

Министерство образования и науки Российской Федерации

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САРАТОВСКИЙ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра теоретических основ
компьютерной безопасности и
криптографии

**Построение системы распознавания речи на основе Скрытых Марковских
Моделей**

АВТОРЕФЕРАТ

дипломной работы

студентки 6 курса 631 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Мавлютовой Юлии Альбертовны

Научный руководитель

Доцент, к.п.н

А.С. Гераськин

31.12.2016 г.

Заведующий кафедрой

профессор, к.ф.-м.н.

В.Н. Салий

31.12.2016 г.

Саратов 2017

ВВЕДЕНИЕ

Распознавание речи стремительно набирает популярность на мировом рынке. Одной из причин тому является его активное использование в голосовом управлении мобильными устройствами. Покупатели все более склонны выбирать мобильные устройства, оснащенные приложениями распознавания речи, например, для отправки сообщений, управления музыкальным плей-листом, совершения звонка абоненту из списка без физического контакта с телефоном и т.д.

Однако кроме спроса на мобильные приложения с возможностью распознавания голоса, растет и спрос на услуги голосовой биометрии. Связано это с уникальностью голоса каждого человека, что позволяет идентифицировать человека с высокой степенью точности. В том числе и с помощью технологий распознавания голоса.

Эта особенность уже используется при идентификации человека для судебно-медицинских целей. Запись голоса подозреваемого в совершении преступления можно сопоставлять с образцами из голосовой базы данных, или наоборот, можно подсчитать вероятность того, что голос данного подозреваемого совпадает с голосом на записи, относящейся к преступлению. Очевидно, что такое доказательство может хорошо помочь при раскрытии уголовных дел.

Распознавание голоса также используется и в мобильном банкинге. Например, в Северной Америке и Западной Европе возможно осуществление определенных транзакций при аудио-обращении пользователя в банк. Поэтому голосовая аутентификация пользователя является необходимой составляющей, обеспечивающей невозможность нарушителю выдать себя за пользователя мобильного банкинга.

Распознавание речи, а также голосовая идентификация нашли свое применение в военных ведомствах. Благодаря такому подходу обеспечивается

безопасность от несанкционированного проникновения в защищенную зону. Система содержит базу данных голосов военнослужащих, имеющих доступ к защищаемой зоне, и предотвращает допуск людей, чьих голосов в ней нет. К тому же, голосовые команды могут использоваться для управления самолетом, запуском ракет и т.д.

Целью работы является разработка программного модуля, осуществляющего формирование алфавита СММ с лево-правой топологией для отдельного слова, использующего кодирования признаков звукового сигнала на основе линейного предсказания (КЛП).

Задачами, поставленные в данной работе, являются:

- 1) Изучить принципы работы СММ для распознавания речи;
- 2) разработать алгоритм предварительной обработки звукового файла;
- 3) разработать алгоритм выделения признаков полученного слова;
- 4) реализовать и проверить работоспособность разработанных алгоритмов.

Дипломная работа состоит из введения, 3 разделов, заключения, списка использованных источников и 1 приложения. Общий объем работы – 65 страниц, из них 39 – основное содержание, включая 21 рисунок, список использованных источников из 10 наименований.

КРАТКОЕ СОДЕРЖАНИЕ

В разделе 1 «Общие сведения» дается общая информация о системах распознавания речи, а также предлагается своя структура такой системы.

Задача распознавания речи – это выделение, классификация и должная реакция на человеческую речь из входного потока. В подразделе 1.1 «Система распознавания речи» описывается общая структура систем распознавания речи и ее классификация. Также рассматриваются причины ошибок системы при идентификации диктора. К ним относятся:

- 1) **среда записи** (уровень и тип шума);
- 2) **среда представления** (длительность речи, психофизиологическое состояние говорящего (болезнь, эмоциональное состояние и т.д.), язык речевого сообщения, изменение голосового усилия);
- 3) **среда канала** (помехи (импульсные, тональные и т.д.), искажения (амплитудно-частотные характеристики микрофона и канала передачи, вид кодирования в канале и т.д.)).

Поэтому так важно произвести качественную предварительную обработку сигнала – от нее зависит, сможет ли система правильно распознать слово.

