

УДК 004

**ФЕДОРОВ Е.Е., д.т.н., доцент,
Донецкая академия автомобильного транспорта**

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ СОНОРНЫХ СОГЛАСНЫХ ЗВУКОВ КОМАНД ВОДИТЕЛЯ ПОСРЕДСТВОМ НЕРАСШИРЯЮЩИХ РАВНОМЕРНО НЕПРЕРЫВНЫХ ОТОБРАЖЕНИЙ НА ОСНОВЕ MFCC

В статье проводится численное исследование сонорных согласных звуков команд водителя посредством нерасширяющих равномерно непрерывных отображений, основанных на мел-частотных кепстральных коэффициентах для формирования вектора вещественных признаков и преобразовании вектора вещественных признаков к вектору целых признаков посредством нормирования, масштабирования и округления.

Ключевые слова: *распознавание команд водителя, гласные звуки, нерасширяющие равномерно непрерывные отображения, мел-частотные кепстральные коэффициенты, формирование и преобразование вектора признаков*

Постановка проблемы

В современной отечественной и мировой практике активно ведутся разработки интеллектуальных систем, связанных с автомобилем и его водителем и позволяющих снизить количество аварий. Особую важность эти исследования имеют для вербального управления автомобилем. Для эффективной идентификации вербальных команд водителя принятие решений должно производиться с высоким быстродействием и высокой вероятностью. Параметры системы идентификации команд водителя определяются по результатам численного исследования.

Анализ последних исследований и публикаций

Существующие методы и модели распознавания речевых образов обычно основаны на скрытых марковских моделях (СММ), алгоритме динамического программирования DTW, и нейросетях и обладают одним или несколькими из перечисленных ниже недостатков [1-6]:

- длительность обучения;
- хранения большого количества эталонов звуков или слов, а также весовых коэффициентов;
- длительность распознавания;
- неудовлетворительная вероятность распознавания.
- необходимость большого количества обучающих данных.

С другой стороны в литературе обычно не проводятся исследования, связанные с определением множеств векторов значений признаков, относящихся к разным звукам речи. Это связано с тем, что обычно значения признаков являются непрерывными, хотя и ограниченными сверху и снизу. Таким образом, множества векторов признаков разных звуков являются несчетными.

Цель статьи

Для вербального управления транспортными объектами провести численное исследование сонорных согласных речевых команд водителя посредством нерасширяющих равномерно непрерывных отображений, основанных на мел-частотных кепстральных коэффициентах (MFCC).

Структура нерасширяющих равномерно непрерывных отображений

В работе [7] были впервые введены нерасширяющие равномерно непрерывные отображения, действующие в компактных метрических пространствах образцов сигналов. В данной статье эти отображения используются с точки зрения преобразования и классификации образцов

звуков речи. Структура нерасширяющих равномерно непрерывных отображений определена в работе [8] следующим образом:

– нерасширяющее равномерно непрерывное отображение $\varphi: S^N \rightarrow V^K$ соответствует функции формирования вектора признаков, т.е. отображает вектор целых значений дискретного сигнала $\mathbf{s} = (s_1, \dots, s_N)$, в вектор вещественных признаков $\mathbf{v} = (v_1, \dots, v_K)$, причем каждая компонента вектора \mathbf{v} вычисляется как MFCC [8];

– нерасширяющее равномерно непрерывное отображение $\psi: V^K \rightarrow X^K$ соответствует функции нормирования (значения признаков текущего вектора \mathbf{v} делятся на максимальное значение этого вектора), масштабирования и округления вектора вещественных признаков, т.е. отображает вектор вещественных признаков $\mathbf{v} = (v_1, \dots, v_K)$, в вектор целых признаков

$$\mathbf{x} = (x_1, \dots, x_K) \text{ в виде } x_i = \psi_i(\mathbf{v}) = \begin{cases} \left[\frac{v_i - \min_{k \in \{1, \dots, K\}} v_k}{\max_{k \in \{1, \dots, K\}} v_k - \min_{k \in \{1, \dots, K\}} v_k} \cdot \alpha \right], & \max_{k \in \{1, \dots, K\}} v_k > \min_{k \in \{1, \dots, K\}} v_k, i \in \overline{1, K}, \\ 0. & \max_{k \in \{1, \dots, K\}} v_k = \min_{k \in \{1, \dots, K\}} v_k \end{cases}$$

где $\alpha \in \{2^{r_x-1}, \dots, 2^{r_x}\}$ – коэффициент масштаба, $[\]$ означает округление;

r_x – количество разрядов (бит) для одного значения признака.

