

Е. Е. ФЕДОРОВ, канд. техн. наук, доц. ДААТ, Донецк

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ ТРУДНО РАЗДЕЛИМЫХ ВОКАЛЬНЫХ СОГЛАСНЫХ КОМАНД ОПЕРАТОРА НА ОСНОВЕ МЕЛ-ЧАСТОТНЫХ КЕПСТРАЛЬНЫХ КОЭФФИЦИЕНТОВ

У статті проводиться чисельне дослідження важке роздільних вокальних приголосних звуків команд оператора за допомогою нерозширювальних компактних рівномірно безперервних відображень, заснованих на мел-частотних кепстральних коефіцієнтах для формування вектора речовинних ознак і перетворенні вектора речовинних ознак до вектора цілих ознак за допомогою нормування, масштабування й округлення.

Ключові слова: важко роздільні звуки, обробка мовленнєвого сигналу, мел-частотні кепстральні коефіцієнти, ідентифікація вербальних команд, не розширювальні компактні рівномірно безперервні відображення.

В статье проводится численное исследование трудно разделимых вокальных согласных звуков команд оператора посредством нерасширяющих компактных 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: трудно разделимые звуки, обработка речевого сигнала, мел-частотные кепстральные коэффициенты, идентификация вербальных команд, нерасширяющие компактные равномерно непрерывные отображения.

Numerical research is carried out article difficultly separable vocal consonants of commands of the operator by means of not expanding compact uniformly continuous mapping based on mel-frequency cepstral factors for formation of a vector of real features and transformation of a vector of real features to a vector of the integer features by means of rationing, scaling and a rounding off.

Keywords: difficultly separable sounds, processing of a speech signal, swept-frequency cepstral factors, identification of the verbal commands, not expanding compact uniformly continuous mapping.

Введение. В современной отечественной и мировой практике активно ведутся разработки интеллектуальных систем, связанных с электромеханическими объектами, управление которыми осуществляет оператор-технолог, и позволяющих снизить количество аварий. Особую важность эти исследования имеют для вербального управления шахтными и металлургическими электромеханическими объектами. Для эффективной идентификации вербальных команд оператора принятие решений должно производиться с высоким быстродействием и высокой вероятностью.

Анализ последних исследований и публикаций. Существующие методы и модели распознавания речевых образов обычно базируются на двух следующих подходах [1-6].

Генеративный (структурный) подход использует детерминированный конечный автомат (например, подход КДП) или стохастический конечный автомат (например, скрытые марковские модели, модели смесей Гаусса).

Дескриптивный подход использует нейроподобные сети (например, многослойный перцептрон и радиально-базисная нейронная сеть).

Оба подхода обладают одним или несколькими из следующих недостатков: 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. Поэтому множества векторов признаков разных звуков являются несчетными.

Цель статьи. Для вербального управления электромеханическими объектами провести численное исследование трудно разделимых вокальных согласных звуков речевых команд оператора на основе мел-частотных кепстральных коэффициентов.

Структура нерасширяющих компактных равномерно непрерывных отображений. В работе [7] были впервые введены нерасширяющие компактные равномерно непрерывные отображения, действующие в компактных польских пространствах образцов сигналов. В данной статье эти отображения используются с точки зрения преобразования и классификации образцов звуков речи. Структура нерасширяющих компактных равномерно непрерывных отображений определена следующим образом.

Нерасширяющее компактное равномерно непрерывное отображение $\phi: S^N \rightarrow V^K$ соответствует функции формирования вектора признаков, т.е. отображает вектор целых значений дискретного сигнала $\mathbf{s} = (s_1, \dots, s_N)$, в вектор вещественных признаков $\mathbf{v} = (v_1, \dots, v_K)$ в виде

$$v_i = \phi_i(\mathbf{s}) = MFCC_{i-1}, i \in \overline{1, K},$$

$$MFCC_0 = \frac{1}{P} \sum_{i=1}^P E_i, MFCC_j = \sum_{i=1}^P E_i \cos(j(i-0.5)\pi/P), j \in \overline{1, K-1},$$

$$E_i = \lg \left(\sum_{k=k_1}^{k_2} \left(\sum_{m=0}^{N-1} (s_{m+1} - 2^{\tau_s}) w_1(m) e^{-j(2\pi/N)km} \right)^2 w_2(k) \right), i \in \overline{1, P},$$

$$w_1(m) = 0.54 + 0.46 \cos \frac{2\pi m}{N},$$

$$w2(k) = \begin{cases} 0, & k < k1_i \\ \frac{k - k1_i}{\Delta k_i / 2}, & k1_i \leq k \leq k1_i + \Delta k_i / 2 \\ \frac{k2_i - k}{\Delta k_i / 2}, & k1_i + \Delta k_i / 2 \leq k \leq k2_i \\ 0, & k > k2_i \end{cases},$$

