

ГОЛОСОВАЯ ИДЕНТИФИКАЦИЯ ПОЛЬЗОВАТЕЛЯ В СИСТЕМАХ КОНТРОЛЯ ДОСТУПА

П.А. Меньшаков, И.А. Мурашко
Кафедра «Информационные технологии»,
Гомельский государственный технический университет имени П.О. Сухого
Гомель, Республика Беларусь
E-mail: pmenshakov@gmail.com, iamurashko@tut.by

В виду постоянного развития любого предприятия, появляются данные, которые необходимо защитить. 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]. Так же биометрия предусматривает и поведенческий анализ объекта. К ним относятся ходьба, жесты и т.п.

Процесс авторизации, при использовании биометрии, довольно прост. При помощи устройства, предназначенного для получения той или иной характеристики, сканируются текущие данные идентифицируемого и сравниваются с полученными ранее данными.

Биометрические системы имеют ряд серьезных преимуществ: биометрия использует свойства человеческого тела и его поведение, что делает эти данные уникальными (для того, чтобы, при помощи своего собственного отпечатка пальцев, предоставить чужой отпечаток пальца или сделать радужную оболочку своего глаза похожей на чью-то другую требуются довольно редкие и сложные технологии); в отличие от бумажных идентификаторов (паспорт, водительские права, удостоверение личности), от пароля или персонального идентификационного номера (ПИН), биометрические характеристики не могут быть подвергнуты краже, не могут быть утеряны или забыты. Довольно длительное время отпечатки пальцев используются для идентификации преступников и предотвращения воровства или мошенничества. Некоторые люди уме-

ют имитировать голоса, но, это требует особых навыков, которые не часто повстречаешь в обычной жизни.

Принцип голосовой идентификации

Для реализации голосовой идентификации 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 рис. 1.

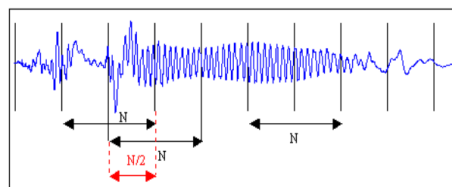


Рис. 1 – График звуковой волны

Данное действие необходимо для более простой работы с записанной звуковой дорожкой. Далее все вычисления будут производиться с каждым кадром в отдельности.

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 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. Они базируются на том, что ухо интерпретирует звуки не линейно, а в логарифмическом масштабе. В виду данных особенностей необходимо привести частотную характеристику каждого кадра к «мелам». Зависимость представлена на рис. 2.

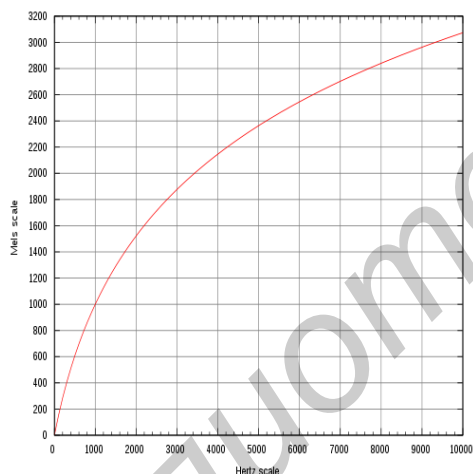


Рис. 2 — График зависимости высоты звука (в мелах) от его частоты

Для перехода к «мел» характеристики 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.

Рекомендуется использовать несколько записей одного и того же голоса. Тогда результирующий вектор представляется как среднее арифметическое векторов, характеризующих отдельные кадры речи. Для повышения точности распознавания просто необходимо усреднять результаты не только между кадрами, но и учитывать показатели нескольких речевых образцов. Имея несколько записей голоса, разумно не усреднять показатели к одному вектору, а провести кластеризацию, например с помощью метода k -средних.

Итогом данного исследования стало модульное приложение, осуществляющее голосовую идентификацию пользователя. Программа состоит из трех основных частей. Первая выполняет добавление пользователей, вторая выполняет идентификацию и третья рассылает информацию об идентифицированных пользователях.

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