

СОЗДАНИЕ ANDROID-ПРИЛОЖЕНИЯ С ФУНКЦИЕЙ РАСПОЗНАВАНИЯ ДИКТОРА ПО ГОЛОСУ

КОРНИЛОВ Н. Я.

*Научный руководитель: Матолыгин А. А., старший преподаватель каф. ЭМИС, ТУСУР
Проект ГПО ЭМИС-1202 «Создание приложений для платформы Android»*

Объектом исследования является система распознавания человека по голосу.

Цель проекта - создание Android-приложения с функцией распознавания диктора по голосу и функцией распознавания речи.

Актуальность

Существует ряд ситуаций, в которых пользователю мобильного устройства необходимо подтвердить свое право на доступ к информации. Подтверждение такого права осуществляется с помощью кодов авторизации: пароля, графического ключа, отпечатка пальца и иных способов. Иногда такие средства верификации личности либо неудобны, либо не обеспечивают необходимой степени защиты. Поэтому, в дополнение к таким традиционным средствам, целесообразно использовать биометрические параметры человека. Преимущество биометрии заключается в том, что эти параметры всегда находятся при человеке, их нельзя забыть, потерять, передать другому человеку, украсть и задача их воспроизведения является чрезвычайно трудоемкой.

Задача распознавания диктора по голосу сводится к тому, чтобы выделить и классифицировать человеческую речь из входного потока. При этом выделяют две подзадачи: идентификация и верификация [1].

Под идентификацией понимают процесс определения человека по образцу голоса. При этом результатом является идентификатор диктора, зарегистрированного в системе, модель голоса которого наиболее вероятно соответствует входному образцу. Верификация является процессом, при котором с помощью сравнения представленного образца с хранимой в базе моделью проверяется запрошенная идентичность. Результатом является подтверждение человека или отрицательный ответ системы.

Новизна работы

В связи с бурным развитием информационных технологий, в частности, рынка мобильных телефонов и смартфонов целесообразно реализовать распознавание диктора по голосу для конечных пользователей мобильных устройств. В настоящее время основную долю рынка занимает платформа Android [2].

Так как приложение разрабатывается для Android-устройств, аппаратом вычисления является мобильный телефон (смартфон) с установленной операционной системой Android.

Новизна проекта заключается в том, что на рынке мобильных приложений подобных проектов не существует. В магазинах приложений для мобильных устройств, а именно: Google Play, Apple App Store, Amazon Appstore и др. не найдены приложения, распознающие человека по голосу.

Метод реализации задачи

В работе [3] по распознаванию диктора доминирует метод кепстрального преобразования спектра речевых сигналов.

Кепстр [4] — энергетический спектр функции $\ln|S(\omega)|^2$, определяемый выражением:

$$C_s(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln|S(\omega)|^2 e^{i\omega q} d\omega,$$

где $|S(\omega)|^2$ — спектральная плотность энергии сигнала.

Другими словами, кепстр определяет последовательность коэффициентов разложения функции $\lg[\Phi(z)]$ в степенной ряд.

Схема этого метода такова (см. приложение А, Б): на интервале времени в 10 – 20 мс вычисляется текущий спектр мощности, затем применяется обратное преобразование Фурье от логарифма этого спектра (кепстр), и находятся коэффициенты кепстра [3]:

$$c_n = \sum_{m=1}^M [\log S(m)] \cos \left[\frac{\pi n}{M} \left(m - \frac{1}{2} \right) \right],$$

где c_n — мел-кепстральный коэффициент под номером n ;

S_m — амплитуда m -го значения в кадре в мелах;

M — наперед заданное количество мел-кепстральных коэффициентов;

$n \in [1, M]$.

Свойства слуха учитываются путем нелинейного преобразования шкалы частот, обычно в шкале *мел*. Шкала мел вычисляется как

$$M(f) = 1125 \cdot \ln \left(1 + \frac{f}{700} \right),$$

где f — частота в герцах;

M — частота в мелах.

Коэффициенты кепстрального преобразования формируют пространство, в котором и производится распознавание диктора. Обычно сохраняется только первые несколько элементов (от 8 до 16), по которым в дальнейшем производится идентификация сигнала [5]. Часто используются первые и вторые разности по времени кепстральных коэффициентов, что втрое увеличивает размерность пространства принятия решений, но улучшает эффективность распознавания диктора [6].

Кепстр описывает форму огибающей спектра сигнала, в которой интегрируются характеристики источников возбуждения (голосового, турбулентного и импульсного) и формы речевого тракта. В экспериментах по субъективному распознаванию голоса было установлено, что огибающая спектра сильно влияет на узнаваемость голоса [7]. Поэтому использование того или иного способа анализа огибающей спектра в целях распознавания диктора оправдано.

Применение

Приложение может найти применение в разблокировке экрана мобильного телефона и голосовом управлении. Более того, если приложение будет обладать достаточной эффективностью, его можно использовать в криминалистике, судебной экспертизе, радио-разведке, контр-разведке, антитеррористическом мониторинге.

