Уральский федеральный университет

X МЕЖДУНАРОДНАЯ
НАУЧНО-МЕТОДИЧЕСКАЯ КОНФЕРЕНЦИЯ
«НОВЫЕ ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ В ВУЗЕ»

Курлов А.С., Пелевин В.Н.

Kurlov A.S., Pelevin V.N.

ИДЕНТИФИКАЦИЯ ЧЕЛОВЕКА ПО ГОЛОСУ ДЛЯ СИСТЕМЫ ДИСТАНЦИОННОГО ОБУЧЕНИЯ

METHOD OF HUMAN IDENTIFICATION BY VOICE AND SPEECH RECOGNITION FOR THE DISTANCE EDUCATION

kurlov@nexusdreamwork.ru
ФГАОУ ВПО «УрФУ имени первого Президента России Б.Н.Ельцина»
г. Екатеринбург

НОТВ-2013

Представлены различные технологии идентификации человека по голосу и распознания речи. Данные разработки могут быть применены в сфере дистанционного обучения.

This article is about identification of human voice and speech recognition. These developments can be applied in the field of distance education.

Область применения

Сегодня существую различные проблемы в сфере дистанционного обучения. Одной из такой проблем является идентификация пользователя системы ДО. Сейчас данный вопрос решается при помощи аутентификации и веб-камеры.

Существует возможность преобразовать в текстовый формат и идентифицировать голос студента во время работы с системой ДО. Таким образом, можно организовать, например, сдачу какого-либо теста в устной форме: человек не щелкает мышкой по верному, на его взгляд, ответу, а называет его (вариант ответа) вслух. Затем можно проанализировать голос на предмет принадлежности «нужному» студенту. Дело в том, что потенциально любой, кто прошёл процесс аутентификации, способен нажать на мышку, в результате нет гарантии, что именной ожидаемый студент проходил процедуру тестирования.

Данная методика увеличивает степень защиты от несанкционированного доступа посторонних лиц к системе ДО. Стоит отметить, что механизм способен работать в течение всего сеанса взаимодействия с системой ДО.

Идентификация голоса и распознание речи может быть достаточно легко имплементирована в систему ДО. Благодаря этому конечному пользователю не понадобится устанавливать различные сторонние программные продукты.

Технический аспект

Существует возможность воспользоваться готовыми сервисами, которые способны хорошо определять произнесённые слова.

Рассмотрим, как можно осуществить преобразование речи в текстовый формат при помощи Google Speech. Взаимодействовать с данным сервисом необходимо через Google API (application programming interface – набор готовых методов, функций, процедур или переменных). Алгоритм работы может выглядеть следующим образом:

1. Записываем речь в формат, с которым способен работать Google API.
2. Отправляем полученный файл.
3. Проанализировать полученный ответ, в котором содержится записанная речь в текстовом формате.

На самом деле данный алгоритм можно упростить за счёт отправки речи непосредственно в Google Speech без предварительной записи файла на компьютер. Запись в файл уместна на ранних стадиях разработки подобной методики для более наглядного анализа работы сервиса распознания речи.

Рассмотрим инструменты, на основе которых можно построить собственный сервис по переводу голосового сигнала в текстовый вид. Примером такого инструмента может послужить Microsoft Speech SDK, iSpeech API, Dragon SDK, Modular Audio Recognition Framework[1].

Самые распространённые алгоритмы, которые используются для трансформации речи в текстовый формат являются: Скрытая Модель Маркова, динамическое программирование, нейронные сети.

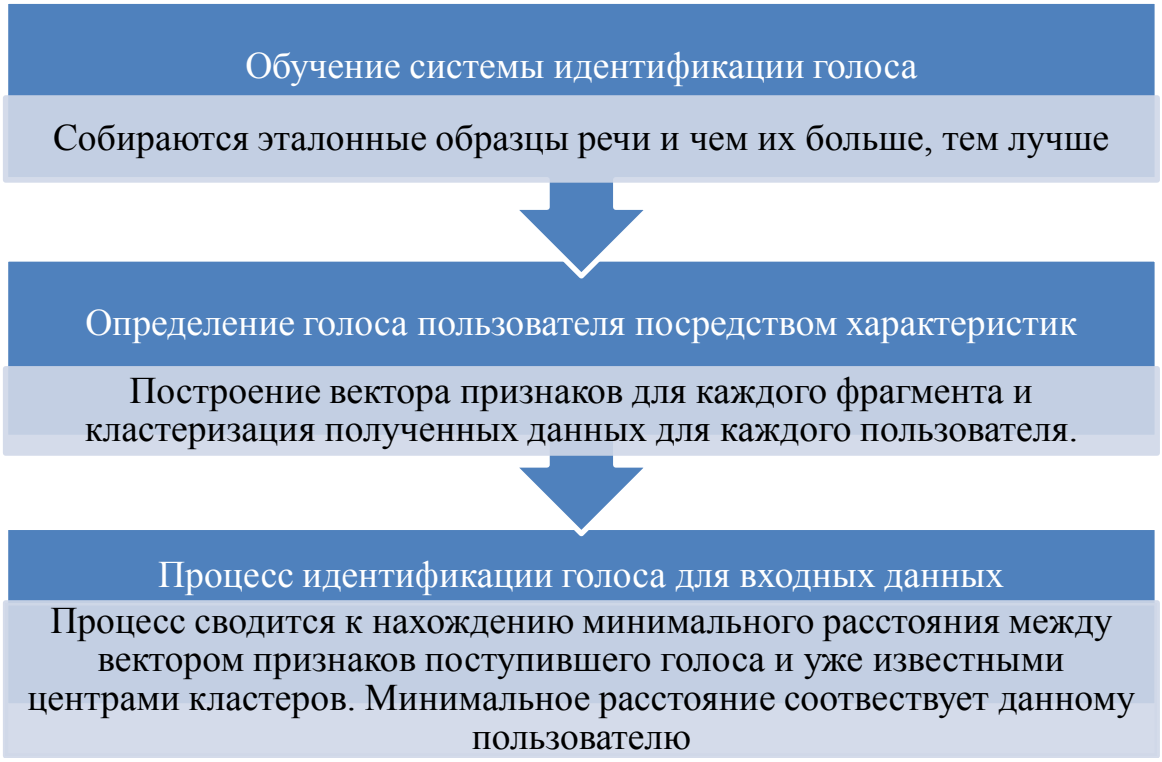
Скрытая Модель Маркова – статистическая модель, которая способна на выходе предоставлять последовательности из символов и чисел. Скрытая Модель Маркова используется для преобразования речи в текстовый формат по причине того, что речь может быть представлена в виде статического сигнала на коротком промежутке времени. В результате весь фрагмент голосового сигнала может быть представлен как стационарный процесс.

Метод распознавания речи при помощи динамического программирования заключается в измерении различий между отдельными речевыми фрагментами. Метод динамического программирования уступает Скрытой Модели Маркова.

Нейронные сети используются в различных аспектах распознавания речи. В отличии от Скрытой Модели Маркова, нейронные сети не работают со статистическими свойствами голосового сигнала. Метод использования нейронных сетей сводится к обучению системы. Существует возможность использовать метод нейронных сетей для первичной обработки голосового сигнала, а затем анализировать информацию при помощи Скрытой Модели Маркова.

Рассмотрим один из способов идентификации пользователя по голосу, использующий Мел-частотные кепстральные коэффициенты[2]. Мел – единица измерения высоты звука. Кепстр – служит определением последовательности разложений коэффициентов функции в степенной ряд. При помощи Мел-частотных кепстральных коэффициентов можно не только идентифицировать голос, но и построить систему трансформации голосового сигнала в поток текстовых символов.

Процесс идентификации голоса пользователя можно разделить на следующие этапы:



Вектор признаков для каждого фрагмента голосового сигнала состоит из Мел-кепстральных коэффициентов. Над полученными данными удобно провести кластеризацию при помощи метода К-средних.

Стоит отметить, что можно установить некоторое допустимое значение для вычисляемого расстояния. Таким образом, могут возникать ситуации, когда голос человека не относящийся к системе ДО, будет распознан как неизвестный и будет произведён отказ в доступе.

В завершение хочу добавить, что для достижения хорошей точности распознавания голоса, рекомендовано использовать от четырёх фрагментов записи голоса, которые можно однажды проанализировать и использовать полученные данные неограниченное количество раз.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Modular Audio Recognition Framework [Электронный ресурс]. – Режим доступа: http://marf.sourceforge.net/docs/marf/0.3.0.6/report.pdf (дата обращения 14.01.2013).
2. Speaker identification using mel frequency [Электронный ресурс]. – Режим доступа: http://ethesis.nitrkl.ac.in/3745/1/final_yr_project__thesis.pdf (дата обращения 14.01.2013).