



OSTIS-2016

(Open Semantic Technologies for Intelligent Systems)

УДК 004.822:514

МЕТОДИКА ГОЛОСОВОЙ ИДЕНТИФИКАЦИИ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

Меньшаков П.А., Мурашко И.А.

*Гомельский государственный технический университет имени П.О. Сухого
Гомель, Республика Беларусь*

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.

Ключевые слова: голосовая идентификация; биометрия; контроль доступа.

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]. Так же биометрия предусматривает и поведенческий анализ объекта. К ним относятся ходьба, жесты и т.п.

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].

1. Аппаратная реализация

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.

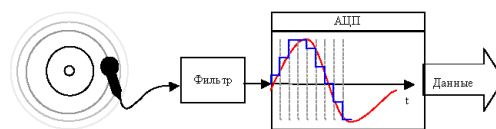


Рисунок 1 – Схема ввода записи голоса

С выхода микрофона сигнал подается на вход блока фильтрации. Следующим этапом является прохождение АЦП. Когда АЦП используется для амплитудного анализа, число, получаемое на выходе АЦП, используется для адресации памяти и называется номером канала, а V – шириной канала. Номер канала несет информацию об амплитудном значении сигнала.

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.

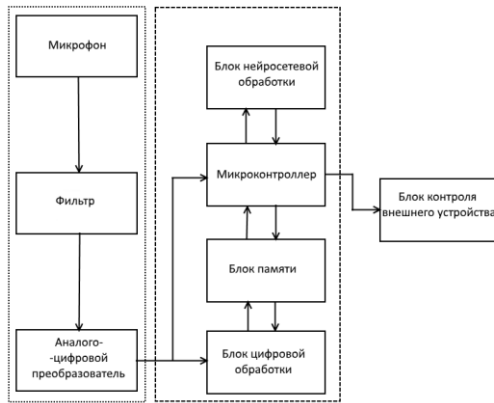


Рисунок 2 – Схема устройства

2. Принцип голосовой идентификации

Сам процесс голосовой идентификации не требователен к ресурсам, и состоит из двух этапов. Сперва, необходимо получить голосовой отпечаток пользователя и преобразовать к виду, в котором его можно будет сравнить с другими. Вторым шагом является сравнение голосовых отпечатков при помощи обученной нейронной сети. Для реализации процесса преобразования необходимо произвести определенный порядок действий.

При помощи микрофона получается запись голоса идентифицируемого и отправляется на ЭВМ. Наиболее оптимальным является получение WAV файла, в виду простоты работы с ним.

Полученную запись голоса необходимо разделить на кадры. Разделение на кадры представлено на рисунке 3. Данное действие необходимо для более простой работы с записанной звуковой дорожкой.

Далее все вычисления будут производиться с каждым кадром в отдельности.

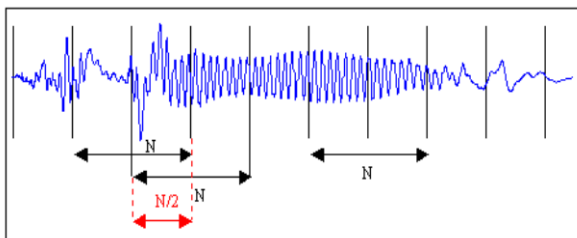


Рисунок 3 – График звуковой волны

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. Существует множество способов, при помощи которых можно уменьшить шумовые эффекты. Мною использовалось 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 преобразуется в их частотную характеристику при помощи прогонки через "Быстрое Преобразование Фурье":

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

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

x_n – амплитуда n -го сигнала,

X_k – 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». Зависимость представлена на рисунке 4.

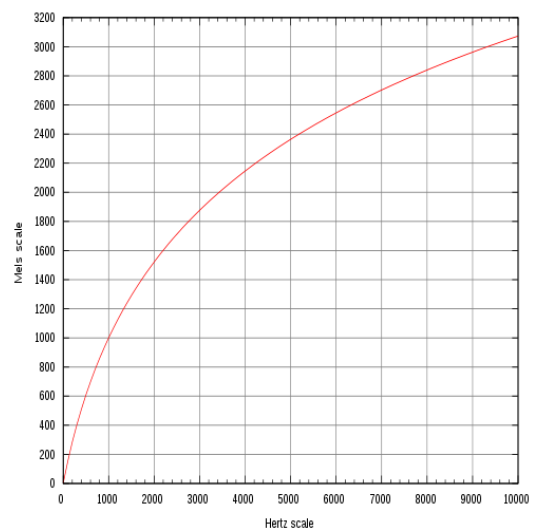


Рисунок 4 – График зависимости высоты звука (в мелах) от его частоты

Для перехода к «мел» характеристике используется следующая зависимость:

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

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.

Однако более оптимальным вариантом является использование нескольких записей одного и того же голоса. Заранее определенное количество образцов голоса можно использовать для обучения нейронной сети.

В работе использовалось обучение без учителя, так как оно является намного более правдоподобной моделью обучения в биологической системе. Развита Кохоненом и многими другими, она не нуждается в целевом векторе для выходов и, следовательно, не требует сравнения с предопределенными идеальными ответами, а обучающее множество состоит лишь из входных векторов. 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 [3]. Схематически данная сеть изображена на рисунке 5.