Результаты работы

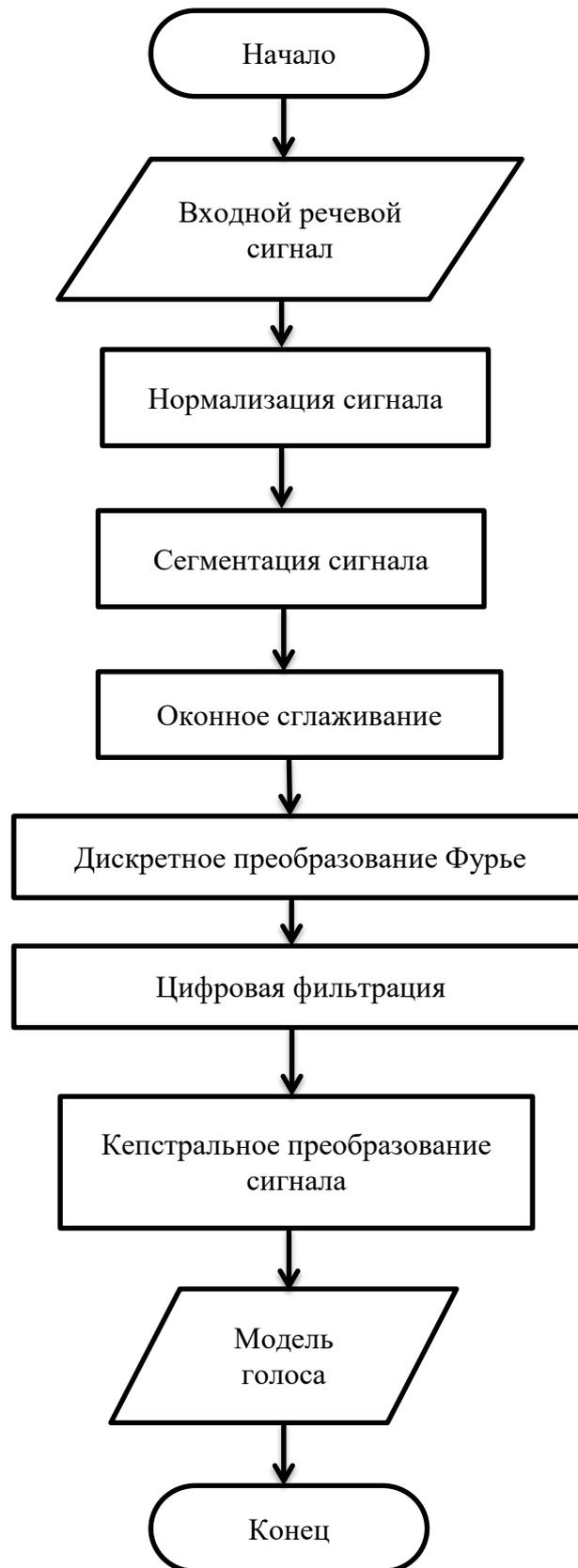
По результатам выполнения проекта сделаны следующие выводы:

- разработан пользовательский интерфейс Android-приложения;
- реализована запись речевого сигнала с микрофона;
- реализована сегментация сигнала и ее обработка в окне во избежание нежелательных эффектов;
- получена спектрограмма речевого сигнала;
- реализовано распознавание речи.

Список использованных источников

1. Теория передачи сигналов: Учебник для вузов Зюко А.Г., Кловский Д.Д., Назаров М.В., Финк Л.М. – Москва: Связь, 1980.-288 с.
2. Strategy Analytics: Android Shipped 1 Billion Smartphones Worldwide in 2014 [Электронный ресурс]: сайт фирмы Strategy Analytics – Режим доступа: <http://www.prnewswire.com/news-releases/strategy-analytics-android-shipped-1-billion-smartphones-worldwide-in-2014-300027707.html> (дата обращения: 13.05.16).
3. В. Н. Сорокин, В. В. Вьюгин, А. А. Тананыкин «Распознавание личности по голосу: аналитический обзор» Информационные процессы, Том 12, №1, стр. 1-30, 2012 г.
4. Кепстр [Электронный ресурс]: электронная энциклопедия – Режим доступа: <https://ru.wikipedia.org/wiki/Кепстр> (Дата обращения: 17.05.16).
5. О.С. Агашин, О.Н. Корелин «Методы цифровой обработки речевого сигнала в задаче распознавания изолированных слов с применением сигнальных процессоров» 2012 г.
6. Zhang Sh.-X, Mak M.-W. (2009). A new adaptation approach to high-level speaker-model creation in speaker verification. *Speech Communication*, v. 51, 534-550.
7. Itoh K. (1992). Perceptual analysis of speaker identity. In: Saito, S. (Ed.), *Speech Science and Technology*. IOS Press, 133-145.

ПРИЛОЖЕНИЕ А
(обязательное)
Алгоритм формирования модели голоса



ПРИЛОЖЕНИЕ Б
(обязательное)

Алгоритм распознавания личности по голосу

