

УСТРОЙСТВО ГОЛОСОВОЙ ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ В СИСТЕМАХ КОНТРОЛЯ ДОСТУПА

П. А. Меньшаков, И. А. Мурашко

Гомельский государственный технический университет имени П.О. Сухого

Гомель, Республика Беларусь

E-mail: pmenshakov@gmail.com, iamurashko@tut.by

На данный момент основным средством контроля доступа являются пропускные пункты, оборудованные различными средствами контроля доступа. Но большинство из средств контроля доступа имеют высокую цену. Причем большая часть расходов приходится на выделение персонального средства идентификации каждому пользователю. Решением данной проблемы может стать голосовая идентификация. Использование биометрии позволяет отказаться от чипов и карт доступа, исключить потерю средства идентификации и его кражу. А использование голоса позволит отказаться от дорогостоящего оборудования для считывания данных

ВВЕДЕНИЕ

В настоящее время, голосовая идентификация, как и биометрия в целом, уже получила широкое распространение. Самый простой пример – сканеры отпечатков пальцев, установленные почти на каждом ноутбуке. Биометрия предполагает систему распознавания людей по одной или более физических или поведенческих черт. В области информационных технологий биометрические данные используются в качестве формы управления идентификаторами доступа и контроля доступа. Также биометрический анализ используется для выявления людей, которые находятся под наблюдением [1]. Довольно длительное время отпечатки пальцев используются для идентификации преступников и предотвращения воровства или мошенничества. Некоторые люди умеют имитировать голоса, но, это требует особых навыков, которые не часто повстречаешь в обыденной жизни [2].

УСТРОЙСТВО ГОЛОСОВОЙ ИДЕНТИФИКАЦИИ

Первоначальным этапом голосовой идентификации является получение голоса пользователя. Для этого необходим микрофон, фильтр и аналого-цифровой преобразователь, для дальнейшей работы с цифровой записью голоса.

В общем виде процесс ввода речевых сообщений приведен на рисунке 1.

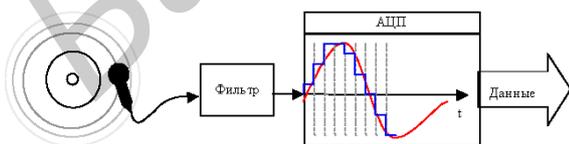


Рис. 1 – Схема ввода записи голоса

С выхода микрофона сигнал подается на вход блока фильтрации. Следующим этапом является прохождение АЦП.

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.

Так же, полученный вектор заносится в энергонезависимую память. Это необходимо для последующего сравнения с полученным отпечатком. После сравнения отпечатка в памяти с полученным отпечатком, микроконтроллер подает команду на блок управления внешним устройством, к примеру, на магнитный дверной замок. Общая схема устройства представлена на рисунке 2.

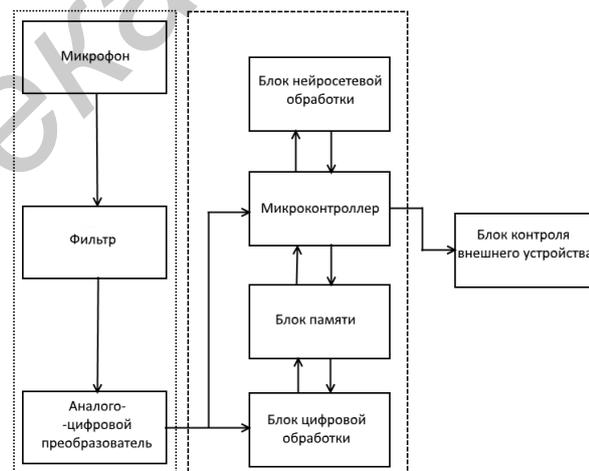


Рис. 2 – Схема устройства

Сам процесс голосовой идентификации не требователен к ресурсам, и состоит из двух этапов. Первый - получить голосовой отпечаток. Вторым шагом является сравнение голосовых отпечатков при помощи обученной нейронной сети.

Полученную запись голоса необходимо разделить на кадры. Данное действие необходимо для более простой работы с записанной звуковой дорожкой.

Следующим этапом является устранение нежелательных эффектов и шумов. Это необходимо для того, чтобы записи, полученные в разное, время соответствовали друг другу независимо от сторонних факторов. Мною использова-

лось умножение каждого кадра на особую весовую функцию "Окно Хемминга":

$$\omega(n) = 0.53836 - 0.46164 * \cos\left(\frac{2\pi n}{N-1}\right) \quad (1)$$

где n — порядковый номер элемента в кадре, для которого вычисляется новое значение амплитуды,

N — длина кадра (количество значений сигнала, измеренных за период).

