

# РАСПОЗНАВАНИЕ ДИКТОРОВ ПО СПЕКТРАЛЬНЫМ ПРИЗНАКАМ РЕЧЕВЫХ СИГНАЛОВ

Безрук В.М., Пресняков С.А., Зарицкий В.И.

Харківський національний університет радіоелектроніки, E-mail: [valerii.bezruk@nure.ua](mailto:valerii.bezruk@nure.ua)

*Аннотация – В работе рассматриваются особенности применения нетрадиционного метода распознавания заданных сигналов при наличии неизвестных сигналов для решения задачи распознавания дикторов по спектральным признакам речевых сигналов. Приводятся результаты экспериментальных исследований распознавания дикторов на выборках реальных сигналов.*

*Ключевые слова – распознавание дикторов, речевые сигналы, спектральные признаки, решающее правило распознавания, неизвестные сигналы.*

## I. Введение

Задача распознавания дикторов является актуальной задачей области речевых технологий [1-3]. Эта задача объединяет идентификацию и верификацию дикторов. Идентификация диктора — процесс определения личности по образцу голоса путём сравнения данного образца с шаблонами, сохранёнными в базе. Результатом процесса идентификации является список кандидатов. Реализующая система может выдавать список фиксированного размера либо принимать решение о включении пользователя в список кандидатов на основании заданного порога. Если предусмотрена возможность того, что в процессе идентификации будет участвовать пользователь, не зарегистрированный в системе, то говорят об идентификации на открытом множестве. В идеальном случае для такого пользователя система должна выдать пустой список. Если все пользователи, проходящие процедуру идентификации, зарегистрированы в системе, то говорят об идентификации на замкнутом множестве. Верификация диктора — процесс, при котором с помощью сравнения представленного образца речевого сигнала с хранимым в базе шаблоном проверяется запрошенная идентичность. Результатом верификации является положительное либо отрицательное решение).

Работа систем распознавания дикторов содержит два основных этапа: регистрация пользователей в системе и сам процесс распознавания (попытка идентификации или верификации). Пользователи предварительно регистрируются в системе, записав свои голоса. Образец голоса каждого диктора обрабатывается с целью извлечения признаков, которые могут быть использованы для распознавания. На основе извлечённых признаков строятся модели («шаблоны») пользователей. Модель представляет собой некоторую структуру (алгоритм), позволяющую при данных признаках оценить степень подобия. Обработка сигнала имеет целью выделить из речевого сигнала информативные признаки, релевантные для задачи распознавания дикторов. Априори невозможно оценить, какие признаки более подходят для распознавания дикторов. Процесс определения подходящих признаков заключается в переборе возможных вариантов признаков с последующей экспериментальной оценкой.

Существуют различные методы построения информативных признаков и правила принятия решений в задаче распознавания [1-4]. В данной работе предлагаются некоторые результаты экспериментальных исследований по распознаванию заданных дикторов с использованием спектральных признаков речевых сигналов и неклассического решающего правила селекции заданных сигналов при наличии неизвестных сигналов.

## II. Решающее правило распознавания заданных сигналов при наличии неизвестных сигналов

В отношении наблюдений речевых сигналов  $\bar{X}$  могут быть выдвигаться  $(M+1)$ -а гипотезы:  $H^i, i = \overline{1, M}$  — для заданных сигналов,  $H^0$  — для неизвестных сигналов, объединённых в  $(M+1)$ -й класс. Плотности вероятности заданных сигналов  $W(\bar{x} | \bar{\alpha}^i), i = \overline{1, M}$  задаются с точностью до векторных параметров  $\bar{\alpha}^i, i = \overline{1, M}$ , а для  $(M+1)$ -го класса плотность вероятности неизвестна. Заданы также априорные вероятности гипотез  $P(H^i) = P_i$ , причём  $\sum_{i=0}^M P_i = 1$ . Полагается также, что заданы обучающие выборки реализаций  $M$  сигналов  $\{\bar{x}_r^i, r = \overline{1, n_i}; i = \overline{1, M}\}$ , а обучающая выборка для  $(M+1)$ -го класса неизвестных сигналов ( $i = 0$ ) отсутствует.