где $MFCC_0$ – мел-частотные кепстральные коэффициенты (MFCC);

E – логарифмированная энергия i -й мел-частотной полосы Фурье-спектра;

$w1(k)$ – окно Хемминга;

$w2(k)$ – треугольное окно;

P – количество мел-частотных полос спектра;

r_s – количество разрядов (бит) для одного значения сигнала;

$k1_i, k2_i$ – границы частотных диапазонов i -й полосы, $\Delta k_i = k2_i - k1_i$.

Неразширяющее компактное 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 (в приведенной ниже формуле обозначается как $[\]$), т.е. отображает вектор вещественных признаков $\mathbf{v} = (v_1, \dots, v_K)$, в вектор целых признаков $\mathbf{x} = (x_1, \dots, x_K)$ в виде

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

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

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

Неразширяющее компактное 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 каждая i -я компонента вектора \mathbf{y} вычисляется в виде

$$y_i = \varphi_i(\mathbf{x}) = \begin{cases} 1, & \inf_{\tilde{\mathbf{x}} \in X_i^K} \rho(\mathbf{x}, \tilde{\mathbf{x}}) = 0 \\ 0, & \inf_{\tilde{\mathbf{x}} \in X_i^K} \rho(\mathbf{x}, \tilde{\mathbf{x}}) = 1 \end{cases}, i \in \overline{1, M},$$

$$\rho(\mathbf{x}, \tilde{\mathbf{x}}) = \begin{cases} 1, & d(\mathbf{x}, \tilde{\mathbf{x}}) \geq 0 \\ 0, & d(\mathbf{x}, \tilde{\mathbf{x}}) < 0 \end{cases},$$

$$d(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{k=1}^K |x_k - \tilde{x}_k|,$$

где X_i^K – класс эквивалентности, соответствующий i -му звуку речи.

Формирование областей трудно разделяемых вокальных согласных звуков в признаковом пространстве. Области трудно разделяемых вокальных согласных звуков речи в признаковом пространстве X^K формируются на основе обучающего множества образцов соответствующих звуков. На рис. 1-4 приведены слова «методический», «ножницы», «выключить», «гидравлический», в которых выделены непереходные (центральные) части звуков [м], [н], [в], [г] соответственно, используемые для формирования обучающих образцов. Параметры для нерасширяющих компактных равномерно непрерывных отображений определены следующим образом: $K = 13, r_x = 4, \alpha = 10$.

На рис. 5 приведены обучающие образцы звука [м] слова «методический» после выполнения нормирования, масштабирования и округления. Каждый образец представляет собой вектор признаков, вычисленный на основе MFCC. Как видно из рис. 5, область звука [м] в признаковом пространстве представлена в виде

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

На рис. 6 приведены обучающие образцы звука [н] слова «ножницы» после выполнения нормирования, масштабирования и округления. Каждый образец представляет собой вектор признаков, вычисленный на основе MFCC. Как видно из рис. 6, область звука [н] в признаковом пространстве представлена в виде

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

На рис. 7 приведены обучающие образцы звука [в] слова «выключить» после выполнения нормирования, масштабирования и округления. Каждый

образец представляет собой вектор признаков, вычисленный на основе MFCC. Как видно из рис. 7, область звука |в| в признаковом пространстве представлена в виде

$$A_6 = \{x \mid 0 \leq x_1 \leq 1, 2 \leq x_2 \leq 7, x_3 = 10, 2 \leq x_4 \leq 4, 5 \leq x_5 \leq 9, 0 \leq x_6 \leq 2, 5 \leq x_7 \leq 9, 1 \leq x_8 \leq 3, 5 \leq x_9 \leq 8, 0 \leq x_{10} \leq 2, 4 \leq x_{11} \leq 7, 0 \leq x_{12} \leq 2, 3 \leq x_{13} \leq 6\}.$$

На рис. 8 приведены обучающие образцы звука |г| слова «гидравлический» после выполнения нормирования, масштабирования и округления. Каждый образец представляет собой вектор признаков, вычисленный на основе MFCC. Как видно из рис. 8, область звука |г| в признаковом пространстве представлена в виде

$$A_2 = \{x \mid 0 \leq x_1 \leq 1, 1 \leq x_2 \leq 5, x_3 = 10, 1 \leq x_4 \leq 5, 7 \leq x_5 \leq 10, 0 \leq x_6 \leq 2, 6 \leq x_7 \leq 9, 0 \leq x_8 \leq 2, 5 \leq x_9 \leq 8, 0 \leq x_{10} \leq 3, 6 \leq x_{11} \leq 9, 1 \leq x_{12} \leq 4, 4 \leq x_{13} \leq 8\}.$$

