

М.С. Шляпугин

Научный руководитель: канд. техн. наук, А. В. Астафьев

Муромский институт (филиал) федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых»

602264, Владимирская область, г. Муром, ул. Орловская, д.23

E-mail:

Обзор и анализ технологии распознавания речи для систем аутентификации пользователей

Интернет становится всё более мобильным. Прирост пользователей за год составляет около 38%. Классические интерфейсы ввода, такие как клавиатуры, мыши и джойстики отходят на второй план, уступая место более естественным методам использования: touch, жесты или голос. По статистике исследовательской компании SEA Market Research, использование голосовых интерфейсов по отношению к классическому вводу текста на мобильных устройствах превысило 50%, что по статистике означает то, что в среднем каждый пользователь мобильных устройств хотя бы раз в месяц пользуется голосовыми интерфейсами. Востребованность данной технологии увеличивается каждый год в геометрической прогрессии, поэтому обходить её стороной больше нельзя. Исходя из вышесказанного, можно сделать вывод, что анализ технологии распознавания речи для систем аутентификации пользователей является актуальной научно-технической задачей.

Целью данной научно-технической работы является проведение обзора и анализа технологии распознавания речи для систем аутентификации пользователей, общий обзор сути данной технологии, анализ её преимущества над остальными системами аутентификации и запись вывода всего исследования, а также проведение краткого обзора существующих на рынке технологий распознавания речи.

Обычный человек вполне способен распознавать речь другого человека – это означает, что на биологическом уровне есть определенные алгоритмы, технически следуя которым, появляется возможность восстановить сказанное. Исследование таких алгоритмов – цель акустического моделирования. Человеческая речь состоит из фонем, их можно представить базовыми смысловыми единицами языка. Чтобы наиболее правильно смоделировать фонему, используется три её состояния — начало, середина и конец. Эти части фонем, называемые сенонами, составляют фонетический алфавит. Небольшой участок сигнала называют «фреймом». Нельзя абсолютно точно определить какая фонема была произнесена на каком-либо фрейме – многие фонемы очень схожи друг с другом. В данном вопросе рассуждают в ключе теории вероятности – для одного сигнала одни фонемы наиболее вероятны, другие менее, третьи можно убрать из самого рассмотрения. Соответственно, определение акустической модели — это функция, принимающая на вход небольшой участок акустического сигнала, называемого фреймом и выводящая вероятности различных фонем на этом фрейме. Рассматривая всё это технически получается, что акустическая модель принимает на вход звук, а на выходе дает распределение вероятностей по сенонам из которой получается амплитуда колебаний по времени. Это не самая информативная форма представления акустического сигнала. Намного лучше – спектр сигнала. На практике обычно используется логарифмированный и отмасштабированный спектр, что соответствует закономерностям человеческого слухового восприятия. Полученные величины подвергаются дискретному косинусному преобразованию, и в результате получается вектор из 13 вещественных чисел. Они могут использоваться, как вход акустической модели. Исходя из полученного системой на вход кодового слова, идёт его представление в виде полученного спектра сигнала, после чего происходит его сравнение с базой данных, в случае сходства происходит положительная аутентификация пользователя.

Аутентификация по голосу является самой распространенной и используемой. Она занимает около 59% от всей используемой биометрической аутентификации. Анализ основных

преимуществ такой системы вынесен в таблицу и позволяет наглядно оценить преимущества речевых технологий перед остальными:

Параметры	Аутентификация			
	Отпечаток пальца	Радужная оболочка глаза	3-D распознавание лица	Голос
Уровень равной ошибки (EER)	2 – 3,3%	4,1 – 4,6%	4,1%	0,1 – 0,86%
Стоимость системы	Высокая	Очень высокая	Высокая	Низкая
Отказ в доступе пользователю	4%	7%	~0%	2-3%
Вероятность взлома	2,5%	6%	4%	0,75%
Вероятность пропуска ложной цели	0,1%	0,001%	10%	0,75%

В настоящее время существует две основных интерфейса распознавания речи:

1) Yandex.SpeechKit.

В области речевых технологий Яндекс занимается с 2012 года. На Yet another Conference 2013 компания представила разработчикам новую библиотеку Yandex.SpeechKit. Это публичный API для распознавания речи, который могут использовать разработчики. Яндекс довольно долго развивал эту систему и сейчас правильно распознается 94% слов в Навигаторе и Мобильных Картах, а также 84% слов в Мобильном Браузере. Это уже очень достойное качество, и компания активно работает над дальнейшими улучшениями характеристик их библиотеки..

2) Google Voice Search

Это продукт компании Google в технологии распознавания речи. Данная технология довольно давно уже интегрирована в мобильные телефоны и компьютеры, где можно ввести информация с помощью голоса. С 14 июня 2011 года Google объявила об интеграции речевого движка в Google Search и с тех пор он работает в стабильном режиме с этого времени. Эта технология на персональных компьютерах поддерживается только браузером Google Chrome. Функция включена по умолчанию в сборках, но может быть включена вручную добавлением командного флага. Есть также функция голосового управления для введения речевых команд на телефонах с ОС Android.

Вывод: технология распознавания речи не является априори самой лучшей системой для аутентификации пользователей. Силовые, правоохранительные и государственные структуры могут использовать в своем распоряжении намного более высокоточные системы аутентификации. Однако для среднего пользователя или компании, рассматриваемая технология является наиболее приемлемой в соотношении цены и качества и предоставляет собой наилучший выбор для использования.

Список использованной литературы:

1. Маркел Дж., Грей А. Х. Линейное предсказание речи / Пер. с англ. М.: Связь, 1980
2. Л.Р. Рабинер, Р.В. Шафер, Цифровая обработка речевых сигналов, М.: Радио и связь, 1981. - 495с.
3. Молдовян АА., Молдовян Д.Н., Левина А.Б. Протоколы аутентификации с нулевым разглашением секрета, Университет ИТМО, 2016 – 49с.