УДК 004.822

ГОЛОСОВАЯ ИДЕНТИФИКАЦИЯ ПОЛЬЗОВАТЕЛЯ В СИСТЕМАХ КОНТРОЛЯ ДОСТУПА

Меньшаков П.А., Мурашко И.А.
Гомельский государственный технический университет имени П.О. Сухого
E-mail: pmenshakov@gmail.com; iamurashko@tut.by

Abstract. *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 melcepstral 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.*

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 1. Данное действие необходимо для более простой работы с записанной звуковой дорожкой.

Далее все вычисления будут производится с каждым кадром в отдельности.

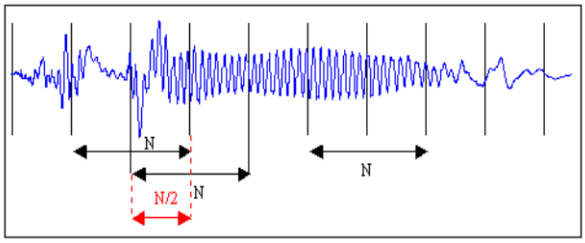

Рисунок 1 – График звуковой волны

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 весомую функцию "Окно Хемминга":

$$\omega(n) = 0.53836 - 046164 * \cos(\frac{2\pi n}{N-1}), \quad (1)$$

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":

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

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. 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».

Для перехода к «мел» характеристике используется следующая зависимость:

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

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:

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

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. Развитая 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.