Вычисление межклассового расстояния. Минимальное межклассовое расстояние между звуками вычисляется по формуле

$$\min_{\substack{i \in 1, |X_i^K| \\ j \in 1, |X_j^K|}} d(x_i, x_j) = \min_{\substack{i \in 1, |X_i^K| \\ j \in 1, |X_j^K|}} \sum_{k=1}^K |x_{ik} - x_{jk}|.$$

Минимальное межклассовое расстояние между звуками |м| и |н| равно 4. Минимальное межклассовое расстояние между звуками |в| и |г| равно 3. Таким образом, классы трудно разделимых звуков не пересекаются, поскольку минимальное межклассовое расстояние превышает 0 и, следовательно, нерасширяющие равномерно непрерывные отображения могут использоваться для таких классов звуков.

Выводы. В статье для вербального управления электромеханическими объектами проведено численное исследование трудно разделимых вокальных согласных звуков речевых команд оператора на основе мел-частотных кепстральных коэффициентов, которые были подвергнуты нормированию, масштабированию и округлению. В результате исследования было установлено, что классы трудно разделимых вокальных согласных звуков, сформированных посредством нерасширяющих равномерно непрерывных отображений, не пересекаются. Полученные результаты подтверждают адекватность выбранной системы признаков, а методы преобразования речевого сигнала, основанные на нерасширяющих компактных равномерно непрерывных отображениях, и определенные области звуков в признаковом пространстве могут использоваться при создании систем человеко-машинного общения для различных отраслей.

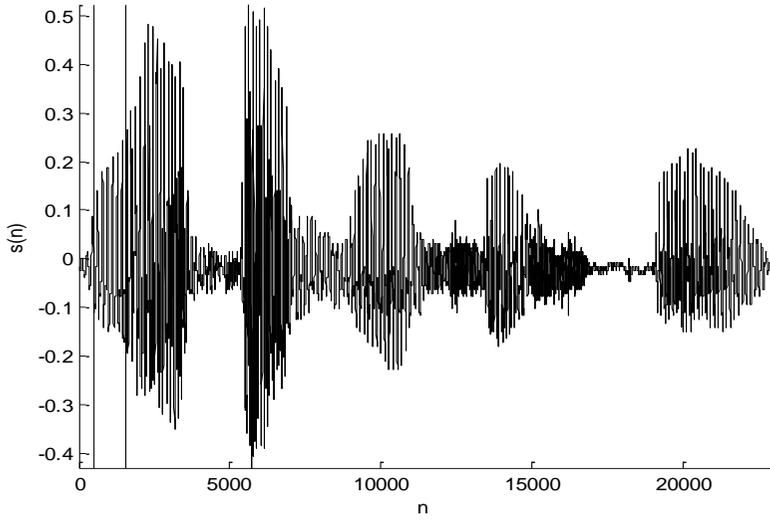


Рис. 1 – Звук $|m|$ слова «методический»

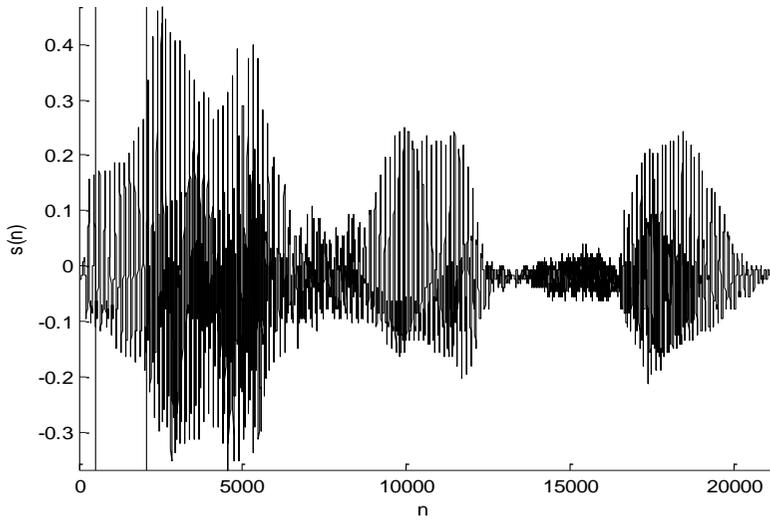


Рис. 2 – Звук $|h|$ слова «ножницы»