В подразделе 1.2 «Структура системы распознавания речи» предлагается один из возможных вариантов структуры системы распознавания речи:

- 1) Речевая база содержит записи слов, которые будут доступны для распознавания;
- 2) модуль анализа признаков формирует алфавит слова, по которому формируется последовательность наблюдений;
- 3) по полученной последовательности наблюдений обучается скрытая марковская модель (СММ), наиболее подходящая из которых сохраняется в базе;

4) из входного звукового потока или файла (в данной работе для простоты будет рассматриваться второй вариант) с помощью предварительной обработки выделяются слова (подавляется шум, отбрасываются лишние звуки и т.д.);

5) аналогично (2) анализируются признаки полученного слова, формируется последовательность наблюдений;

6) полученная последовательность поочередно подается моделям СММ, хранящимся в базе. Вычисляется вероятность того, что данная последовательность подходит для очередной СММ. Слово, которому принадлежит СММ с наибольшей вероятностью, считается искомым словом.

В разделе 2 «Математические основы распознавания речи» представлены все сведения, необходимые для выполнения поставленных задач.

В подразделе 2.1 «Алгоритм выделения отдельных слов из единого wav-файла» приведена блок-схема алгоритма предварительной обработки звукового файла, содержащего несколько раз произнесенных одним диктором требуемого слова, основанный на фильтрации и выделении участков файла соответствующих повторяющимся словам. При этом для моделирования отдельного слова была выбрана лево-правая СММ, предполагающая что в каждый момент времени система переходит в новое состояние.

В подразделе 2.2 «Алгоритм анализа признаков слов» представлена блок-схема кодирования, основанного на линейном предсказании (КЛП анализа), а также приведена блок-схема алгоритма получения алфавита слова, а также последовательности наблюдений, основанного на КЛП анализе.

В подразделе 2.3 «Понятие Марковских моделей» описывается, что такое скрытая марковская модель (СММ), а также приводятся параметры, необходимые для задания любой СММ, и условия, которым они должны отвечать.

Алгоритм позволяет уточнять параметры модели таким образом, чтобы вероятность того, что данная наблюдаемая последовательность построена

именно для данной модели, $p(O|\lambda)$ увеличивалась. Итеративное применение алгоритма до схождения в одной точке позволяет максимизировать $p(O|\lambda)$. Сам алгоритм приведен в подразделе 2.4 «Алгоритм метода Баума-Велша обучения СММ системы».

В разделе 3 «Реализация алгоритма выделения отдельных слов из единого wav-файла» представлена программная реализация алгоритма, блок-схема которого приводилась в подразделе 2.1. Из единого wav-файла, содержащего десять повторений одного и того же слова, сказанного одним диктором, выделяются команды, отсеиваются лишние шумы и паузы. Команды дискриминируются по времени.

В разделе 4 «Реализация алгоритма построения последовательности наблюдений по выделенному отдельному слову» приводится программная реализация алгоритма, представленного в подразделе 2.2. Также приводятся некоторые статистические данные, доказывающие, что разработанному и реализованному алгоритму можно верить.

Работа алгоритма проверяется на двух типах слов: периодическом (то есть когда слово состоит из двух и более повторяющихся частей) и непериодическом. На вход алгоритму подавались тридцать звуковых файлов, каждый из которых содержит десять повторений одной и той же команды. Все команды произносились одним и тем же диктором.

В качестве непериодического слова была выбрана команда «вперед». Для получения хороших результатов необходимо правильно подобрать значение Eps минимального расстояния между состояниями слова, при котором состояния считаются одинаковыми. Например, при значении $Eps=0,37$ из 30 экспериментов всего 2 результата не отвечали требованиям лево-правой топологии модели (то есть число состояний совпадало с числом значений). При значении $Eps=0,4$ не отвечали требованиям лево-правой топологии уже 7 результатов.

Также была исследована правдоподобность получаемых количеств состояний. Для этого по выборке из 30 экспериментов при показателе надежности 95% среднее значение было равно 19, доверительный интервал которого лежит в отрезке $[19-0,775; 19+0,775]$. Иными словами, вероятность ошибки составляет всего 4%. Следовательно, полученным данным можно верить.

Полученные последовательности наблюдений подавались на вход модулю обучения СММ. При этом для последовательности наблюдений каждого эксперимента была обучена СММ, то есть вероятность того, что сгенерированная СММ подходит для данной последовательности наблюдений, значительно увеличивалась.

В качестве периодической команды была выбрана команда «на-на». В команде есть повторяющиеся части, следовательно, она заведомо не может удовлетворять предположению о лево-правой модели. Это подтверждают и результаты экспериментов, где всего лишь в четырех из них число состояний совпадало с числом значений. При этом даже на подходящих последовательностях наблюдений (удовлетворяющих лево-правой топологии) невозможно было добиться хороших результатов обучения, то есть вероятность того, что сгенерированная СММ подходит для данной последовательности наблюдений, крайне мала.

