

Arduino MEGA 2560. Рассматриваемая система за счет представленного вышеалгоритма обеспечивает отличную защиту от несанкционированного доступа в помещение, звуковую и визуальную сигнализацию, при попытке несанкционированного доступа, оперативную возможность смены секретной кодовой комбинации и простой, интуитивно понятный для пользователя интерфейс – всё это выделяет ее на фоне других систем.

Список использованных источников:

1. Соммер У. Программирование микроконтроллерных плат Arduino/Freeduino / СоммерУ // Для радиолюбителей. – СПб.: БХВ – Петербург, 2012 г. – 256 с.
2. Сайт, посвященный программированию платформы Arduino [Электронный ресурс]: 2016 г. URL: <http://arduino.ru/>
3. Сайт, посвященный программированию платформы Arduino [Электронный ресурс]: 2016 г. URL: <http://arduino.cc/>

СИСТЕМА ДЕТЕКТИРОВАНИЯ КЛЮЧЕВЫХ ФРАЗ В РЕЧЕВОМ СИГНАЛЕ

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Суша А. В.

Вашкевич М. И. – к-т. техн. наук, доцент

В работе приводится краткая характеристика основных методов детектирования (поиска) ключевых фраз в речевом сигнале. Так же описывается алгоритм работы системы, построенной с применением метода поиска по шаблону. Сравнение шаблона ключевой фразы с фрагментом сигнала осуществляется с помощью DTW-алгоритма.

Поиск ключевых фраз (англ. – keyword spotting) – речевая технология, связанная с решением задачи поиска и локализации заданного слова или целой фразы в речевом потоке. Задача поиска ключевых слов в речевом потоке может быть разделена на несколько направлений:

- непосредственно выявление и определение места ключевого слова или словосочетания (поиск в фонограмме речи, речевых базах, поиск в реальном времени);
- распознавание команд в слитном потоке речи;
- понимание смысла речи посредством поиска ключевых слов или фраз, для диалоговых систем.

Для решения задачи поиска ключевых фраз было разработано множество различных способов, из которых можно выделить основные методы. Ниже приведены широко используемые в настоящее время методы поиска ключевых фраз [1].

Метод *поиска по шаблону* является наиболее распространённым. Суть его заключается в том, что для каждого ключевого слова записывается несколько вариантов его произнесения различными дикторами, на основе чего создаётся шаблон, который используется для организации поиска данного слова в речевом потоке. Главным недостатком такого подхода является то, что создание шаблона требует запись каждого слова микрофоном (слово нельзя просто ввести с клавиатуры).

Скрытые марковские модели (сокращённо – СММ) широко используются для решения задачи поиска ключевых фраз. Для каждого ключевого слова создаётся соответствующая СММ, которая используется при поиске в речевом потоке, и одну «модель мусора» для всех остальных слов. Временная последовательность символов ключевых слов и мусор символов формируется в результате распознавания речевой последовательности. Недостатком метода является то, что для каждого нового ключевого слова нужно не только обучать новую СММ модель, но также нужно заново обучать модель мусора.

Стоит отметить метод *поиска ключевых фраз по фонемной (слоговой) решётке*. Метод основывается на конструировании большой сети возможных звуков (слов) в различные моменты времени, по которой осуществляется поиск возможных произношений слова или фразы. Основное преимущество этого метода в том, что он обладает большой гибкостью: даже если фонема ключевого слова не является лучшей гипотезой между узлами решётки, она все равно сохраняется в результате распознавания. Но в то же время, построение такой решётки довольно трудоёмкий процесс.

В качестве основы был выбран метод поиска ключевых фраз по шаблону. Данный метод был выбран в связи с его относительной простотой и отсутствием необходимости предварительного обучения и переобучения. Описанный далее алгоритм работы системы детектирования ключевых фраз в речевом сигнале основывается на алгоритмах, описанных в публикациях [2, 3]. Структурная схема алгоритма приведена на рисунке 1. Рассмотрим алгоритм работы системы по шагам.