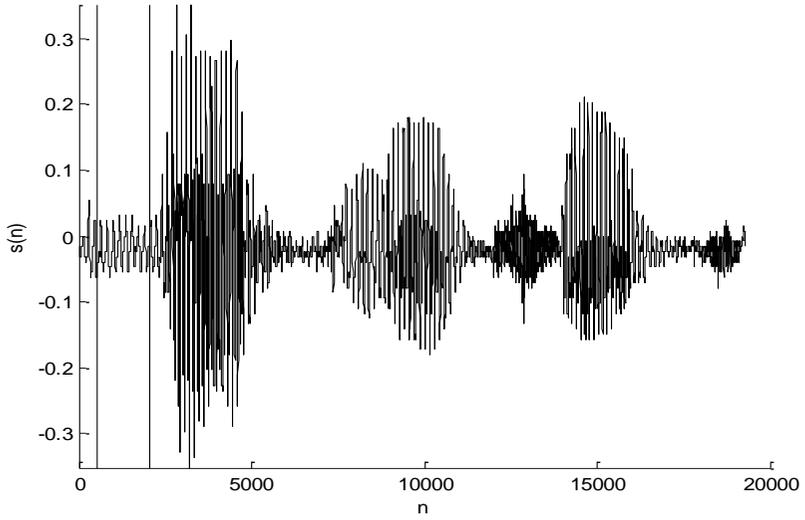


Рис. 3 – Звук [в] слова «выключить»

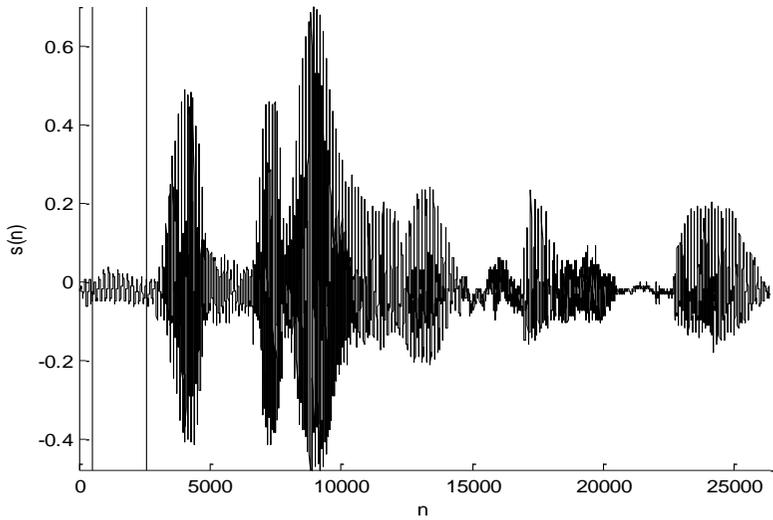


Рис. 4 – Звук [г] слова «гидравлический»

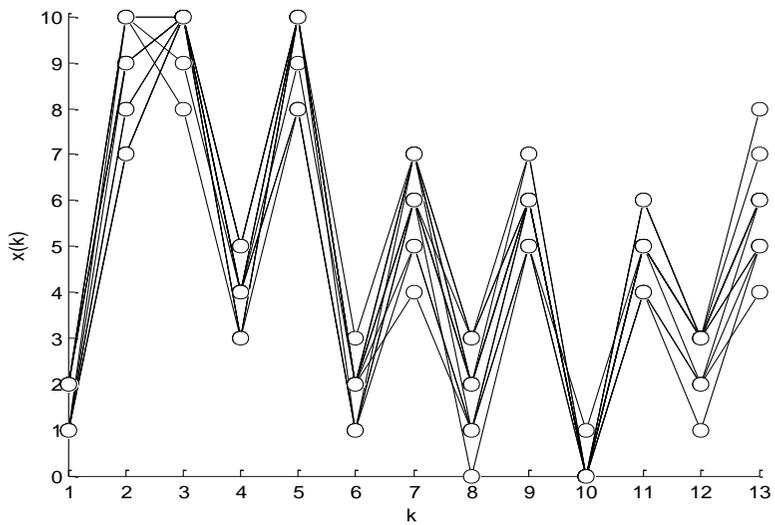


Рис. 5 – Обучающие образцы звука [м] слова «методический»