– нерасширяющее равномерно непрерывное отображение $\varphi: X^K \rightarrow Y^M$ соответствует функции классификации вектора целых признаков, т.е. отображает вектор целых признаков $\mathbf{x} = (x_1, \dots, x_K)$ в номер класса образца звука речи, представленный булевым вектором с одной ненулевой компонентой $\mathbf{y} = (y_1, \dots, y_M)$, причем классификация (т.е. сопоставление с обучающим образцом) проводится на основе дискретной метрики [8].

Формирование областей сонорных согласных звуков в признаковом пространстве

Области сонорных согласных звуков речи в признаковом пространстве X^K формируются на основе обучающего множества образцов соответствующих звуков. В статье исследуются сонорные согласные звуки |м|, |н|, |л|, |р|, |й|, не содержащие переходных участков. Параметры для нерасширяющих равномерно непрерывных отображений определены в работе [7] следующим образом: $K = 13, r_x = 4, \alpha = 10$. Структура нерасширяющих равномерно непрерывных отображений определена в работе [8].

В табл.1 и на рис.1 приведены обучающие образцы звука |м| слова «методический» после выполнения нормирования, масштабирования и округления. Каждый образец представляет собой вектор признаков, вычисленный на основе логарифмированной меры контрастности. Область звука |м| в признаковом пространстве представлена в виде

$$A_m = \{ \mathbf{x} \mid x_1 \in \{1,2\}, x_2 \in \{7,8,9,10\}, x_3 \in \{8,9,10\}, x_4 \in \{3,4,5\}, x_5 \in \{9,8,10\}, x_6 \in \{1,2\}, x_7 \in \{4,5,6,7\}, x_8 \in \{0,1,2,3\}, x_9 \in \{5,6,7\}, x_{10} \in \{0,1\}, x_{11} \in \{4,5,6\}, x_{12} \in \{1,2,3\}, x_{13} \in \{4,5,6,7,8\} \}$$

Обучающие образцы звука |м| слова «методический»

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1	1	7	10	4	8	1	5	1	5	0	4	2	4
2	1	7	10	4	8	1	6	2	6	0	5	2	5
3	1	7	10	4	10	2	5	1	6	0	4	1	5
4	1	8	10	3	8	1	7	1	5	0	6	3	6
5	1	8	10	4	9	2	6	2	6	0	5	3	5
6	1	9	10	3	10	2	6	0	5	1	5	3	6
7	1	10	10	3	10	2	7	2	7	0	5	3	8
8	2	9	10	4	10	2	6	2	6	0	5	3	5
9	2	9	10	5	10	3	7	3	6	0	5	3	6
10	2	9	10	5	10	2	7	2	6	0	6	3	7
11	2	10	8	3	10	2	4	1	6	0	4	2	6
12	2	10	9	4	10	1	6	3	6	0	5	3	6
13	2	10	10	5	10	2	7	3	7	0	6	3	6

