

**РАСПОЗНАВАНИЕ ГОЛОСА В СИСТЕМАХ БЕЗОПАСНОСТИ**Н.Я. Корнилов

Научный руководитель: старший преподаватель каф. ЭМИС А.А. Матолыгин

Томский университет систем управления и радиоэлектроники,

Россия, г. Томск, пр. Ленина, 40, 634050

E-mail: [centauri368@gmail.com](mailto:centauri368@gmail.com)**VOICE RECOGNITION IN SAFETY SYSTEMS**N.Ya. Kornilov

Scientific Supervisor: Senior teacher of the dep. EMIS A.A. Matolygin

Tomsk University of Control Systems and Radioelectronics,

Russia, Tomsk, Lenin str., 40, 634050

E-mail: [centauri368@gmail.com](mailto:centauri368@gmail.com)

***Abstract.** This article discusses the problem of biometric way of identifying a person, in particular, voice recognition for safety system. The algorithm for obtaining the speaker's voice model is described. In this paper, the method of Mel-frequency cepstral coefficients (MFCC) is considered for distinguishing the distinctive features of the speaker. Besides, speaker recognition by voice may be used in criminal investigations, forensics and radio reconnaissance.*

**Введение.** 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.

Системы распознавания говорящего по голосу строятся на принципе отличия голосов разных людей друг от друга. Распознавание говорящего – это процесс определения на основе характеристик речевого сигнала, принадлежит ли данное высказывание конкретному диктору. Системы распознавания человека по голосу подразделяются на два вида: идентификацию и верификацию говорящего [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.

**Материалы и метод исследования.** В работе [2] по распознаванию диктора доминирует метод кепстрального преобразования спектра речевых сигналов.

Кепстр [3] – 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 (рис. 1): 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 [2]:

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

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. Мел – психофизическая единица высоты звука, основанная на восприятии звука слуховым аппаратом человека. Шкала мел вычисляется как

$$M(f) = 1127 \cdot \ln\left(1 + \frac{f}{700}\right),$$

где  $f$  – частота в герцах;

$M$  – частота в мелах.

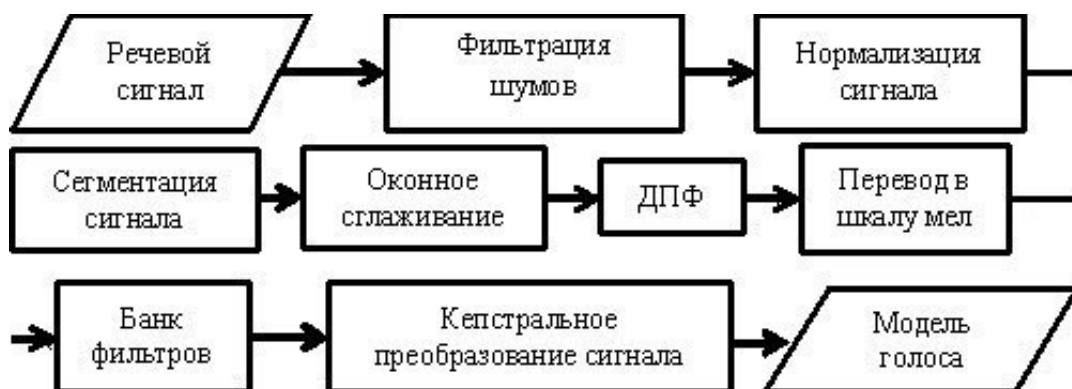


Рис. 1. Алгоритм формирования модели голоса

Коэффициенты кепстрального преобразования формируют пространство, в котором и производится распознавание диктора.

Кепстр описывает форму огибающей спектра сигнала, в которой интегрируются характеристики источников возбуждения и формы речевого тракта. В экспериментах по субъективному распознаванию голоса было установлено, что огибающая спектра сильно влияет на узнаваемость голоса [4]. Поэтому

использование того или иного способа анализа огибающей спектра в целях распознавания диктора оправдано.

Наиболее популярны методы аппроксимации плотности вероятности в пространстве признаков взвешенной смесью нормальных распределений (GMM – Gauss Mixture Models), метод опорных векторов (SVM – Support Vector Machines), метод скрытых Марковских моделей (HMM – Hidden Markov Models), искусственные нейронные сети, а также модификации факторного анализа [2].

Хотя распознавание по голосу не может гарантировать стопроцентную правильность результата, оно может довольно эффективно использоваться в таких областях, как криминалистика и судебная экспертиза; разведка; антитеррористический мониторинг; безопасность; банковское дело [2].

**Результаты.** В данной работе система распознавания диктора по голосу разрабатывается для мобильных устройств с операционной системой Android. Результаты выполнения работы:

- разработано приложение на платформе Android;
- разработан пользовательский интерфейс Android-приложения;
- реализована запись речевого сигнала с микрофона;
- реализована сегментация сигнала и ее обработка в окне;
- получена спектрограмма речевого сигнала;
- преобразован спектр в шкалу частот мел.

**Заключение.** Распознавание людей по голосу в наше время является одним из важных приложений речевых технологий. Это обусловлено развитием информатизации общества, постоянным наращиванием потоков передачи данных по различным каналам связи, которые требуют защиты от несанкционированного доступа, ростом значения автоматической обработки данных, а также нарастающим использованием автоматических речевых технологий. Системы распознавания говорящего по индивидуальным характеристикам вызывают большой интерес – как в научных, так и коммерческих кругах. Задача распознавания человека по голосу была поставлена более 40 лет назад, но все еще остается актуальной для исследований.

#### СПИСОК ЛИТЕРАТУРЫ

1. Зюко А.Г., Кловский Д.Д., Назаров М.В., Финк Л.М. Теория передачи сигналов: Учебник для вузов. – М.: Связь, 1980. – 288 с.
2. Сорокин В.Н. Распознавание личности по голосу: аналитический обзор / В.Н. Сорокин, В.В. Вьюгин, А.А. Тананыкин // Информационные процессы. – 2012. – Т. 12. – № 1. – С. 1–30.
3. Кепстр: электронная энциклопедия [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki/Кепстр>. – 17.05.2016.
4. Itoh K. Perceptual analysis of speaker identity / K. Itoh // Speech Science and Technology. – 1992. – P. 133–145.