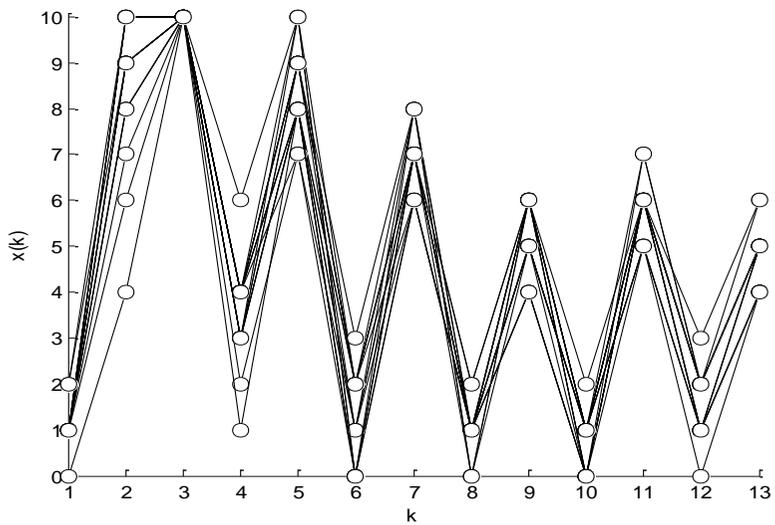


Рис. 6 – Обучающие образцы звука [н] слова «ножницы»

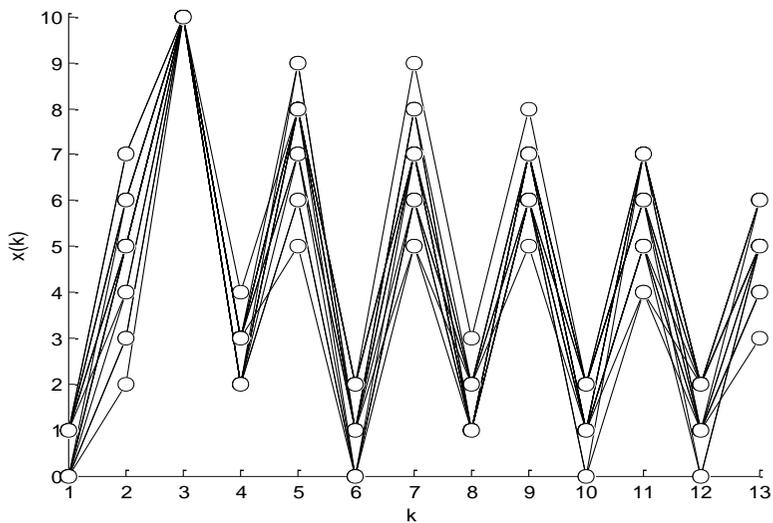


Рис. 7 – Обучающие образцы звука [в] слова «выключить»

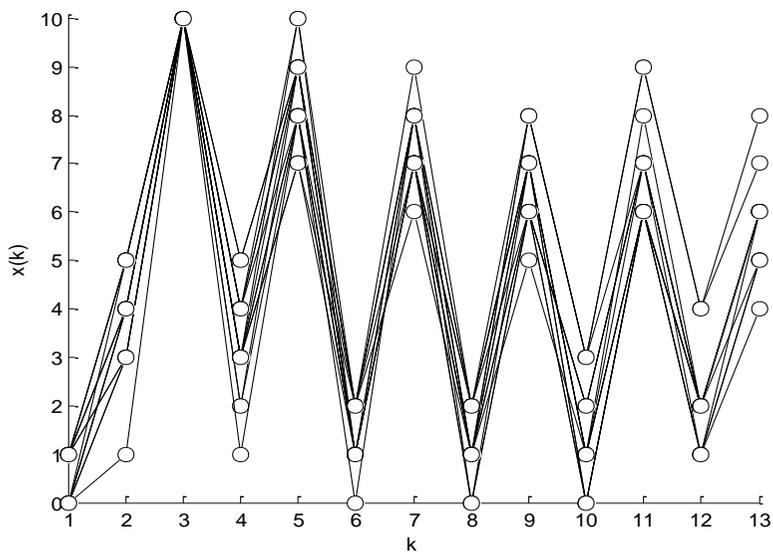


Рис. 8 – Обучающие образцы звука [r] слова «гидравлический»

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. : Вильямс, 2006. – 1104. 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 : Ноулидж, 2011. – 356 с. 7. *Федоров Е. Е.* Метод обработки сигнала на основе нерасширяющих равномерно непрерывных отображений // Труды междунар. науч.-техн. конф. «Радіотехнічні поля, сигнали, апарати та системи (теорія, практика, історія, освіта)». – К. : НТУ «КПІ», 2012. – С. 343–346.

Надійшла до редколегії 12.12.2012