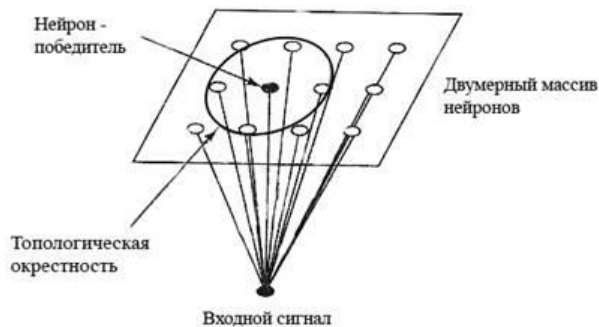


Рисунок 5 – Схема сети Кохонена

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. Этот механизм называется латеральным. 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. Таким образом, срабатывает нейрон, для которого вектор входа ближе всего к вектору весов связей.

Если сеть находится в режиме обучения, то для выигравшего нейрона происходит коррекция весов матрицы связи по формуле:

$$w_n = w_n + \alpha (x - w_n). \quad (5)$$

где w_n – новое значение веса;

w_m – старое значение;

α – скорость обучения;

x – величина входа.

Геометрически это правило иллюстрирует рисунок 6.

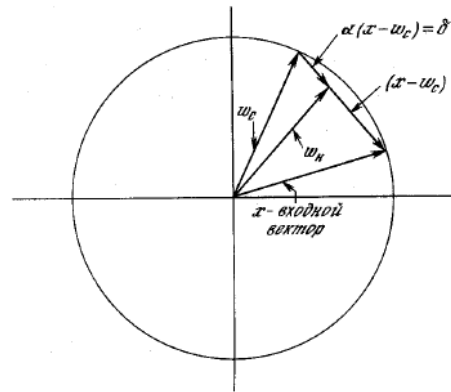


Рисунок 6 – Коррекция весов нейрона Кохонена

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.

Заключение

Итогом данного исследования стало модульное приложение, осуществляющее голосовую идентификацию пользователя, использующее модернизированный алгоритм вычисления нейронов в слое Кохонена. Программа состоит из трех основных частей. Первая выполняет добавление пользователей, вторая выполняет идентификацию и третья хранение голосовых записей.

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 усовершенствования и добавления новых функций, что делает его не только выгодным программным продуктом, но и перспективным проектом для развития и получения прибыли.

Библиографический список

[Bosi M., 2003] Introduction to digital audio coding and standards / M. Bosi, R.E. Goldberg - Springer Science+Business, Media USA. - 2003. - 434 p.

[You Y., 2010] AudioCoding: Theory and Applications / Y. You - NY: Springer-er, 2010 - 349 p.

[Загуменнов, А. П., 1999] Компьютерная обработка звука./ А. П. Загуменнов - М. : ДМК, 1999. - 384 с

VOICE USER IDENTIFICATION IN ACCESS CONTROL SYSTEMS

Menshakov P.A., Murashko I.A.

*Department of Information Technology
Gomel State Technical University named by P.O.*

*Suhoi
Gomel, Belarus*

pmenshakov@gmail.com

iamurashko@tut.by

At the moment, the primary means of access control checkpoints are equipped with a variety of means of access control. But most of the access controls are expensive. And most of the costs fall on the allocation of personal identification means for each user. The solution to this problem is to voice recognition. The use of biometrics eliminates the chips and access cards, identification means to eliminate loss and theft. And the use of the voice will eliminate the expensive equipment to read the data.

Introduction

Currently, voice recognition, biometrics as a whole, already widespread. The simplest example - fingerprint scanners installed on almost every laptop. Biometrics recognition system involves people on one or more physical or behavioral traits. In the field of information technology, biometric data is used as a form of identity management and access control access. Also, biometric analysis is used to identify people who are under the supervision of. Just provide biometrics and behavioral analysis of the object. These include walking, gestures, etc. Authorization process, the use of biometrics, is quite simple. Using an apparatus for varying the characteristics of an identified current data scanned and compared with previous data. Biometric systems have a number of important advantages. Biometrics uses the properties of the human body and its behavior, which make them unique. Unlike paper identifiers from a password or personal identification number (PIN), biometric characteristics can not be subjected to theft, can not be lost or forgotten. Quite a long time fingerprints are used to identify criminals and prevent theft or fraud. Some people are able to mimic the voice,

but it requires special skills that are not often met in everyday life.

Main Part

To implement voice recognition is necessary to make a specific course of action. With a microphone turns voice recording identified and sent to the computer. The optimal reception is WAV file, since handling ease. The resulting voice recording should be divided into frames.

The next step is to eliminate the undesirable effects and noises. It is necessary to record obtained at different time correspond to each other, regardless of external factors. There are many ways in which to reduce the effects of noise.

To date, the most successful are the voice recognition system, using the knowledge of the hearing aid device. They are based on the fact that the ear interprets sounds not linearly but in a logarithmic scale. In view of these features is necessary to bring the frequency response for each frame of mels.

This is the last step required for the subsequent conversion to vector features, which, compared to the base of voice recordings. The vector will comprise mel-coefficient coefficients.

The resulting feature vector is added to the database for later comparison. But a more accurate alternative is to use multiple entries of the same voice. A predetermined number of voice samples may be used to train the neural network. We used learning without a teacher, because it is much more plausible model of learning in the biological system. Kohonen developed and many others, it does not need to output the target vector and therefore.

Conclusion

The result of this study is a modular application that voice recognition by using an upgraded algorithm for computing the neurons in the Kohonen layer. The program consists of three main parts. The first performs the addition of users, the second and third authenticates voice recordings.

The study showed that the resulting algorithm can significantly speed up the program of voice identification. This upgrade allows you to use the program at the enterprises with the influx of users.

The same software system is very flexible and has plenty of room for further improvements and adding new features, making it not only profitable software product, but promising project for development and profit.