Таким образом, для периодических слов было принято решение о необходимости обработки только одного периода слова, то есть число состояний было уменьшено, в данном случае, вдвое. Тогда при правильно подобранном $Eps=0.34$ всего 1 полученная последовательность наблюдений не удовлетворяла предположению о лево-правой топологии. Хотя при $Eps=0.35$ полученные результаты не удовлетворяют предположению о лево-правой топологии уже в 6 случаях.

Для этого случая также была исследована правдоподобность получаемых количеств состояний. По выборке из 30 экспериментов при показателе

надежности 95% среднее значение было равно 18, доверительный интервал которого лежит в отрезке $[18-0,199; 18+0,199]$. Иными словами, вероятность ошибки составляет всего 1,1%. Следовательно, полученным данным можно верить.

Обучение СММ для случая периодичной команды с числом состояний вдвое меньше числа значений также дает хорошие результаты.

ЗАКЛЮЧЕНИЕ

В данной работе были исследованы системы распознавания речи. В результате анализа теоретического материала была определена задача, которая заключается в разбиении звукового файла на отдельные слова. В результате исследования был разработан алгоритм, решающий поставленную задачу и реализованный в виде программного модуля.

Также был разработан алгоритм, на основе которого был реализован и отлажен программный модуль, осуществляющий для отдельного слова формирование алфавита СММ с лево-правой топологией, использующего кодирования признаков звукового сигнала на основе линейного предсказания (КЛП).

Для проверки работоспособности алгоритмов было проведено исследование, показывающее, что полученным данным можно верить.

В работе был изучен алгоритм метода Баума-Велша обучения СММ по заданной последовательности наблюдений. Полученные результаты были проверены на модуле, реализующем данный алгоритм.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Перспективы развития систем распознавания речи (выдержка из исследования) [Электронный ресурс] // Хабрахабр [Электронный ресурс]. URL: <https://habrahabr.ru/post/232613/> (дата обращения: 27.09.2016). Загл. с экрана. Яз. рус.

2 Распознавание речи. Часть 1. Классификация систем распознавания речи [Электронный ресурс] // Хабрахабр [Электронный ресурс]. URL: <https://geektimes.ru/post/64572/> (дата обращения: 27.09.2016). Загл. с экрана. Яз. рус.

3 Матвеев, Ю.Н. Технологии биометрической идентификации личности по голосу и другим модальностям/ Ю.Н. Матвеев // Вестник МГТУ им. Н.Э. Баумана. Сер. «Приборостроение». 2012. С. 46-61.

4 Рабинер, Л. Р. Скрытые Марковские модели и их применение в избранных приложениях при распознавании речи: Обзор/ Л.Р. Рабинер // ТИИЭР, 1989, т. 77, № 2, С. 86–120.

5 Суворов, В.Н. О кепстральном анализе в популярной форме/ В.Н. Суворов // Ви Тэк, Санкт-Петербург, 2006, № 4, С. 52-53.

6 Адлер, Ю.П. Планирование эксперимента при поиске оптимальных условий/ Ю.П. Адлер, Е.В. Маркова, Ю.В. Грановский. М.: Наука, 1976 – 279 с.

7 Савин, А.Н. Разработка компонентов программного комплекса для потоковой фильтрации аудиоконтента на основе использования скрытых марковских моделей / А.Н. Савин, Н.Е. Тимофеева, А.С. Гераськин, Ю.А. Мавлютова // Известия Саратовского университета. Новая серия. Серия Математика. Механика. Информатика, 2015, С. 340–350.

8 Центральная предельная теорема [Электронный ресурс] // Википедия [Электронный ресурс]. URL: https://ru.wikipedia.org/wiki/%D0%A6%D0%B5%D0%BD%D1%82%D1%80%D0%B0%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D0%BF%D1%80%D0%

В5%D0%B4%D0%B5%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D1%82
%D0%B5%D0%BE%D1%80%D0%B5%D0%BC%D0%B0 (дата обращения:
25.12.2016). Загл. с экрана. Яз. рус.

9 Федосов, В.П. Цифровая обработка сигналов в LabVIEW: учеб. пособие / В.П. Федосов, А.К. Нестеренко. М.: ДМК Пресс, 2007 – 456с.

10 Лупов, С.Ю., LabVIEW в примерах и задачах. Учебно-методические материалы по программе повышения квалификации «Обучение технологиям National Instruments». / С.Ю. Лупов, С.И. Муякшин, В.В. Шарков. Нижний Новгород, 2007 – 101 с.