

ИДЕНТИФИКАЦИЯ ЛИЧНОСТИ ПО ГОЛОСУ В ОХРАННЫХ СИСТЕМАХ

В настоящее время все более популярными становятся охранные системы, например, системы управления доступом, в частности, системы идентификации личности. Существуют различные способы решения задачи определения индивидуальных особенностей человека. В данной работе решалась задача идентификации личности по голосу.

Разрабатываемую систему предполагается применять при скрытом прослушивании помещений с использованием миниатюрных микрофонов (Sensor) и специальных плат (Sound card) аудиоввода (например, «Шорох-5» и «Ewclid-A» соответственно), как показано на рисунке 1.

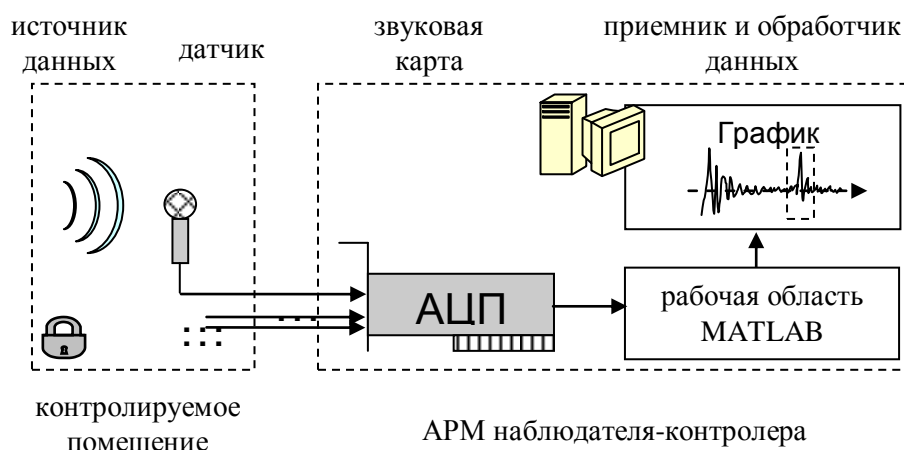


Рис.1. Схема работы системы идентификации диктора

Основными проблемами идентификации дикторов являются:

- выделение параметров речевого сигнала, однозначно определяющих индивидуальные особенности голоса конкретного человека;
- создание и обучение нейронной сети, адекватно с заданной точностью распознающей учителя.

В качестве признаков, определяющих индивидуальные особенности речи, были выбраны: основной тон и формантные частоты (первые четыре). Основной тон для гласных определяется с помощью действительного кепстра: $c_x := \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} \log(|X(e^{j\omega})|) \cdot e^{j \cdot \omega \cdot n} d\omega$

Формантные частоты для речевого сигнала $s(i)$ определяются с помощью линейного предсказания:

- исходя из решения уравнения Юла-Уолкера для авторегрессивной модели:

$$\sum_{k=1}^p a_k \cdot R(j-k) := -R(j), \text{ где } R(j-k) := \sum_n s(n-j) \cdot s(n-k), \quad R(j) := \sum_n s(n) \cdot s(n-j)$$

- путем аппроксимации спектра, где модель частотного спектра сигнала $s(i)$ имеет вид:

$$|H(j \cdot \omega)| := \frac{G^2}{\left(1 + \sum_{k=1}^p a_k \cdot e^{-j \cdot k \cdot \omega}\right)^2}, \text{ где } G - \text{коэффициент усиления}$$

Система идентификации голоса моделировалась в программном пакете MATLAB с использованием комплекса программ пакетов Data Acquisition Toolbox и Neural Network Toolbox. Для решения задач идентификации была выбрана двухслойная нейронная сеть прямого распространения, обучаемая по процедуре обратного распространения ошибки (newff: feed-forward backpropagation network). В качестве метода обучения использовался LM алгоритм (reduced memory Levenberg-Marquardt algorithm) как наиболее быстрый и менее ресурсоемкий. Размер скрытого слоя нейронной сети (10 нейронов) и объем необходимых данных для обучения определялись эмпирически:

- размер сети должен быть минимально необходимым для решения поставленной задачи;
- объем данных и размер сети должны быть достаточны для симуляции обучения.

В случае невыполнения второго требования, может проявляться эффект «переобученности» нейронной сети. Указанный эффект можно описать как отсутствие существенной подстройки весовых коэффициентов сети, что проявляется в быстром достижении заданного уровня допустимой ошибки и возникает по причине скорого нахождения пути прохождения через все указанные точки.

В качестве функций активации используются логистические функции гиперболического тангенса \tansig и логарифма \logsig . Обучение строится на основании соответствия выборки индивидуальных параметров речи определенным дикторам. Объем выборки r определяется исходя из следующего правила: $r \approx w * n * 10$, где n – число признаков речи, w – размер нейронной сети. Выделение данных для обучения нейронной сети и идентификации дикторов осуществляется из определенных слов с ярко выраженными гласными фонемами.

В результате работы получена система обработки речевого сигнала, решающая следующие задачи:

- выделение времени основного тона и формантных частот в реальном времени;
- запись по требованию соответствующих индивидуальных параметров речи в файл, по данным которого обучается нейронная сеть;
- идентификация дикторов с помощью обученной нейронной сети с точностью, определяемой уровнем допустимой ошибки.

По результатам моделирования, для выборки в 500 вектор-признаков для каждого из дикторов вероятность ошибки идентификации заданного диктора составляет не более 5%.

Созданная система верификации диктора может найти свое применение в охранных системах контроля доступа, а также при негласном прослушивании, что позволит отследить наличие определенных зарегистрированных лиц в контролируемых помещениях.

В настоящее время разработанная система модифицируется с целью создания удобного пользователю интерфейса на Windows платформе. Предполагается развить данную работу для определения содержания речи дикторов.

Во время разработки системы использовались следующие продукты Microsoft:

- Microsoft Office 2003;
- Microsoft Windows NT4.0, 2000, XP;
- Блокнот и другие приложения, входящие в операционную систему Microsoft Windows.