

УДК 004.522:004.934

Д.Г. Савкова, О.И. Федяев

Донецкий национальный технический университет
кафедра прикладной математики и информатики

АКУСТИКО-ЛИНГВИСТИЧЕСКАЯ МОДЕЛЬ РАСПОЗНАВАНИЯ РЕЧИ

Аннотация

Савкова Д.Г., Федяев О.И. Акустико-лингвистическая модель распознавания речи. Разработана и исследована акустико-лингвистическая модель распознавания речи, основанная на скрытых марковских моделях. С помощью этой модели проведены эксперименты по автоматическому распознаванию лексем языка программирования в дикторонезависимом и дикторозависимом режимах.

Ключевые слова: автоматическое распознавание речи, акустико-лингвистическая модель, скрытые марковские модели

Проблема интеллектуализации взаимодействия «человек-компьютер». Устная речь является наиболее естественным и простым для человека способом общения. Сейчас речевые технологии широко используются в робототехнике и управлении различными компьютерными системами, потому что коммуникация с техническими устройствами на уровне речевых актов для человека является более удобной [1]. Ключевым элементом любого речевого интерфейса является система автоматического распознавания речи, основу которой составляет акустико-лингвистическую модель. Из-за нестабильности человеческой речи проблема качества её распознавания по-прежнему остаётся актуальной.

Цель статьи. Разработать в современной инструментальной среде CMU Sphinx [2] акустико-лингвистическую модель и провести анализ её эффективности на примере автоматического распознавания слов языка программирования.

Задача исследования. Необходимо оценить качество распознавания изолированных слов языка программирования Паскаль на дикторонезависимой акустической модели, построенной с помощью инструментария Sphinx на словарях разного объёма.

Распознавание речи на основе скрытых марковских моделей. Марковская модель – это вероятностный автомат с конечным числом состояний, изменяющий своё состояние один раз в единицу времени. При этом

наблюдателю известны состояния и вероятности переходов между состояниями (матрица переходов). Таким образом, марковская модель описывает некоторый вероятностный процесс. Каждому наблюдаемому событию этого процесса соответствует одно из состояний модели.

Рассмотрим процесс автоматического распознавания речи на основе скрытых марковских моделей, используемых в инструментальной системе Sphinx (рис.1). Пользователь произносит слова в микрофон, а звуковая карта преобразовывает звук в цифровой сигнал. Сигнал не обрабатывается системой в таком виде, она преобразует его в последовательность векторов характеристик. В них выделяются отрезки, на которых присутствуют слова. Каждое слово разбивается на фонемы и им в соответствие сопоставляются наиболее вероятные состояния скрытой марковской модели. Так находятся текстовые представления каждому речевому образу.

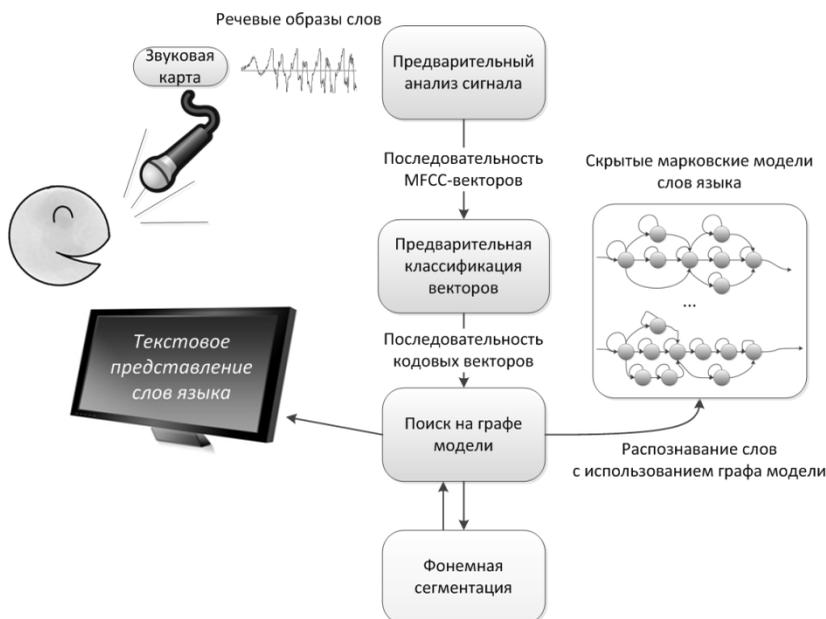


Рисунок 1 – Функциональная схема системы распознавания речи

Акустико-лингвистическая модель системы распознавания слов. В качестве входных данных для построения акустической модели распознавания слов служат: словарь, языковая модель, список фонем и прочих звуков. В словаре содержится список слов и транскрипции к ним. Транскрипции должны состоять исключительно из фонем, которые присутствуют в списке фонем. Помимо слов есть и другие звуки, не несущие в себе смысловой нагрузки: звуки дыхания, различный шум. Языковая модель – совокупность

вероятностей появления слов в речи. В соответствии со всеми вышеперечисленными характеристиками речи записывается аудиобазы и она оформляется в виде обучающих примеров с ответами. Каждая аудиозапись должна иметь своё текстовое представление. Чем больше материала в аудиобазе, тем лучше качество распознавания. Элементы информационной структуры лингвистической модели показаны на рис. 2.

Чтобы собрать модель на Sphinxtrain последней версии необходимо запустить python-скрипт. Для нахождения неизвестных параметров скрытой марковской модели Sphinx использует алгоритм Баума-Велша.

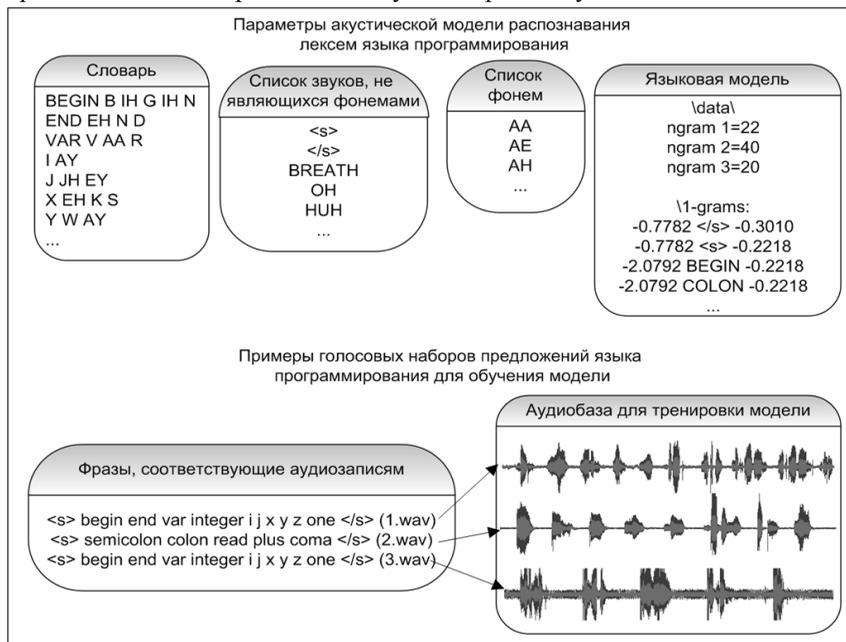


Рисунок 2 – Информационная структура лингвистической модели

Исследование качества модели на примере речевых слов языка. Для исследования качества модели был проведен ряд экспериментов. Было создано 5 словарей разного объема: на 20, 40, 60, 80 и 100 слов. Они содержали английские слова – лексемы языка программирования Паскаль. Были использованы две акустические модели: дикторонезависимая Voxforge [3] и собственная дикторозависимая. В дикторозависимой участвовал один диктор, каждое слово из 100 было произнесено 5 раз. В экспериментах распознавались изолированные слова без грамматики. Аудиобазы для тестирования были одинаковой для всех словарей. Она содержала все слова словаря на 20 слов (по 4 повтора в разном порядке). График зависимости качества распознавания от

объема словаря показан на рис. 3. Модель, обученная на определенного диктора, имеет лучшее качество распознавания, чем дикторозависимая.

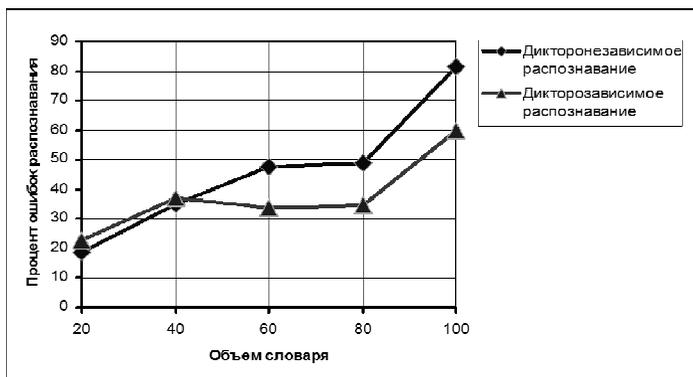


Рисунок 3 – График зависимости качества распознавания от объема словаря

Это объясняется не только тем, что система лучше распознает диктора, на которого обучалась, но и тем, что база Voxforge обучалась на американских дикторах, а записи для тестирования записывались русскоговорящим диктором.

Выводы. Анализ работы модели показывает, что качество распознавания изолированных слов, не связанных между собой грамматикой, не удовлетворяет практическим требованиям. Целесообразно тренировать акустическую модель на конкретного диктора, т.к. в этом случае качество распознавания улучшается на 20% (для качественной настройки достаточно 5 прочтений всех слов словаря).

Список литературы

1. Савкова Д.Г., Бондаренко И.Ю. Опыт применения инструментальной системы Sphinx для решения задачи распознавания речевых команд управления компьютерными системами // Сборник материалов 3-й Всеукраинской научно-практической конференции «Информационные управляющие системы и компьютерный мониторинг» ИУС КМ-2012. – Донецк: ДонНТУ. – 2012. – с. 111-117.

2. CMU Sphinx Open Source Toolkit For Speech Recognition Evaluation [Electronic resource] / Интернет-ресурс. - Режим доступа: <http://cmusphinx.sourceforge.net/> [проверено 1.04.2013]. - Загл. с экрана.

3. Welcome – Russian Evaluation [Electronic resource] / Интернет-ресурс. - Режим доступа: <http://www.voxforge.org/ru> [проверено 1.04.2013]. - Загл. с экрана.