

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.418:656.073.9

Савицкий
Ян Валерьевич

Повышение надежности и функциональных характеристик идентификационной
речевой системы

АВТОРЕФЕРАТ
на соискание академической степени
магистра технических наук

1-59 81 01 – Управление безопасностью производственных процессов

Научный руководитель
В. В. Радыгина, кандидат
биологических наук, доцент

Минск 2019

ВВЕДЕНИЕ

Системы идентификации по голосу – это вычислительные системы, которые могут определять голосовой отпечаток говорящего из общего потока. Распознавание голоса используется в биометрических целях безопасности, чтобы определить голос конкретного человека. Эта технология стала очень популярной в мобильном банкинге, который требует идентификации подлинности пользователей, а также для других голосовых команд, чтобы помочь им совершать сделки. Также с развитием умных колонок, ботов-помощников и начала интенсивного развития в области голосовых интерфейсов, технология идентификация людей по голосу стала особенно популярна.

При идентификации по средством голоса большие трудности представляют собой процессы сравнения эталонного голосового отпечатка с предоставляемым. От того, насколько эффективно будет решена эта задача, самым непосредственным образом зависит процент правильной идентификации по голосу.

Данная работа написана на тему «Повышение надежности и функциональных характеристик идентификационной речевой системы».

Актуальность научной работы обусловлена развитием голосовым интерфейсов, необходимостью идентификации людей посредством голоса, возможностью осуществлять работы системы без использования сервера, при этом не требуется специальных устройств, так как система оптимально организует взаимодействие человека и системы, что позволяет использовать менее сложную технику с минимальным количеством программного обеспечения.

Целью работы является изучить особенности функционирования голосовых интерфейсов и разработать систему речевой идентификации с повышенной надежностью и функциональными характеристиками.

Задачи исследования:

1. Провести анализ научно-технической литературы по системам идентификации по голосу.
2. Разработать систему идентификации пользователя посредством голоса с повышенной надежностью и функциональными характеристиками.
3. Провести испытание и обоснование эргономичности системы.

Результатом магистерской диссертации стала разработка идентификационной речевой системы с повышенной надежностью и функциональными, предназначенной для идентификации или авторизации пользователя посредством голоса.

КРАТКОЕ СОДЕРЖАНИЕ

Распознавание по голосу – одна из форм биометрической аутентификации, позволяющая идентифицировать личность человека по совокупности уникальных характеристик голоса. Относится к динамическим методам биометрии. Однако, поскольку голос человека может меняться в зависимости от возраста, эмоционального состояния, здоровья, гормонального фона и целого ряда других факторов, не является абсолютно точным. По мере развития звукозаписывающей и воспроизводящей техники, технология распознавания применяется с различным успехом в сфере защиты информации, охраны и систем доступа, криминалистике.

Весь процесс обработки речевого сигнала можно разбить на несколько главных этапов:

- предобработка сигнала;
- выделение критериев;
- распознавание диктора.

Для получения и предобработки сигнала необходимы микрофон, фильтр и аналого-цифровой преобразователь для дальнейшей работы с цифровой записью голоса. С выхода микрофона сигнал подается на вход блока фильтрации. Следующим этапом является прохождение аналого-цифровой преобразователь. Далее оцифрованный сигнал попадает в блок цифровой обработки. В блоке цифровой обработки сигнал фильтруется и преобразуется в вектор, с которым в дальнейшем будет работать микропроцессор и нейросетевой обработчик. Для последующего сравнения с сохраненным ранее голосовым отпечатком диктора полученный вектор заносится в энергонезависимую память. После сравнения отпечатка в памяти с полученным отпечатком микроконтроллер подает команду на блок управления внешним устройством, к примеру, на магнитный дверной замок. Сам процесс голосовой идентификации не требователен к ресурсам и состоит из двух этапов. На первом этапе осуществляется получение голосового отпечатка диктора и преобразование к виду, в котором его можно будет сравнить с другими. Второй этап заключается в сравнении голосовых отпечатков при помощи обученной нейронной сети.

Нейронная сеть – это последовательность нейронов, соединенных между собой синапсами. Структура нейронной сети пришла в мир программирования прямоком из биологии. Благодаря такой структуре, машина обретает способность анализировать и даже запоминать различную информацию. Нейронные сети также способны не только анализировать входящую информацию, но и воспроизводить ее из своей памяти. Другими словами, нейронная сеть машинная интерпретация мозга человека, в котором находятся

миллионы нейронов, передающих информацию в виде электрических импульсов.

Нейронные сети используются для решения сложных задач, которые требуют аналитических вычислений подобных тем, что делает человеческий мозг.

Одним из наиболее сложных аспектов в разработке систем идентификации по голосу является широкая междисциплинарность задачи. При работе над такими системами затрагиваются вопросы теории обработки сигналов, математического анализа, лингвистики, теории коммуникаций, физиологии и в некоторых случаях психологии. И для построения успешной системы распознавания необходимо рассмотреть такой круг дисциплин, который один человек охватить не в состоянии. Следовательно, для разработчика становится особенно важно понимание основ речевого распознавания без необходимости быть экспертом в каждой из сторон проблемы.

Идентификация по голосу – это не простая задача, особенно усложненная тема, что существует множество источников вариативности, связанных с речевым сигналом.

Основными проблемами систем идентификации по голосу:

- необходимо более помехоустойчивое распознавание информативных признаков, связанных с характеристиками голосового источника;
- наличие полноценной речевой базы для тестирования и обучения систем идентификации по голосу с большим числом дикторов, различного возраста, акцента, особенностями произношения, эмоционального состояния, записанных на различных условиях записи и с разной частотой дискретизации.
- уменьшение требований к аппаратному обеспечению для работы системы.

Рассмотренные проблемы распознавания речи позволяют сформулировать следующие задачи:

- 1 Реализация алгоритма получения голосового отпечатка пользователя;
- 2 Разработка нейронной сети с архитектурой для решения задачи сравнения векторов данных и нахождения общих признаков.
- 3 Обучение нейронной сети на большой обучающей выборке данных, полученной из хранилища данных Яндекса;
- 4 Проверка работы нейронной сети на тестовой выборке.
- 5 Разработка программного модуля с открытым интерфейсом и возможностью получать и использовать для анализа голосовой

отпечаток из других источников, что позволит сделать систему распределенной;

6 Разработка демонстрационного android-приложения.

Для реализации процесса сравнения голосовых отпечатков было решено использовать нейросетевое сравнение при помощи самоорганизующейся сети Кохонена, так как данная нейронная сеть обучается без учителя, с применением модификации, позволяющей ускорить ее работу с минимальной потерей точности. Создание в TensorFlow сети с выбранной архитектурой не составляет проблем, т.к. данная архитектура входит в базовый пакет данного фреймворка. Далее проходим по каждому слою и добавляем один нейрон для реализации алгоритма автоэнкодера и вынужденной корреляции соседних слоев.

Обучающая выборка, для используемой нейронной сети, состоит лишь из входных векторов, а обучающий алгоритм подстраивает веса сети так, чтобы получались согласованные выходные векторы, т. е. чтобы предъявление достаточно близких входных векторов давало одинаковые выходы. Процесс обучения, следовательно, выделяет статистические свойства обучающего множества и группирует сходные векторы в классы. Предъявление на вход вектора из данного класса даст определенный выходной вектор.

В результате работы был разработан модуль анализа звуковой дорожки и получения голосового отпечатка, разработана и обучена нейронная сеть на большой выборке данных. Большая выборка данных позволяет уменьшить вероятность ошибки идентификации связанных с качеством звука, наличием посторонних шумов, акцентом, эмоциональной окраской в голосе и других признаков. Архитектура системы позволяет сделать ее распределенной, что позволит уменьшить время обработки запроса и уменьшить требования к аппаратному обеспечению.

ЗАКЛЮЧЕНИЕ

Для решения поставленных задач была изучена научно-техническая литература по системам идентификации по голосу и выявлены современные проблемы в системах идентификации. В итоге разработана библиотека на основе положительных качеств уже существующих систем и алгоритмов распознавания, отличающаяся повышенным функционалом и высокой вероятностью распознавания, на качество которого эмоциональное состояние диктора и качество записи в меньшей степени влияет на вероятность распознавания. В качестве языка программирования был выбран Java. В качестве алгоритма обработки звуковой записи и получения голосового отпечатка использовался кепстральный анализ с применением наиболее быстрых методов обработки голосового сигнала. В качестве платформы для разработки нейронной сети использовался TensorFlow. Нейронная сеть была представлена в виде самоорганизующейся карты Кохонена. В результате множества попыток обучения были выбраны наилучшие значения порогов для скорости обучения, количества эпох и функции изменения скорость обучения.

Повышенная надежность и функциональные характеристики системы были достигнуты путем обучения нейронной сети на обучающей выборке, с большим многообразием голосовых характеристик, а также добавления возможности приема и обработки уже разложенного результирующего вектора характеристик, что позволяет поддерживать распределенную систему идентификации по голосу устройствам с малой производительностью, а также использовать иные методы расчета вектора характеристик голоса.

Тема работы актуальна, особенно на нынешнем этапе развития голосовых интерфейсов.

Результатом магистерской диссертации стала разработка системы идентификации по голосу с повышенной надежностью и функциональными характеристиками, предназначенной для интегрирования в системы, где необходима идентификация или авторизация пользователя по средством голоса.