Полученные кадры преобразуются в их частотную характеристику при помощи прогонки через "Быстрое Преобразование Фурье":

$$X_k = \sum_{i=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad (2)$$

где N — длина кадра (количество значений сигнала, измеренных за период),

x_n — амплитуда n -го сигнала,

$X_k \sim N$ — комплексных амплитуд синусоидальных сигналов, слагающих исходный сигнал.

На сегодняшний день наиболее успешными являются системы распознавания голоса, использующие знания об устройстве слухового аппарата. В виду данных особенностей необходимо привести частотную характеристику каждого кадра к «мелам».

Для перехода к «мел» характеристике используется следующая зависимость:

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right) = 1127 \log_e\left(1 + \frac{f}{700}\right) \quad (3)$$

где m — частота в мелах,

f — частота в герцах.

Это последнее действие, необходимое для последующего преобразование в вектор характеристики, который, впоследствии, сравнивается с базой голосовых записей. Вектор будет состоять из мел-кепстральных коэффициентов, получить которые можно по следующей формуле:

$$c_n = \sum_{k=1}^K (\log S_k) \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right] \quad (4)$$

где c_n — мел-кепстральный коэффициент под номером n ,

S_k — амплитуда k -го значения в кадре в мелах,

K — наперед заданное количество мел-кепстральных коэффициентов $n \in [1, K]$.

Полученный вектор характеристик добавляется в базу данных, для последующего сравнения с ним.

В работе использовалась нейронная сеть с обучением без учителя, так как оно является намного более правдоподобной моделью обучения в биологической системе. Процесс обучения, выделяет статистические свойства обучающего множества и группирует сходные векторы в классы.

Предъявление на вход вектора из данного класса даст определенный выходной вектор [3]. Схематически данная сеть изображена на рисунке 3.

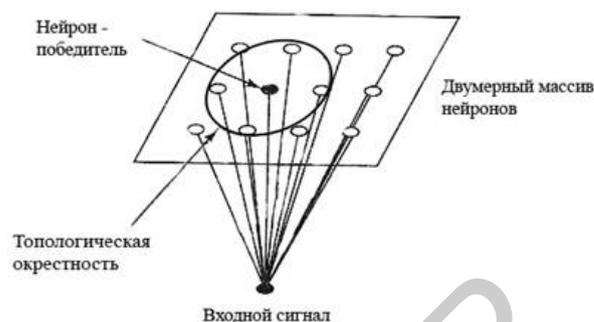


Рис. 3 – Схема сети Кохонена

Распространение сигнала в такой сети происходит следующим образом: входной вектор нормируется на 1.0 и подается на вход, который распределяет его дальше через матрицу весов W . Каждый нейрон в слое Кохонена вычисляет сумму на своем входе и в зависимости от состояния окружающих нейронов этого слоя становится активным или неактивным (1.0 и 0.0). Нейроны этого слоя функционируют по принципу конкуренции, т. е. в результате определенного количества итераций активным остается один нейрон или небольшая группа. Так как обработка этого механизма требует значительных вычислительных ресурсов, в моей модели он заменен нахождением нейрона с максимальной активностью и присвоением ему активности 1.0, а всем остальным нейронам 0.0.

Так как входной вектор x нормирован, т. е. расположен на гиперсфере единичного радиуса в пространстве весов, то при коррекции весов по этому правилу происходит поворот вектора весов в сторону входного сигнала. Постепенное уменьшение скорости поворота, позволяет произвести статистическое усреднение входных векторов, на которые реагирует данный нейрон.

Как показало исследование, полученный алгоритм позволяет значительно ускорить работу программы голосовой идентификации. Данная модернизация позволяет использовать программу на предприятиях с большим потоком пользователей.

Так же программный комплекс очень гибок и имеет большое пространство для дальнейшего совершенствования и добавления новых функций, что делает его не только выгодным программным продуктом, но и перспективным проектом для развития и получения прибыли.

1. Introduction to digital audio coding and standards / M. Bosi, R.E. Goldberg – Springer Science+Business, Media USA. – 2003. – 434 p.
2. AudioCoding: Theory and Applications / Y. You – NY: Springer, 2010 – 349 p.
3. Компьютерная обработка звука./ А. П. Загуменнов – М. : ДМК, 1999. – 384 с