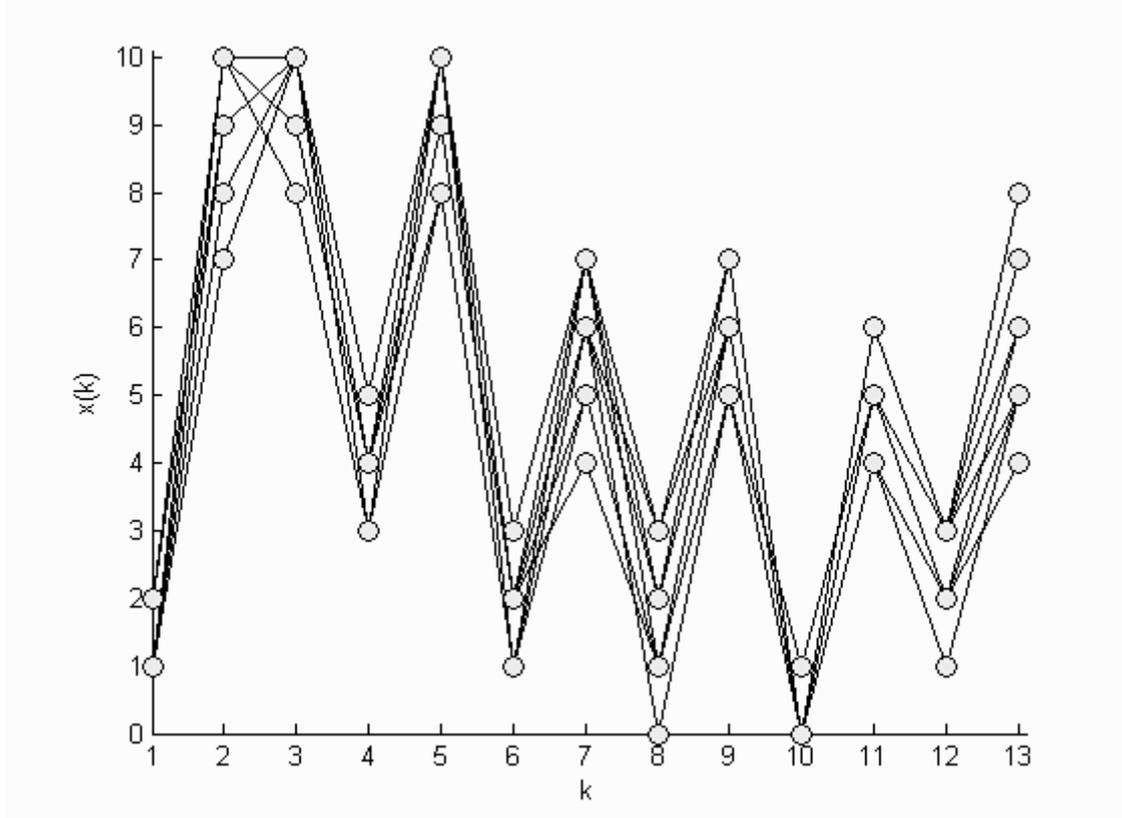


Рис. 1. Обучающие образцы звука |м| слова «методический»

В табл.2 и на рис.2 приведены обучающие образцы звука |н| слова «ножницы» после выполнения нормирования, масштабирования и округления. Каждый образец представляет собой вектор признаков, вычисленный на основе логарифмированной меры контрастности. Область звука |н| в признаковом пространстве представлена в виде

$$A_n = \{ \mathbf{x} \mid x_1 \in \{0,1,2\}, x_2 \in \{4,6,7,8,9,10\}, x_3 \in \{10\}, x_4 \in \{1,2,3,4,6\}, x_5 \in \{7,8,9,10\}, x_6 \in \{0,1,2,3\}, x_7 \in \{6,7,8\}, x_8 \in \{0,1,2\}, x_9 \in \{4,5,6\}, x_{10} \in \{0,1,2\}, x_{11} \in \{5,6,7\}, x_{12} \in \{0,1,2,3\}, x_{13} \in \{4,5,6\} \}$$

Обучающие образцы звука |н| слова «ножницы»

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1	0	4	10	4	8	2	6	1	6	1	5	1	4
2	1	6	10	6	10	0	6	1	5	1	6	2	5
3	1	7	10	4	8	1	6	1	4	0	5	1	4
4	1	8	10	4	9	1	7	2	6	0	6	2	5
5	1	8	10	4	8	0	6	1	5	1	6	1	4
6	1	8	10	4	7	0	7	1	5	1	5	1	5
7	1	9	10	3	9	2	7	1	6	0	6	1	5
8	1	9	10	3	8	1	7	1	5	1	5	0	4
9	1	9	10	3	8	2	7	1	4	0	6	2	5
10	1	9	10	4	10	2	7	1	4	0	7	2	5
11	1	9	10	4	9	2	7	1	6	0	6	2	5
12	1	10	10	2	7	1	8	0	6	1	6	2	5
13	1	10	10	1	8	0	7	1	6	1	6	1	4
14	1	10	10	3	10	0	7	2	6	1	7	2	5
15	1	10	10	3	8	0	8	1	6	2	6	2	5
16	1	10	10	4	9	2	8	1	5	0	6	3	6
17	2	9	10	4	9	3	8	0	5	1	6	2	6
18	2	10	10	3	9	2	7	0	6	1	6	1	5

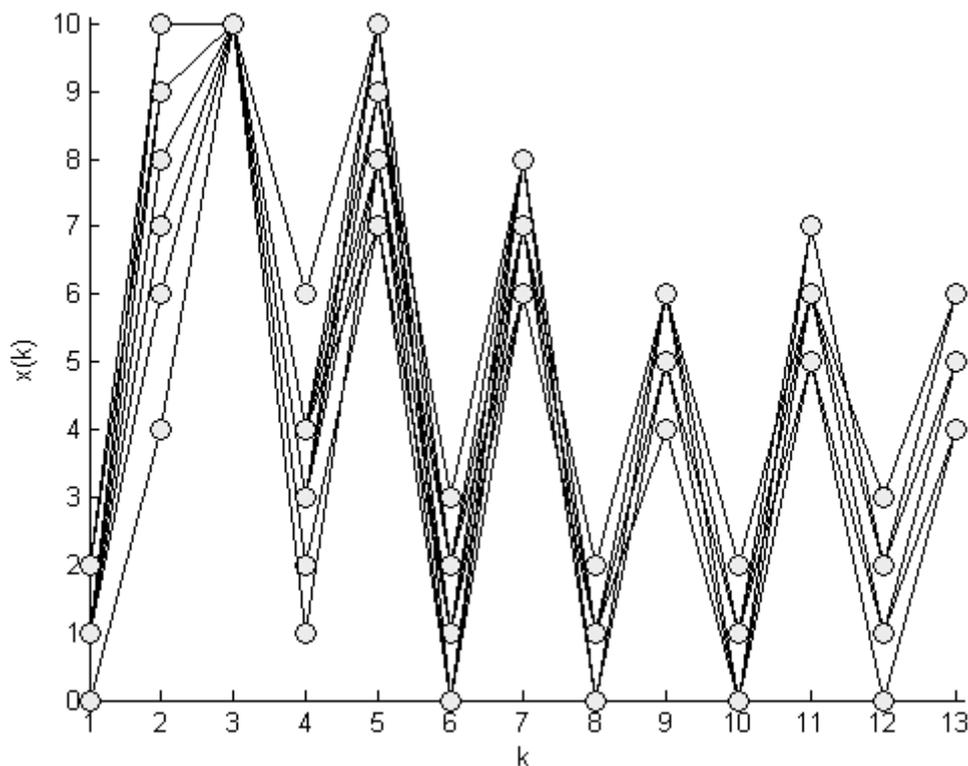


Рис. 2. Обучающие образцы звука |н| слова «ножницы»

В табл.3 и на рис.3 приведены обучающие образцы звука |л| слова «включить» после выполнения нормирования, масштабирования и округления. Каждый образец представляет собой вектор признаков, вычисленный на основе логарифмированной меры контрастности. Область звука |л| в признаковом пространстве представлена в виде

$$A_l = \{ \mathbf{x} \mid x_1 \in \{2,4\}, x_2 \in \{7,8,10\}, x_3 \in \{6,8,9,10\}, x_4 \in \{4,5,6,7,8\}, x_5 \in \{8,9,10\}, x_6 \in \{0,1,2,3,4\}, x_7 \in \{6,7,8\}, x_8 \in \{0,1\}, x_9 \in \{6,7,8\}, x_{10} \in \{2,3,4,5\}, x_{11} \in \{3,4,5,6,7,8\}, x_{12} \in \{0,1,2,3,4,5\}, x_{13} \in \{4,6,7,8\} \}$$

Таблица 3

Обучающие образцы звука |л| слова «включить»

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1	2	7	6	5	10	2	6	1	6	2	4	0	4
2	2	8	6	7	10	0	7	1	7	3	3	1	6
3	2	8	6	8	10	0	7	1	6	2	3	1	6
4	4	10	8	4	8	3	8	0	6	3	6	3	6
5	4	10	8	6	9	4	8	0	7	5	7	4	7
6	4	10	9	5	10	1	7	0	7	2	5	2	7
7	4	10	9	5	8	3	7	0	6	4	6	3	6
8	4	10	9	6	10	0	7	1	7	3	6	2	8
9	4	10	9	6	9	4	8	0	7	5	7	5	7
10	4	10	10	4	9	2	7	0	8	5	8	4	8
11	4	10	10	7	9	3	8	0	8	5	6	3	7

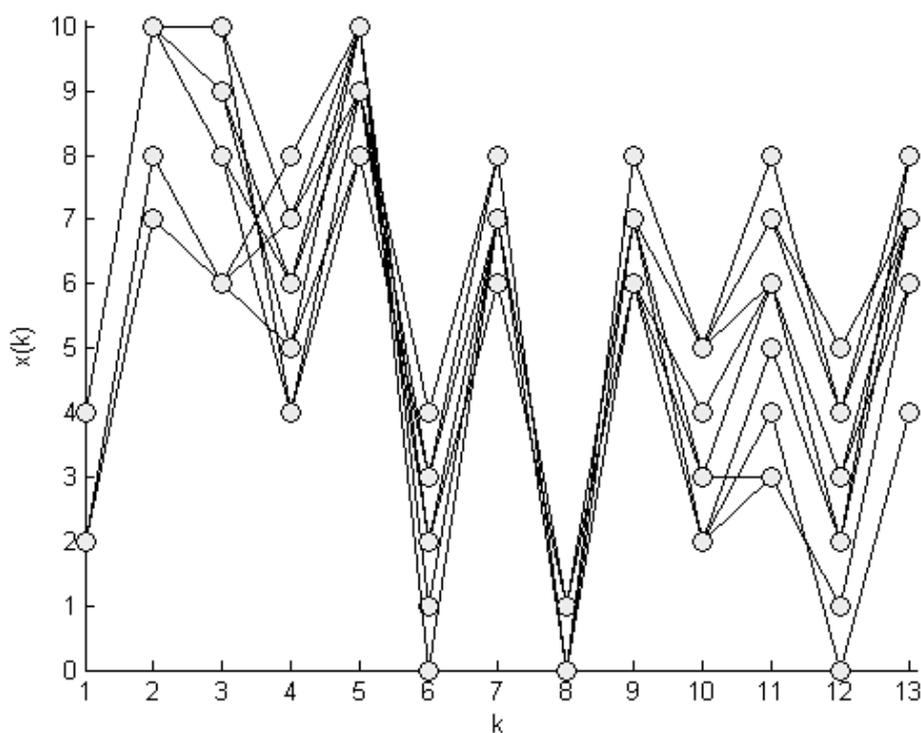


Рис. 3. Обучающие образцы звука |л| слова «включить»

В табл.4 и на рис.4 приведены обучающие образцы звука |р| слова «рез» после выполнения нормирования, масштабирования и округления. Каждый образец представляет собой вектор признаков, вычисленный на основе логарифмированной меры контрастности. Область звука |л| в признаковом пространстве представлена в виде

$$A_p = \{ \mathbf{x} \mid x_1 \in \{2,3,4\}, x_2 \in \{10\}, x_3 \in \{6,7,8,9\}, x_4 \in \{0,1,2\}, x_5 \in \{5,6,7,8\}, x_6 \in \{0,1,2\}, x_7 \in \{3,4,5,6\}, x_8 \in \{0,1,2\}, x_9 \in \{4,5,6\}, x_{10} \in \{2,3,4\}, x_{11} \in \{6,7,8\}, x_{12} \in \{2,3,4\}, x_{13} \in \{4,5,6,7\} \}$$

Обучающие образцы звука [p] слова «рез»

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1	2	10	7	1	7	0	3	1	5	2	6	2	4
2	2	10	7	0	6	1	5	1	4	2	6	3	5
3	2	10	7	0	6	1	5	1	5	2	7	4	6
4	2	10	7	1	6	0	5	2	5	3	7	3	5
5	2	10	8	1	7	0	4	1	5	2	6	3	4
6	2	10	8	1	7	1	5	0	4	2	6	2	4
7	3	10	6	1	6	0	5	2	6	4	8	4	6
8	3	10	6	0	6	2	5	2	5	3	6	4	6
9	3	10	6	0	5	1	4	1	5	2	6	4	5
10	3	10	7	1	6	0	6	2	5	3	7	4	5
11	3	10	7	2	8	1	4	0	5	2	7	3	5
12	3	10	7	0	5	1	5	1	5	4	7	4	6
13	3	10	9	1	6	2	5	0	4	3	7	3	4
14	4	10	7	0	5	2	6	2	5	4	7	4	7

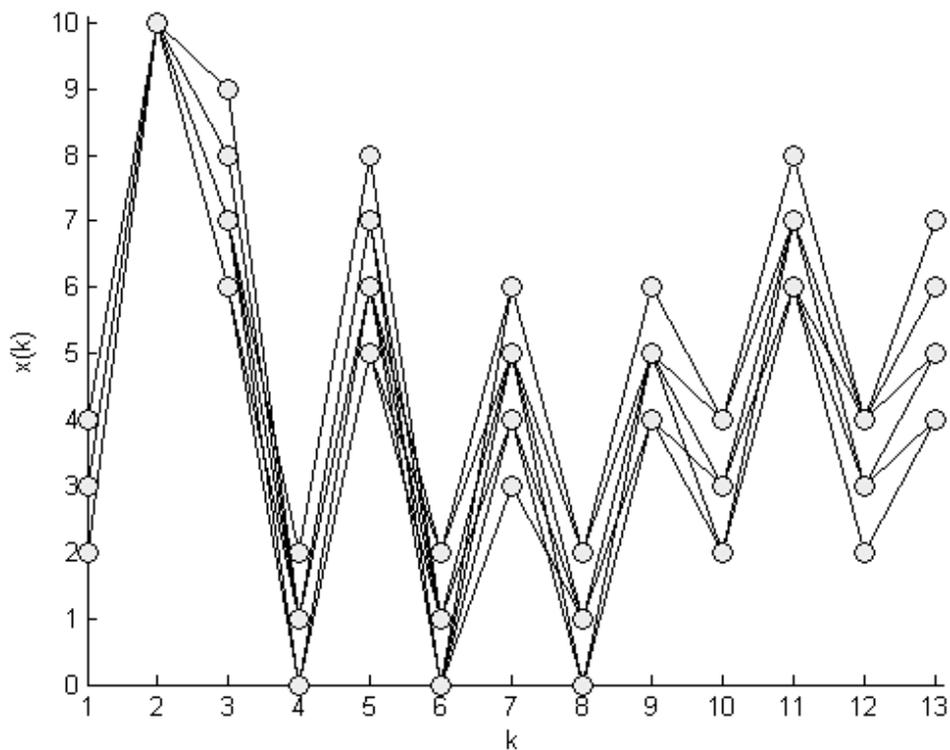


Рис. 4. Обучающие образцы звука [p] слова «рез»

В табл.5 и на рис.5 приведены обучающие образцы звука [й] слова «яков» после выполнения нормирования, масштабирования и округления. Каждый образец представляет собой вектор признаков, вычисленный на основе логарифмированной меры контрастности. Область звука [й] в признаковом пространстве представлена в виде

$$A_{\text{й}} = \{ \mathbf{x} \mid x_1 \in \{1,2,3,4\}, x_2 \in \{4,5,6,7,8,10\}, x_3 \in \{4,6,7,8,9\}, x_4 \in \{4,5,6,7,8,9,10\}, x_5 \in \{7,8,9,10\}, \\ x_6 \in \{0,1\}, x_7 \in \{7,8,10\}, x_8 \in \{0,1,2\}, x_9 \in \{5,6,7\}, x_{10} \in \{2,3,4,5\}, x_{11} \in \{4,5,6,7\}, x_{12} \in \{0,1,2,3\}, \\ x_{13} \in \{6,7,8,9\} \}$$

Обучающие образцы звука |й| слова «яков»

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1	1	4	8	9	8	0	10	1	5	3	5	1	6
2	1	4	9	7	7	1	10	0	6	3	4	1	6
3	2	5	8	7	7	1	10	0	7	5	6	2	8
4	2	6	7	8	8	0	10	1	7	4	6	2	8
5	2	8	8	7	8	0	10	1	7	4	6	1	8
6	2	8	9	7	9	0	10	1	6	4	7	2	8
7	2	10	7	6	9	0	10	1	6	2	5	1	7
8	2	10	8	5	9	1	7	1	5	2	6	0	6
9	3	6	8	10	8	1	10	0	7	4	4	3	8
10	3	6	8	9	9	0	10	2	6	5	6	3	7
11	3	6	9	10	9	1	10	0	6	4	6	3	8
12	3	10	6	7	9	0	10	2	6	2	6	3	8
13	3	10	7	9	10	0	10	1	6	3	6	2	8
14	3	10	7	6	9	0	8	1	5	2	6	2	8
15	3	10	9	6	8	0	8	2	6	2	5	2	7
16	4	7	7	10	9	1	10	0	7	5	4	3	9
17	4	10	4	4	9	0	7	1	5	2	5	2	7

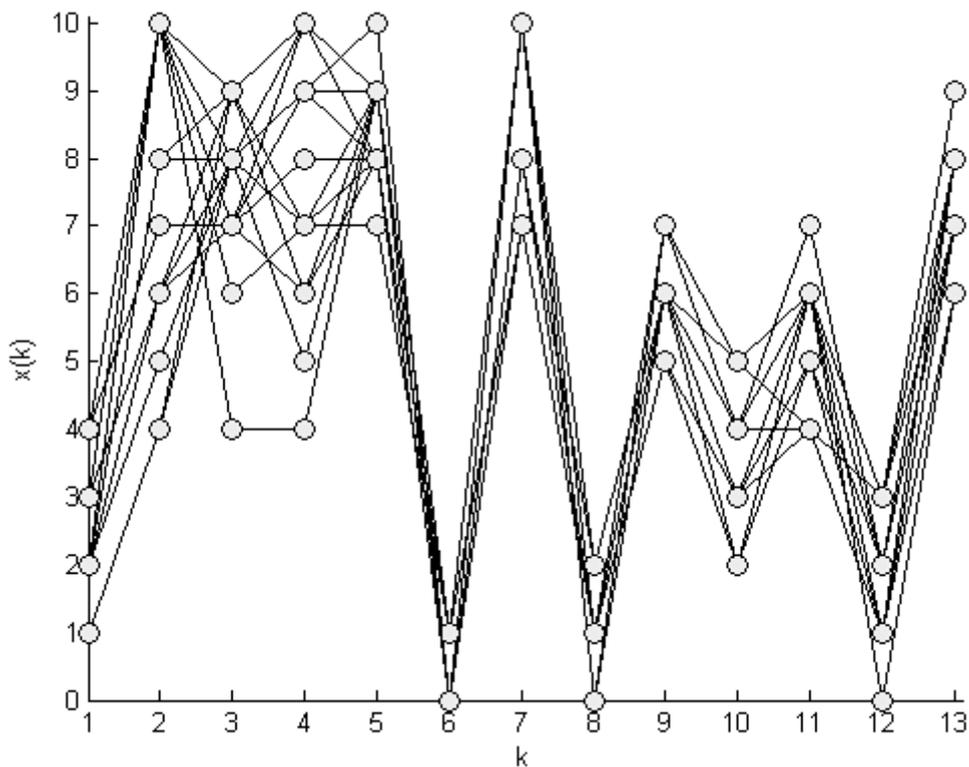


Рис. 5. Обучающие образцы звука |й| слова «яков»

Выводы

Научная новизна. В статье для вербального управления транспортными объектами проведено численное исследование сонорных согласных речевых команд водителя посредством нерасширяющих равномерно непрерывных отображений на основе MFCC. В результате исследования определены области каждого сонорного согласного звука в признаковом пространстве.

Практическое значение. Система идентификации речевых команд водителя, разработанная на основе созданных обучающих образцов, может использоваться в системах человеко-машинного общения для различных отраслей.

Список литературы

1. Осовский С. Нейронные сети для обработки информации / С. Осовский. – М.: Финансы и статистика, 2002. – 344 с.
2. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – М.: Издательский дом «Вильямс», 2006. – 1104 с.
3. Комарцова Л.Г. Нейрокомпьютеры / Л.Г. Комарцова, А.В. Максимов. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2002. – 320 с.
4. Каллан Р. Основные концепции нейронных сетей / Р. Каллан. – М.: Издательский дом «Вильямс», 2001. – 288 с.
5. Малла С. Вэйвлеты в обработке сигналов / С. Малла. – М.: Мир, 2005. – 671 с.
6. Федоров Е.Е. Методология создания мультиагентной системы речевого управления: монография / Е.Е. Федоров. – Донецк: изд-во «Ноулидж», 2011. – 356 с.
7. Федоров Е.Е. Метод обработки сигнала на основе нерасширяющих равномерно непрерывных отображений / Е.Е. Федоров // Труды Междунар. науч.-техн. конф. «Радіотехнічні поля, сигнали, апарати та системи (теорія, практика, історія, освіта)» (РТПСАС'2012). – Киев. – 2012. – С. 343-346.
8. Федоров Е.Е. Численное исследование гласных звуков команд водителя посредством нерасширяющих равномерно непрерывных отображений на основе MFCC / Е.Е. Федоров // Вісник Донецької академії автомобільного транспорту. – 2013. – №1. – С. 19-30.

Федоров Є.Є. Чисельне дослідження сонорних приголосних звуків команд водія за допомогою нерозширюючих рівномірно безперервних відображень на підставі MFCC

Анотація. У статті проводиться чисельне дослідження сонорних приголосних звуків команд водія за допомогою нерозширювальних рівномірно безперервних відображень, заснованих на мел-частотних кепстральних коефіцієнтах (MFCC) для формування вектора речовинних ознак і перетворенні вектора речовинних ознак до вектора цілих ознак за допомогою нормування, масштабування й округлення.

Ключові слова: розпізнавання команд водія, голосні звуки, нерозширюючі рівномірно безперервні відображення, мел-частотні кепстральні коефіцієнти, формування та перетворення вектора ознак

Fedorov E.E. Numerical research of resonant consonant sounds of commands of the driver through of not expanding uniformly continuous mapping based on MFCC

Abstract. In the article carried out numerical research of resonant consonant sounds of commands of the driver through not expanding uniformly continuous mapping, based on mel-frequency cepstral coefficients (MFCC) for formation of a vector of real features and transformation of a vector of real features to a vector of integer features by means of rationing, scaling and a rounding.

Keywords: recognition of driver's commands, vowels, non-expandable uniformly continuous maps, mel-frequency cepstral coefficients, formation and transformation of the feature vector

Стаття надійшла до редакції 12.05.2013 р.