Подготовительный шаг. Осуществляется фрагментация шаблона ключевого слова на M фреймов, длиной 30 мс. На основании из каждого i -го фрейма формируется вектор характеристических признаков $\bar{a}_i, i = 0, M - 1$. В качестве этих признаков были выбраны *мел-частотные кепстральные коэффициенты* (сокращённо – MFCC) [4]. Длина вектора характеристических признаков выбирается произвольно, в работе использовались 13 коэффициентов. В результате этих действий, формируется последовательность векторов характеристических признаков $A_M = \{\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{M-1}\}$.

Шаг 1. Из речевого потока выбирается фрагмент данных, равный по длительности $2M$ фреймов шаблона ключевого слова. Далее осуществляется формирование векторов MFCC $\bar{b}_j, j = \overline{0, 2M-1}$. В результате получаем последовательность векторов характеристических признаков $B_{2M}^n = \{\bar{b}_1, \bar{b}_2, \dots, \bar{b}_{2M-1}\}$ фрагмента речевого сигнала в некоторый момент времени n .

Шаг 2. Из последовательности B_{2M}^n в момент времени n выбираются несколько подпоследовательностей $B_L^i, i = \overline{0, H-1}$, таких, что $L \in 0,5M; 2M$. Число этих подпоследовательностей выбирается произвольно, но следует отметить, что большое их число существенно замедлит работу алгоритма.

Шаг 3. Каждая i -я подпоследовательность B_L^i сравнивается с последовательностью ключевого слова A_M посредством *DTW-алгоритма* [5], в результате строится таблица минимальных дистанций $d_k, k = \overline{0, K-1}$ между последовательностями $D_K^i = \{d_0, d_1, \dots, d_{K-1}\}$. Стоит отметить, что DTW-алгоритм позволяет сравнивать последовательности разных размеров, а также помимо привычных скалярных значений этих последовательностей может оперировать и векторами, при условии, что имеется функция расчёта расстояния между ними. На основании этой таблицы рассчитывается оценка расстояния между последовательностями по формуле:

$$D^i = \frac{1}{N+L} \cdot \sum_{k=0}^{K-1} d_k \quad (1)$$

Шаг 4. На основании полученных значений оценочных расстояний определяется значение локальной минимальной дистанции $LMD\ n = \min D^i, i = \overline{0, H-1}$, где H – число подпоследовательностей, а n – рассматриваемый момент времени.

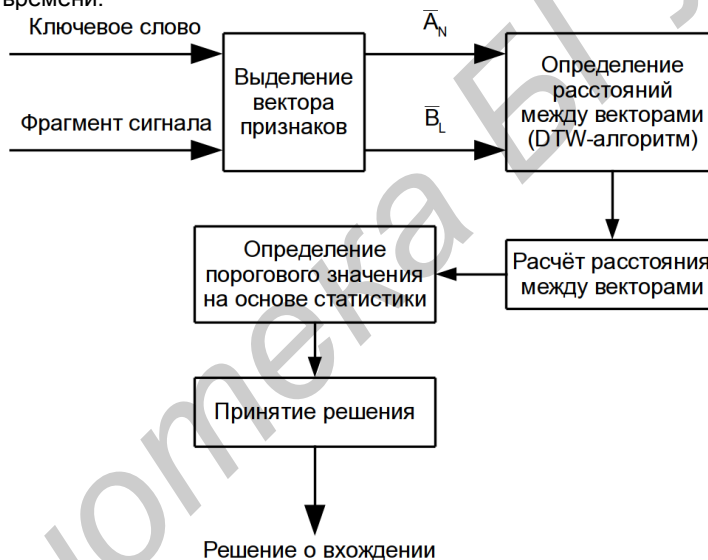


Рис. 2. Структурная схема алгоритма работы системы детектирования ключевых фраз в речевом сигнале

Шаг 5. Действия, описанные в шагах 1-4 выполняются до полной проверки речевого потока. Фрагменты сигнала выбираются с некоторым перекрытием.

Шаг 6. На основании полученных значений локальных минимальных дистанций определяется пороговое значение, которое используется для последующего определения вхождения ключевого слова:

$$TH = M \cdot LMD - c \cdot \sigma, \quad (2)$$

где M – математическое ожидание;

σ – среднеквадратичное отклонение.

Шаг 7. Используя полученное пороговое значение определяется вхождение ключевого слова в рассмотренные ранее моменты времени n , выбранные с перекрытием. Для этого рассчитывается число «попаданий» ключевого слова в некоторой окрестности момента n , с максимальным удалением S :

$$hit_l = \sum_{l=n-S}^{n+S} TH - LMD\ l. \quad (3)$$

Значение hit_l указывает на вхождение ключевого слова в фрагмент речевого сигнала в некоторый момент времени n .

