

УДК 004.934.1'1

Е.Е. Федоров  
Донецкая академия автомобильного транспорта  
fee75@mail.ru

## Методика идентификации диктора на основе модифицированной вероятностной нейронной сети

*Для повышения точности и надежности идентификации диктора, повышения быстродействия процедуры обучения в статье предлагается методика идентификации диктора на основе модифицированной вероятностной нейронной сети, базирующаяся на формировании системы векторов признаков речевого сигнала, векторов параметров моделей классификаторов дикторов, моделей классификаторов дикторов, правил принятия решения. Для предложенной методики проводится численное исследование.*

**Ключевые слова:** модифицированная вероятностная нейросеть, идентификация диктора, система векторов признаков, модели классификаторов дикторов, правила принятия решений.

### Введение

*Общая постановка проблемы.*

В современной отечественной и мировой практике активно ведутся разработки интеллектуальных систем поддержки принятия решений, использующих естественно-языковой интерфейс. Для получения доступа к управлению такими системами необходимо предварительно осуществить идентификацию диктора. Кроме того, идентификация диктора играет важную роль в фоноскопической экспертизе.

*Анализ исследований.*

Согласно [1], многие современные методы идентификации диктора базируются на анализе формант речевого сигнала (например, Р-метод). Однако такой подход обладает следующими недостатками: сложность точного выделения формант, не используются настройка весовых коэффициентов признаков, многие признаки являются коррелированными. Другой подход [1] заключается в анализе геометрических характеристик участков речевого сигнала, однако используемые в нем признаки очень чувствительны к физическому состоянию человека и окружающей среде. Третий подход [2] заключается в использовании признаков, выделяемых с помощью метода линейного прогнозирования (ЛП). Четвертый подход [3] основан на мел-частотных кепстральных коэффициентах (MFCC). Третий и четвертый подходы для сопоставления распознаваемого вектора признаков и эталона обычно используют плотность нормального распределения [4]. В этих подходах также не используются настройка весовых коэффициентов признаков, многие признаки являются коррелированными.

*Цель и постановка задач исследования.*

Цель – разработать методику идентификации диктора на основе модифицированной вероятностной нейронной сети.

Решаемые задачи:

- формирование системы векторов признаков;
- формирование векторов параметров модели классификаторов дикторов;
- формирование моделей классификаторов дикторов и решающих правил;
- проведение численного исследования предложенной методики.

### 1. Формирование системы векторов признаков

В режиме обучения для фиксации речевых особенностей дикторов формируются система из 14 векторов признаков речевого сигнала [5]:

- коэффициенты ЛП (линейного прогнозирования);
- коэффициенты отражения ЛП;
- автокорреляция ЛП;
- кепстр ЛП;
- площади поперечных сечений акустической трубы;
- нормированная автокорреляция;
- нормированный энергетический спектр ЛП;
- меры контрастности ЛП;
- нормированный энергетический спектр ДПФ (дискретного преобразования Фурье);
- меры контрастности ДПФ;
- меры контрастности ДВП (дискретного вейвлет-преобразования);
- меры контрастности НВП (непрерывного вейвлет-преобразования);
- количество импульсов равной длины;
- MFCC (мел-частотные кепстральные

коэффициенты).

Сформированные вектора признаков отражают частотные, автокорреляционные и энергетические характеристики речи дикторов, его речевого тракта и слухового аппарата и используются для создания векторов параметров моделей классификаторов дикторов.

## 2. Формирование векторов параметров моделей классификаторов дикторов

Для каждого диктора записывается обучающий набор реализаций речевого сигнала, причем каждая реализация разбивается на фреймы (участки равной длины). Для каждого  $i$ -го фрейма на основе вектора признаков  $u$ -го типа

$$\mathbf{x}_{ui} = (x_{ui1}, \dots, x_{uik}, \dots, x_{uiN_u}), \quad u \in \overline{1,14}, \quad i \in \overline{1, N_u^{(1)}}$$

создается эталон в виде

$$\mathbf{m}_{ui} = (m_{ui1}, \dots, m_{uik}, \dots, m_{uiN_u}), \quad u \in \overline{1,14}, \quad i \in \overline{1, N_u^{(1)}},$$

$$m_{uik} = \frac{x_{uik} - x1_{uk}}{x2_{uk} - x1_{uk}},$$

где  $m_{uik}$  – нормированное численное значение  $k$ -й компоненты  $u$ -го вектора признаков  $i$ -го фрейма,

$x1_{uk}$  – минимальное значение  $k$ -й

компоненты  $u$ -го вектора признаков,

$x2_{uk}$  – максимальное значение  $k$ -й

компоненты  $u$ -го вектора признаков,

$N_u$  – длина  $u$ -го вектора признаков,

$N_u^{(1)}$  – количество эталонов для  $u$ -го вектора признаков.

Вектор параметров модели  $u$ -го классификатора дикторов образован набором эталонов фреймов

$$\mathbf{m}_u = (\mathbf{m}_{u1}, \dots, \mathbf{m}_{uN_u^{(1)}}), \quad u \in \overline{1,14}.$$

На основе сформированных векторов параметров создаются модели классификаторов дикторов.

## 3. Формирование модели классификатора дикторов

На основании сформированных векторов параметров в качестве модели классификатора дикторов выбрана модель авторской модифицированной вероятностной нейросети (MPNN), представленная на рис. 1.

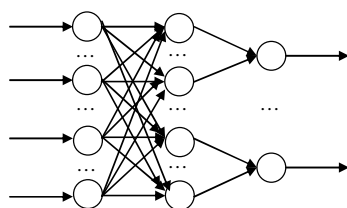


Рисунок 1 – Модифицированная вероятностная нейросеть

Предложенная нейросеть формируется по следующему принципу. Компонентам вектора признаков речевого сигнала  $\mathbf{x}_u$  соответствуют нейроны входного слоя; эталонам (параметрам модели)  $\mathbf{m}_{ui}$  соответствуют нейроны скрытого слоя; номерам дикторов соответствуют нейроны выходного слоя. Компоненты эталонов  $m_{uik}$  могут рассматриваться как весовые коэффициенты скрытого слоя.

Вместо модели вероятностной нейросети (PNN) вида [6, 7, 8]

$$f_j(\mathbf{x}) = \frac{1}{n_j} \sum_{i=1}^{N^{(1)}} w_{ij} G_i(\mathbf{x}), \quad j \in \overline{1, N^{(2)}},$$

$$G_i(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}_i^{-1}(\mathbf{x} - \mathbf{m}_i)\right)}{\sqrt{(2\pi)^N \det \mathbf{C}_i}},$$

или

$$G_i(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2} \frac{(\mathbf{x} - \mathbf{m}_i)^T (\mathbf{x} - \mathbf{m}_i)}{\sigma^2}\right)}{\sqrt{(2\pi)^N \sigma^2}},$$

для  $u$ -го классификатора дикторов используется модель MPNN вида

$$f_{uj}(\mathbf{x}_u) = \max_{i \in \overline{1, N_u^{(1)}}} (w_{uij} G_i(\mathbf{x}_u)), \quad j \in \overline{1, N^{(2)}}, \quad u \in \overline{1,14},$$

$$G_i(\mathbf{x}) = \exp\left(-\frac{1}{2}(\mathbf{x}_u - \mathbf{m}_{ui})^T (\mathbf{x}_u - \mathbf{m}_{ui})\right),$$

где  $\mathbf{x}_u$  – нормированный входной вектор признаков размерности  $N_u$ ,

$N_u$  – количество нейронов входного слоя,

$N_u^{(1)}$  – количество нейронов в первом (скрытом) слое,

$N^{(2)}$  – количество нейронов во втором (выходном) слое,

$w_{uij} \in \{1,0\}$  – весовые коэффициенты выходного слоя.

В отличие от обычной PNN для MPNN не требуется эмпирически подбирать ковариационную матрицу  $\mathbf{C}_i$  или среднеквадратичное отклонение  $\sigma$ , что упрощает и ускоряет процедуру обучения. Также считается, что все дикторы имеют одинаковую априорную вероятность и цену ошибки классификации.

Принятие решений  $u$ -м классификатором осуществляется в соответствии с решающим правилом

$$j^* = \arg \max_{j \in \overline{1, N^{(2)}}} f_{uj}(\mathbf{x}_u) \rightarrow result_u = j^*, \quad u \in \overline{1,14}$$

#### 4. Численное исследование

Для выбора лучших классификаторов из 14 предложенных и сопоставления авторской модифицированной вероятностной нейросети (MPNN) с вероятностной нейросетью (PNN) и алгоритмом динамического программирования DTW было проведено численное исследование. Длина тестовой выборки определялась как  $M=100$ .

Структура PNN и MPNN была определена следующим образом:

– количество нейронов в первом (скрытом) слое –  $N_u^{(1)} = 3000$ ;

– количество нейронов во втором (выходном) слое –  $N^{(2)} = 100$ .

Алгоритм DTW был выбран детерминированным и симметричным [9, 10].

Надежность классификации (идентификации) дикторов оценивалась по функционалу

$$J = \frac{m}{n},$$

где  $m$  – количество правильных классификаций,  
 $n$  – общее количество классификаций.

Результаты численного исследования идентификации диктора на основе MPNN, PNN и DTW по ключевому слову «Саша», фонеме |а| и фонеме |ш|, выделенным из этого слова, приведены в табл.1, табл.2 и табл.3 соответственно. Как видно из табл.1-3, наибольшую вероятность правильной идентификации по ключевому слову «Саша» и фонеме |а| имеет авторская сеть MPNN. Полученные результаты (табл. 1, 3) подтверждают, что для MPNN и DTW с

увеличением длины ключевого слова (от одной фонемы до четырех) возрастает вероятность распознавания (классификации) диктора. Как следует из табл.1, лучшими векторами признаков являются меры контрастности НВП (на основе вейвлета Морле), MFCC и коэффициенты отражения ЛП. Эти вектора разной физической природы – первый и второй учитывают особенности слухового аппарата, третий – речевого аппарата, поэтому предлагается их совместное использование. Таким образом, для MPNN рекомендуется использовать нейросетевые классификаторы 2, 12 и 14.

#### Выводы

*Новизна.* В статье предложена методика идентификации (классификации) диктора на основе модифицированной вероятностной нейросети (MPNN). Эта методика включает в себя формирование системы векторов признаков, векторов параметров моделей классификаторов дикторов, моделей классификаторов дикторов. Предложенные нейросетевые классификаторы диктора обладает следующими достоинствами: высокое быстродействие обучения и адаптации за счет отсутствия длительной настройки параметров модели; высокое быстродействие принятия решения за счет параллельной обработки информации; высокая точность и надежность принимаемых решений за счет выбора лучших векторов признаков.

*Практическое значение.* Основные положения данной работы предназначены для реализации в интеллектуальных системах поддержки принятия решений.

#### Литература

1. Бондаренко М.Ф. Ідентифікація людини за параметрами мовних сигналів (проблеми та шляхи їх вирішення) / М.Ф. Бондаренко, О.Я. Дрюченко, С.Ф. Коряк, Ю.П. Шабанов-Кушнарченко. – Х.: «Компанія СМІТ», 2006. – 260 с.
2. Рабинер Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер. – М.: Радио и связь, 1981. – 496 с.
3. Biatov K.A high speed unsupervised speaker retrieval using vector quantization and second-order statistics / K.A. Biatov // Праці десятої всеукраїнської міжнар. конф. «УкрОБРАЗ'2010». – К. – 2010. – С. 45-48.
4. Атал Б.С. Автоматическое опознавание дикторов по голосам / Б.С.Атал // ТИИЭР. – 1976. – Т. 64, №4. – С. 48-66.
5. Федоров Е.Е. Методика формирования акустических характеристик эталонов речи / Е.Е. Федоров. – Донецк: изд-во «Вебер», 2008. – 282 с.
6. Specht D.F. Probabilistic neural networks / D.F. Specht // Neural Networks. – 1990. – Vol. 3. – P. 109-118.
7. Каллан Р. Основные концепции нейронных сетей / Р. Каллан. – М.: Издательский дом «Вильямс», 2001. – 288 с.
8. Руденко О.Г. Штучні нейронні мережі: навчальний посібник / О.Г. Руденко, Є.В. Бодяньський. – Х.: ТОВ «Компанія СМІТ», 2006. – 404 с.
9. Sakoe H. Dynamic programming algorithm optimization for spoken word recognition / H. Sakoe, S. Chiba // IEEE Transactions on Acoustics, Speech and Signal Processing. – 1978. – Vol. 26, №1. – P. 43- 49.
10. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк. – К.: Наук. думка, 1987. – 261 с.

Таблица 1. Результаты численного исследования систем признаков, используемых при идентификации диктора посредством MPNN

Фонема / Слово	Система признаков	Вероятность идентификации
слово «Саша»	коэффициенты ЛП	0.66
	коэффициенты отражения ЛП	0.94
	автокорреляция ЛП	0.66
	кепстр ЛП	0.72
	площади поперечных сечений акустической трубы ЛП	0.72
	нормированная автокорреляция	0.54
	нормированный энергетический спектр ЛП	0.18
	меры контрастности ЛП	0.26
	нормированный энергетический спектр ДПФ	0.8
	меры контрастности ДПФ	0.64
	меры контрастности ДВП	0.56
	меры контрастности НВП	0.99
	нормированное количество импульсов равной длины	0.58
	MFCC	0.96
фонема  a	коэффициенты ЛП	0.3
	коэффициенты отражения ЛП	0.74
	автокорреляция ЛП	0.3
	кепстр ЛП	0.34
	площади поперечных сечений акустической трубы ЛП	0.7
	нормированная автокорреляция	0.28
	нормированный энергетический спектр ЛП	0.1
	меры контрастности ЛП	0.12
	нормированный энергетический спектр ДПФ	0.72
	меры контрастности ДПФ	0.48
	меры контрастности ДВП	0.36
	меры контрастности НВП	0.92
	нормированное количество импульсов равной длины	0.24
	MFCC	0.82
фонема  ш	коэффициенты ЛП	0.44
	коэффициенты отражения ЛП	0.8
	автокорреляция ЛП	0.44
	кепстр ЛП	0.62
	площади поперечных сечений акустической трубы ЛП	0.28
	нормированная автокорреляция	0.46
	нормированный энергетический спектр ЛП	0.1
	меры контрастности ЛП	0.12
	нормированный энергетический спектр ДПФ	0.36
	меры контрастности ДПФ	0.22
	меры контрастности ДВП	0.16
	меры контрастности НВП	0.72
	нормированное количество импульсов равной длины	0.24
	MFCC	0.66

Таблица 2. Результаты численного исследования систем признаков, используемых при идентификации диктора посредством PNN

Фонема / Слово	Система признаков	Вероятность идентификации
слово «Саша»	коэффициенты ЛП	0.06
	коэффициенты отражения ЛП	0.08
	автокорреляция ЛП	0.06
	кепстр ЛП	0.06
	площади поперечных сечений акустической трубы ЛП	0.02
	нормированная автокорреляция	0.08
	нормированный энергетический спектр ЛП	0.02
	меры контрастности ЛП	0.04
	нормированный энергетический спектр ДПФ	0.04
	меры контрастности ДПФ	0.1
	меры контрастности ДВП	0.02
	меры контрастности НВП	0.12
	нормированное количество импульсов равной длины	0.12
	MFCC	0.32
фонема  a	коэффициенты ЛП	0.28
	коэффициенты отражения ЛП	0.52
	автокорреляция ЛП	0.28
	кепстр ЛП	0.16
	площади поперечных сечений акустической трубы ЛП	0.2
	нормированная автокорреляция	0.16
	нормированный энергетический спектр ЛП	0.02
	меры контрастности ЛП	0.14
	нормированный энергетический спектр ДПФ	0.32
	меры контрастности ДПФ	0.24
	меры контрастности ДВП	0.14
	меры контрастности НВП	0.2
	нормированное количество импульсов равной длины	0.22
	MFCC	0.62
фонема  ш	коэффициенты ЛП	0.28
	коэффициенты отражения ЛП	0.54
	автокорреляция ЛП	0.28
	кепстр ЛП	0.5
	площади поперечных сечений акустической трубы ЛП	0.12
	нормированная автокорреляция	0.16
	нормированный энергетический спектр ЛП	0.12
	меры контрастности ЛП	0.22
	нормированный энергетический спектр ДПФ	0.1
	меры контрастности ДПФ	0.24
	меры контрастности ДВП	0.16
	меры контрастности НВП	0.3
	нормированное количество импульсов равной длины	0.22
	MFCC	0.72

Таблиця 3. Результати численного исследования систем признаков, используемых при идентификации диктора посредством DTW

Фонема / Слово	Система признаков	Вероятность идентификации
слово «Саша»	коэффициенты ЛП	0.86
	коэффициенты отражения ЛП	0.98
	автокорреляция ЛП	0.86
	кепстр ЛП	0.82
	площади поперечных сечений акустической трубы ЛП	0.7
	нормированная автокорреляция	0.76
	нормированный энергетический спектр ЛП	0.6
	меры контрастности ЛП	0.6
	нормированный энергетический спектр ДПФ	0.86
	меры контрастности ДПФ	0.94
	меры контрастности ДВП	0.78
	меры контрастности НВП	0.82
	нормированное количество импульсов равной длины	0.92
	MFCC	0.98
фонема  a	коэффициенты ЛП	0.36
	коэффициенты отражения ЛП	0.72
	автокорреляция ЛП	0.36
	кепстр ЛП	0.4
	площади поперечных сечений акустической трубы ЛП	0.52
	нормированная автокорреляция	0.54
	нормированный энергетический спектр ЛП	0.24
	меры контрастности ЛП	0.2
	нормированный энергетический спектр ДПФ	0.78
	меры контрастности ДПФ	0.66
	меры контрастности ДВП	0.54
	меры контрастности НВП	0.58
	нормированное количество импульсов равной длины	0.58
	MFCC	0.9
фонема  ш	коэффициенты ЛП	0.68
	коэффициенты отражения ЛП	0.86
	автокорреляция ЛП	0.68
	кепстр ЛП	0.58
	площади поперечных сечений акустической трубы ЛП	0.32
	нормированная автокорреляция	0.34
	нормированный энергетический спектр ЛП	0.3
	меры контрастности ЛП	0.22
	нормированный энергетический спектр ДПФ	0.22
	меры контрастности ДПФ	0.4
	меры контрастности ДВП	0.38
	меры контрастности НВП	0.6
	нормированное количество импульсов равной длины	0.46
	MFCC	0.86

Надійшла до редакції 12.11.2010