Такие исходные данные для распознавания сигналов могут быть охарактеризованы термином «повышенная априорная неопределённость» и определяют нетрадиционную задачу селекции и распознавания заданных сигналов при наличии неизвестных сигналов [4].

Нерандомизированное решающее правило распознавания осуществляет разбиение выборочного пространства сигналов на  $(M+1)$ -ну непересекающуюся область. При этом для задачи распознавания заданных сигналов при наличии неизвестных сигналов имеют место такие составляющие вероятности принятия ошибочных решений: за счёт перепутывания  $M$  заданных сигналов между собой, за счёт отнесения заданных сигналов к  $(M+1)$ -му классу неизвестных сигналов, за счёт отнесения сигналов из  $(M+1)$ -го класса неизвестных сигналов к  $M$  заданным сигналам.

Для решения такой неклассической задачи селекции и распознавания заданных сигналов при наличии неизвестных сигналов может быть использовано следующее решающее правило [5]:

при выполнении условий:

$$H^i : P_i W(\bar{x} / \bar{\alpha}^i) \geq \lambda^i, i = \overline{1, M} \quad (1)$$

$$P_i W(\bar{x} / \bar{\alpha}^i) \geq P_l W(\bar{x} / \bar{\alpha}^l), l = \overline{1, M}, l \neq i, \quad (1b)$$

принимается гипотеза о действии  $i$ -го заданного сигнала, если выполняются неравенства

$$H^0 : P_i W(\bar{x} / \bar{\alpha}^i) < \lambda^i, i = \overline{1, M}, \quad (1c)$$

то принимается гипотеза  $H^0$  о действии неизвестных сигналов;

Здесь пороговое значение  $\lambda$  определяется из условия обеспечения заданной вероятности правильного распознавания заданных сигналов.

В решающем правиле (1) не используется информация о плотности распределения  $(M+1)$ -го класса сигналов и не требуется его обучающая выборка. Постановка и решение рассмотренной задачи распознавания — это формализация требования содержательного характера о необходимости выделить (селектировать) и распознать  $M$  заданных сигналов и отнести в  $(M+1)$ -й класс неизвестные сигналы, информация о которых отсутствует.

Будем рассматривать случай, когда для описания речевых сигналов используются спектральные признаки. В частности, рассмотрим представление сигналов в виде отсчетов амплитудного спектра в базисе дискретных экспоненциальных функций (ДЭФ). В предположении гауссовского распределения и некоррелированности координат вектора спектральных отсчетов решающее правило (1) упрощается и принимает следующий вид [5]:

$$H^1: \sum_{j=1}^N \frac{(c_j - \mu_{jc}^i)^2}{(\sigma_{jc}^i)^2} \leq \Delta_c^i, \quad i = \overline{1, M}$$

$$\sum_{j=1}^N \frac{(c_j - \mu_{jc}^1)^2}{(\sigma_{jc}^1)^2} \leq \sum_{j=1}^N \frac{(c_j - \mu_{jc}^k)^2}{(\sigma_{jc}^k)^2}, \quad k = \overline{1, M}$$

$$H^0: \sum_{j=1}^N \frac{(c_j - \mu_{jc}^i)^2}{(\sigma_{jc}^i)^2} > \Delta_c^i, \quad i = \overline{1, M}$$

Здесь  $c_j$  – отсчеты амплитудного спектра сигналов в базисе ДЭФ,  $\mu_{jc}^i$ ,  $(\sigma_{jc}^i)^2$  – оценки математических ожиданий и дисперсий спектральных отсчетов  $c_j$ , которые оцениваются по обучающим выборкам заданных сигналов;  $N$  – размерность спектрального представления сигналов;  $\Delta_c^i$  – некоторые пороговые значения, выбираемые из условия обеспечения заданной вероятности правильного распознавания заданных сигналов.

Указанное решающее правило использовано для решения задачи распознавания дикторов по спектральным признакам речевых сигналов.

### III. Результаты экспериментальных исследований распознавания дикторов

Для проведения экспериментальных исследований были накоплены выборки из 15 реализаций речевых сигналов для 30 дикторов, которые произносили слово «Hello». Записи реализаций речевых сигналов оцифровывались с частотой дискретизации 8 кГц и представлялись в виде векторов наблюдений размерностью 4096 отсчетов. Сигналы были записаны с помощью программы Recorder для Iphone 7 (IOS Operationsystem) со следующими параметрами: Format - Wav, Samplerate - 8000 Hz, Channel - Mono, Depth - 32 bit. Далее для каждой реализации сигналов вычислялся амплитудный спектр в базисе ДЭФ размерностью  $N=4096$ .

Исследования были проведены путем статистических испытаний. При этом решающее правило было программно реализовано в среде MATLAB. Обучающие выборки реализаций речевых сигналов для заданных дикторов использовались для оценивания параметров решающего правила – средних значений и дисперсий спектральных отсчетов речевых сигналов. Контрольные выборки реализаций сигналов использовались для получения оценок качества распознавания заданных сигналов при наличии неизвестных сигналов. В качестве заданных сигналов выбраны речевые сигналы 5-ти дикторов. Остальные 25 дикторов рассматривались как неизвестные. На рис. 1 приведены усредненные амплитудные спектры речевых сигналов для заданных дикторов.

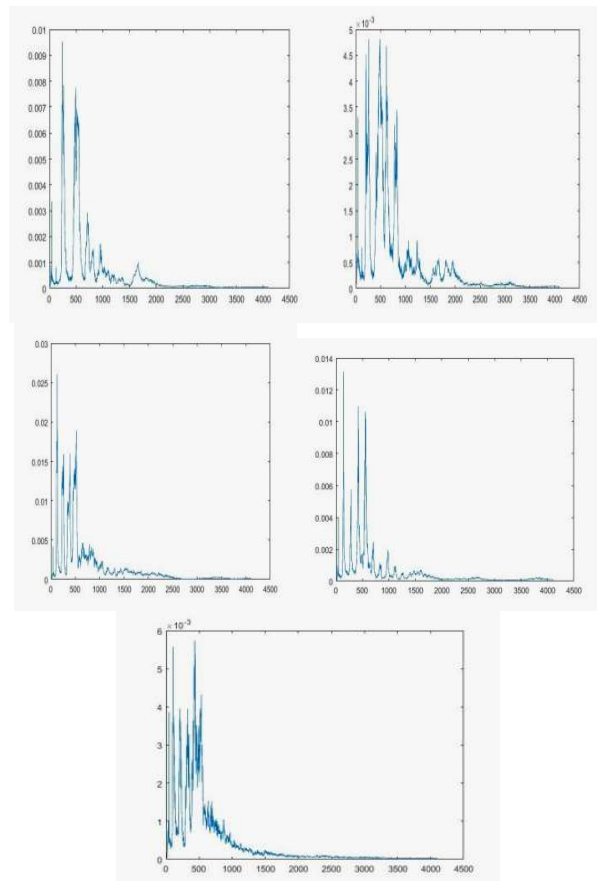


Рис. 1. Усредненные амплитудные спектры речевых сигналов для заданных дикторов

В результате статистических испытаний по контрольным выборкам реализаций были получены оценка средней вероятностей правильного распознавания заданных дикторов – 0.93, оценка средней вероятности ошибочного отнесения заданных дикторов к классу неизвестных – 0.07, а также оценка средней вероятности ошибочного отнесения неизвестных дикторов к заданным – 0.01.

Таким образом, экспериментальные исследования на выборках реальных речевых сигналов подтверждают возможность решения задачи распознавания заданных дикторов при наличии класса неизвестных дикторов по спектральным признакам речевых сигналов.

### IV. Список литературы

1. Furui S. Digital Speech Processing. Synthesis and Recognition // Marcel Dekker. - New York, 1989.
2. Campbell W., Campbell J., Reynolds D., Jones D., Leek T. Phonetic speaker recognition with support vector machines // Advances in Neural Information Processing Systems. Vol.16. - MIT Press, Cambridge, 2004.
3. Ing-Jr Ding, Chih-Ta Yen, Yen-Ming Hsu. Developmentsof Machine Learning Schemes for Dynamic Time-Wrapping-Based Speech Recognition // Mathematical Problemsin Engineering. - 2013.
4. Watanabe S. Methodologies of pattern recognition. - Academic Press, 2014.
5. Безрук В.М., Певцов Г.В. Теоретические основы проектирования систем распознавания сигналов для автоматизированного радиоконтроля. - Харьков: Коллегиум, 2007.