет создание базовых вариантов систем анализа речи.

Для демонстрации функциональности построенного комплекса на его основе была реализована система автоматического распознавания фонем татарского языка. С помощью инструментов комплекса были проведены такие этапы работы, как формирование алфавита фонем татарского языка, выделение основных акустических особенностей татарской речи, создание системы автоматического транскрибирования, подготовка и запись речевого корпуса. Полученные результаты (качество распознавания фонем 61,25%) позволяют сделать вывод о том, что комплекс может быть применён при построении подобных систем, а возможности добавления новых и расширения

функциональности уже существующих модулей позволяют дорабатывать системы для решения конкретных задач.

Литература

1. Vimala C. A Review on Speech Recognition Challenges and Approaches / C. Vimala, V. Radha // World of Computer Science and Information Technology Journal (WCSIT). – 2012. – Vol. 2, № 1. – P. 1–7.
2. Alumäe T. Open and Extendable Speech Recognition Application Architecture for Mobile Environments / T. Alumäe, K. Kaljurand // Conference Proceedings of The 3rd Workshop on Spoken Languages Technologies for Under-resourced languages (SLTU–2012). 7–9 May 2012, Cape Town. – Cape Town: North-West University, South Africa, 2012. – P. 15–18.
3. Хусаинов А.Ф. Прототип платформы анализа речи на татарском языке / А.Ф. Хусаинов, Д.Ш. Сулейманов // Матер. III Междунар. научн.-техн. конф. «Открытые семантические технологии проектирования интеллектуальных систем OSTIS – 2013». Минск, 21–23 февраля 2013 г. – Минск: БГУИР, 2013. – С. 361–368.
4. Сизов А.Г. Оценка качества голоса лиц, страдающих дисфонией / А.Г. Сизов, Р.В. Мещеряков, С.Д. Тиунов // Труды СПИИРАН. – 2012. – № 1. – С. 138–152.
5. Мещеряков Р.В. Структура систем синтеза и распознавания речи // Изв. Том. политех. ун-та. – 2009. – Т. 315, № 5. – С. 121.
6. Мещеряков Р.В. Диалог как основа построения речевых систем / Р.В. Мещеряков, В.П. Бондаренко // Кибернетика и системный анализ. – 2008. – № 2. – С. 30.
7. The НТК book (for НТК Version 3.4) [Электронный ресурс]. – Режим доступа: <http://speech.ee.ntu.edu.tw/courses/DSP2011spring/hw2/НТКBook-3.4.1.pdf>, свободный (дата обращения: 01.02.2013).
8. Практическая фонетика татарского языка [Электронный ресурс]. – Режим доступа: <http://tatar.com.ru/fonetika.php>, свободный (дата обращения: 04.06.2013).

Хусаинов Айдар Фаилович

Аспирант Института вычислительной математики и информационных технологий
Казанского (Приволжского) федерального университета
Тел.: +7-987-205-13-80
Эл. почта: khusainov.aidar@gmail.com

Khusainov A.F.

Software complex for speech analysis (the case of Tatar phonemes recognition system)

This paper describes software toolkit for speech analysis designing, creating and managing purposes. This toolkit also implements basic functionality (i.e. building transcriber tool, designing and recording speech corpora) and provides ability to edit and add new components. Proposed toolkit employed for creating automatic phoneme recognition system for the Tatar language, experimental results indicate that toolkit can be successfully used for creation of speech analysis systems for under-resourced languages.

Keywords: speech analysis, design tools, phoneme recognition, the Tatar language.