Стоит отметить, что в текущем виде описанный выше алгоритм не пригоден для использования в системах реального времени в связи с необходимостью сбора статистики для определения порогового значения, которое используется при принятии решения. Работы по преодолению данного ограничения производятся.

Список использованных источников:

1. Chandra, Dr.E. Keyword Spotting: An Audio Mining Technique in Speech Processing / Dr.E. Chandra, K.A. Senthildevi // IOSR Journal of VLSI and Signal Processing. – 2015. – №5. – с. 22-27.
2. Barakat, M. S. Keyword spotting based on the analysis of template matching distances / M. S. Barakat, C. H. Ritz, D. A. Stirling // Signal Processing and Communication Systems. – 2011. – с. 1-6.
3. John Sahaya Rani Alex. Spoken Utterance Detection Using Dynamic Time Warping Method Along With a Hashing Technique / John Sahaya Rani Alex, Nithya Venkatesan // International Journal of Engineering and Technology. – 2014. – №2 (6). – с. 1100-1108.
4. Zheng Fang Comparison of Different Implementations of MFCC / Zheng Fang, Zhang Guoliang, Song Zhanjiang // J. Comput. Sci. A Technol. – 2011. – №6 (16). – с. 582-589.
5. Eamonn, J.K. Derivative Dynamic Time Warping / J.K. Eamonn, J.P Michael // SIAM International Conference on Data Mining. – 2001. – с. 11.

СИСТЕМА ИДЕНТИФИКАЦИИ ДИКТОРА ПО ОБРАЗЦУ ГОЛОСА НА БАЗЕ РЕЧЕВЫХ ЗАПИСЕЙ

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Вашкевич Г. С.

Вашкевич М. И. – канд. техн. наук, доцент

В данной работе описывается общий подход к построению систем идентификации диктора, проводится обзор современных методов и подходов к решению данной задачи, а также представляется пример реализации системы идентификации диктора на основе векторного квантования.

В условиях интенсивного развития систем распознавания речи все чаще перед разработчиками стоит задача не только распознать речь, но и установить личность диктора. Такие системы нашли широкое применение в криминалистике, системах контроля доступа и верификации, а также в обработке речевых баз данных.

На данный момент основными методами идентификации дикторов являются методы, основанные на следующих моделях: модели гауссовых смесей (англ. Gaussian Mixture Models, GMM), модель, основанная на методе опорных векторов (англ. support vector machine, SVM), и модель, основанная на векторном квантовании [1].

В методе, использующем гауссовы смеси, модель голоса диктора представляет собой аппроксимацию распределения используемых речевых признаков смесью гауссовых распределений. Недостатком данного метода является его вычислительная сложность, и сложность реализации [2].

Метод опорных векторов используется для разделения сложных областей в пространстве признаков речевого сигнала на два класса оптимальной (с точки зрения алгоритма) гиперплоскостью. В этом методе отыскиваются образцы, находящиеся на границах между двумя классами и через них строятся две параллельные гиперплоскости, с максимальным расстоянием между ними. После этого на равном удалении от каждой из найденных гиперплоскостей строится разделяющая гиперплоскость. Таким образом все обучающие вектора модели диктора находятся по одну сторону от разделяющей гиперплоскости. Если на вход системы подать тестовый вектор, то система будет сравнивать его положение в пространстве с положением гиперплоскости. Если тестовый вектор находится по одну сторону от гиперплоскости вместе с обучающими векторами модели диктора – то система принимает решение о том, что данный тестовый характеристический вектор принадлежит диктору.

Модель, основанная на векторном квантовании, строит модель диктора на основе кодовой книги, полученной путем векторного квантования обучающей выборки характеристических векторов диктора. Векторное квантование в общем случае осуществляется при помощи алгоритма кластеризации K-средних. Принятие решения о принадлежности тестового характеристического вектора диктору осуществляется при превышении порогового значения меры близости между входным вектором и, хотя бы одним вектором из кодовой книги.

В общем случае системы идентификации дикторов имеют структуру, представленную на рисунке 1: