

МЕТОДИ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ



Застосування дзвіноподібних функцій в алгоритмі сегментно-складового синтезу

О.О. САВЕНКОВА, О.М. КАРПОВ

Дніпропетровський національний університет

У статті розглянуто новий підхід до розв'язання задачі розпізнавання мовлення для великих словників на основі сегментно-складового синтезу траєкторії параметрів у класі дзвіноподібних функцій. Для побудови опису мовних одиниць словника запропоновано застосування функцій локон Аньєзі.

В статье рассмотрен новый подход к решению задачи распознавания речи для больших словарей на основе сегментно-слоговой синтеза траектории параметров в классе колоколообразных функций. Для построения описания речевых единиц словаря предложено использование функции локон Аньези.

New approach for the speech recognition problem solving for large vocabularies by segment-syllabic parameters trajectory synthesis in class of bell-shaped functions is considered. Using bell-shaped function of Agnesi for the description of vocabulary speech units is proposed.

Загальна постановка проблеми. Для створення людино-машинних інтерфейсів з мовленнєвим вводом/виводом інформації актуальною є проблема побудови системи розпізнавання мови (СРМ) зі словником максимально великого розміру [1, 2, 3, 4, 5]. Для створення таких СРМ домінуючим є підхід, в якому на першому етапі виконується розпізнавання звукових одиниць, які значно менше слова (фонем), а потім – із використанням різних джерел лінгвістичних знань (фонетичних, синтаксичних) визначаються можливі мовні структури повідомлень [2, 3, 4]. При такому підході в поданій реалізації мовного сигналу (МС) послідовність сегментів розглядається як сукупність незалежних подій, яким відповідають незалежні мовні одиниці (МО). Проте параметри сегментів МС знаходяться в залежності як от попередніх, так і от наступних сегментів, отже необхідно розглядати неперервні послідовності параметрів або траєкторії параметрів (ТП) МС, які враховують залежність між сегментами, та апроксимувати подану реалізацію МС в термінах МО як функцію часу. Цей процес представляє собою синтез великих МО (наприклад, речень) із менших МО, які належать до певного словника, найбільш відповідних для всього мовного повідомлення по всій сукупності МО словника. В процесі синтезу МО важливими є наступні проблеми: вибір МО словника; побудова адекватного опису параметрів МО словника; пошук такої комбінації мовних одиниць словника, яка є близькою за своїми параметрами з параметрами поданого для розпізнавання повідомлення. МО, які використовуються в процесі синтезу, повинні задовольняти вимозі максимальної повноти покриття множини слів даної мови. Цій вимозі задовольняють такі МО, як склади [6, 7, 8]. Отже доцільно накопичити словники складів і використовувати їх в процесі синтезу.

Для знаходження еталонної траєкторії параметрів (ЕТП), для якої досягнута найкраща близькість з

траєкторією параметрів поданого МС по всій сукупності МО, необхідно синтезувати можливі ЕТП, складені із ТП наявних у словнику МО, що вимагає величезних часових затрат. Розв'язання часової складності за допомогою використання стратегій пошуку в просторі станів у ширину, у глибину, із застосуванням евристик досліджено у [9, 10, 11, 12].

Основна проблема синтезу еталонної ТП з ТП мовних одиниць словника це розриви першого роду в точках склеювання суміжних ТП (рис. 1), тобто синтезована ТП не є гладкою функцією.



Рис. 1. Приклад склеювання ТП з розривом першого роду

Для досягнення найкращої близькості синтезованої траєкторії параметрів з траєкторією параметрів поданого МС постає задача побудови опису параметрів МО словника, який би забезпечив гладкість синтезованої траєкторії параметрів.

Постановка задачі. Згідно із загальною постановкою задачі сегментно-складового синтезу [6, 11, 12] сформулюємо наступну задачу.

Нехай є послідовність параметрів у вигляді спектрально-часового подання (або ТП) $XS(\omega, t)$ поданої реалізації мовного сигналу. Траєкторія параметрів $XS(\omega, t)$ сегментована на m_p сегментів-фонем SG_{pi}^x , які можна об'єднати в M груп-складів ($p = 1 \div M, i = 1 \div m_p$).

Нехай заданий словник $\{SL_k\}$, який містить N складів. Для кожного складу SL_k обчислені послідовності параметрів $\{YS_k(\omega, t)\}$ і знайдені межі сегментів-фоном SG_{kj}^y ($k=1 \div N, j=1 \div n_k$).

Необхідно побудувати опис МО словника $\{YS_k^*\}$ у класі аналітичних функцій, які є гладкими функціями на всій області визначення вихідного спектрально-часового подання $\{YS_k(\omega, t)\}$. При склеюванні ТП $\{YS_k^*\}$ необхідно отримати гладку функцію опису еталонної ТП $\tilde{XS}^*(\omega, t)$, для якої досягнута найкраща близькість з ТП $XS(\omega, t)$, тобто

$$d = \tilde{XS}^*(\omega, t) \# XS(\omega, t) \rightarrow \min, \quad (1)$$

де $\#$ – операція зіставлення.

Розглянемо побудову опису ТП МО словника для двох способів спектрально-часового подання [6, 13].

Побудова опису спектрально-смугових параметрів мовних одиниць словника за допомогою поліномів низьких порядків. Розглянемо спектрально-смугове подання ТП елементів словника $YS(\omega, t)$ [6, 13].

На кожній часовій ділянці, яка відповідає сегменту-фонемі, послідовності параметрів спектрально-смугового подання в кожній частотній смузі мають просту форму, яку можна описати поліномами низьких порядків (не вище третього порядку, $n \leq 3$). Таким чином, для кожної k -ї МО словника побудуємо наступну модель опису, яка з достатньою точністю апроксимує послідовності параметрів у кожній частотній смузі l : на кожному сегменті i знаходимо параметри моделі опису даного сегмента поліномом 3-го порядку (кубічна модель)

$$\tilde{Y}_{k,l}^i = a_{k,l}^i \cdot t^3 + b_{k,l}^i \cdot t^2 + c_{k,l}^i \cdot t + d_{k,l}^i, \quad (2)$$

для чого розв'язується задача мінімізації середньоквадратичного наближення з умовами в точках сегментації, які забезпечують гладкість склеювання ТП

$$\sigma_l^2 = \sum_{k=1}^S \left(\sum_{i=N_{k-1}}^{N_k} \left| \tilde{Y}_{k,l}^i - Y_{k,l}^i \right|^2 \right) \rightarrow \min, \quad (3)$$

де N_k – межі сегментації, S – кількість сегментів для k -ї МО словника.

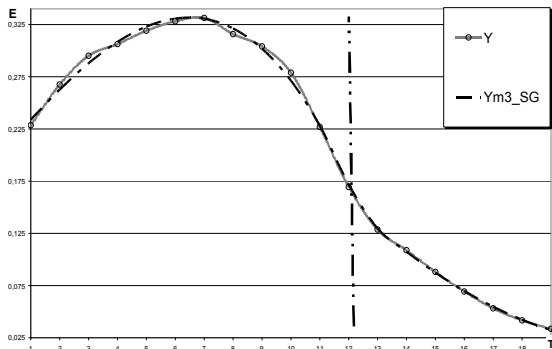


Рис. 2. Приклад побудови моделі опису послідовності параметрів в одній частотній смузі

На рис. 2 наведений результат побудови моделі опису ТП для двохсегментного складу словника «ін» в одній із частотних смуг. Застосовані наступні позначення: Y – вихідна сегментована послідовність параметрів в частотній смузі; $Ym3_SG$ – послідовність параметрів, побудована на основі кубічної моделі опису з вузлом в точці сегментації.

Алгоритм синтезу ЕТП на основі кубічної моделі. Синтез еталонної ТП $\tilde{XS}^*(\omega, t)$ виконується згідно наступної моделі сплайн-опису

$$\tilde{XS}^* = \begin{cases} \tilde{Y}_i^1, & N'_0 \leq i \leq N'_1, \\ \dots \\ \tilde{Y}_i^k, & N'_{k-1} + 1 \leq i \leq N'_k, \\ \dots \\ \tilde{Y}_i^M, & N'_{R-1} + 1 \leq i \leq N'_R, \end{cases} \quad (4)$$

де $k=1 \div R$; R – кількість складів в еталонній ТП $XS^*(\omega, t)$; \tilde{Y}_i^k – ТП мовної одиниці словника; N_k – кількість часових відліків ТП k -ї МО словника, $t_1 \in [1, N_1]$, $t_2 \in [1, N_2]$, ..., $t_k \in [1, N_k]$, ..., $t_R \in [1, N_R]$. Межі складів всередині поточної комбінації МО для еталонної ТП визначаються таким чином:

$$N'_0 = 1; \quad N'_1 = N_1; \quad N'_2 = N_1 + N_2; \quad N'_3 = N_1 + N_2 + N_3; \dots; \quad N'_R = N_1 + N_2 + \dots + N_R.$$

Параметри моделей опису МО уточнюються за умови мінімізації похибки для поточної комбінації МО в кожній частотній смузі l :

$$\sigma_l^2 = \sum_{k=1}^R \left(\sum_{i=N'_{k-1}}^{N'_k} \left| \tilde{Y}_{k,l}^i - Y_{k,l}^i \right|^2 \right) \rightarrow \min. \quad (5)$$

Побудова опису спектрально-часових параметрів мовних одиниць словника у класі дзвіноподібних функцій. Розглянемо спектрально-часове подання (СЧП) ТП елементів словника $YS(\omega, t)$ [6, 13].

Необхідно побудувати опис спектрально-часових параметрів МО словника $YS^*(\omega, t)$ у класі дзвіноподібних функцій. Розглянемо функцію локон Ан'єзі, яка має наступні властивості [14]:

(а) функція має максимум на заданій координаті (частота, час) та асимптотично наближається до площини області визначення в будь-якому напрямку від максимуму;

(б) форма функції визначається деякими параметрами a , b , c і має вигляд

$$Z(x) = \frac{a^3}{c^2 + (x-b)^2}; \quad (6)$$

(в) функція є гладкою на всьому діапазоні визначення.

Нехай в деякій частотно-часовій області $[\omega_0, \omega_M] \times [t_0, t_N]$ таблично задана спектрально-часова функція мовного сигналу $YS(\omega_k, t_l)$, де ω_k – дискретно задана частота, t_l – дискретно заданий час ($k=1..M, l=1..N$).

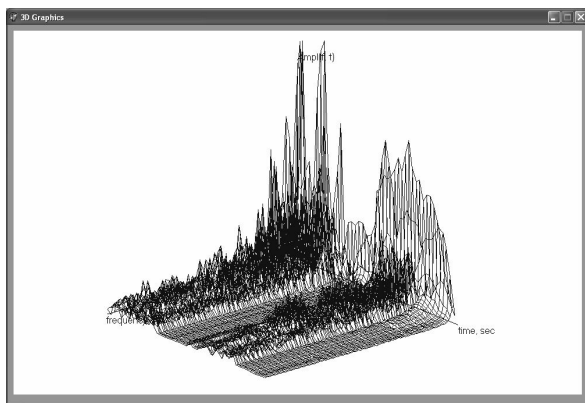
Необхідно для $YS(\omega, t)$ знайти параметри функцій $\{Zt(t_l)\}$, $\{Z\omega(\omega_k)\}$, які мають вище перераховані властивості (а), (б), (в), причому:

– функція $Zt_{(i)}(t_l)$ ($i=1..L$) описує часові властивості компонент мовного сигналу, визначена в діапазоні $Zt \in [0, t_N]$ і має вигляд

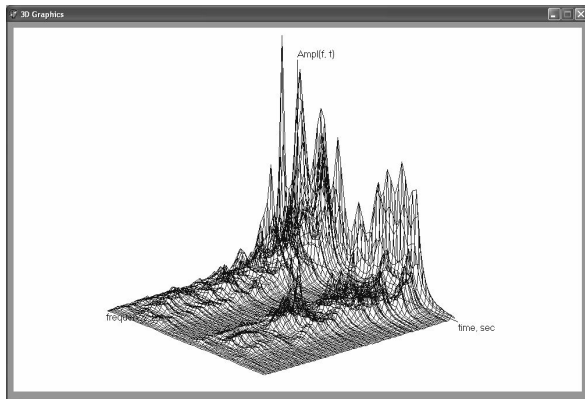
$$Zt_{(i)}(t_l) = \frac{a_{(i)}^3}{c_{(i)}^2 + (t_l - T_{(i)})^2}; \quad (7)$$

– функція $Z\omega_{(i)}(\omega_k)$ ($i=1..L$) описує частотні властивості компонент мовного сигналу, визначена в діапазоні $Z\omega \in [0, \omega_M]$ і має вигляд

$$Z\omega_{(i)}(\omega_k) = \frac{b_{(i)}^3}{d_{(i)}^2 + (\omega_k - \Omega_{(i)})^2}. \quad (8)$$



а)



б)

Рис. 3. а) вихідне СЧП слова "odin", б) опис СЧП у класі дзвіноподібних функцій

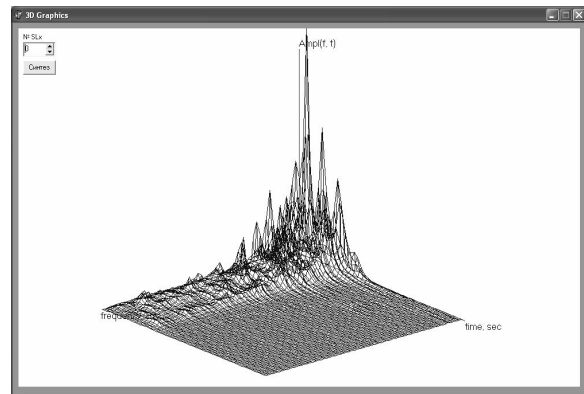
Тоді можна побудувати опис СЧП мовного сигналу $YS^*(\omega_k, t_l)$ у вигляді суперпозиції L добутків дзвіноподібних функцій $Zt_{(i)}(t_l)$, $Z\omega_{(i)}(\omega_k)$ ($k=1..M$, $l=1..N$, $i=1..L$) таким чином

$$YS^*(\omega_k, t_l) = \sum_{i=1}^L Z\omega_{(i)}(\omega_k) \cdot Zt_{(i)}(t_l), \quad (9)$$

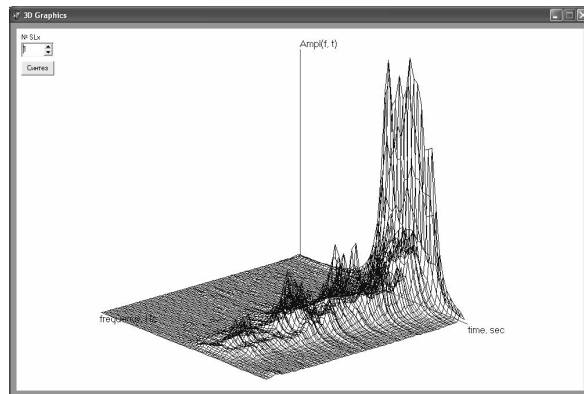
$$YS^*(\omega_k, t_l) = \sum_{i=1}^L \frac{a_{(i)}^3}{c_{(i)}^2 + (t_l - T_{(i)})^2} \cdot \frac{b_{(i)}^3}{d_{(i)}^2 + (\omega_k - \Omega_{(i)})^2}. \quad (10)$$

Визначення невідомих параметрів дзвіноподібних функцій $Zt_{(i)}(t_l)$, $Z\omega_{(i)}(\omega_k)$ ($i=1..L$) виконується за алгоритмом, який запропонований в [14]. Діапазоном визначення добутків функцій $\{Zt(t_l)\}$, $\{Z\omega(\omega_k)\}$ є область D розміром $[0, t_N] \times [0, \omega_M]$.

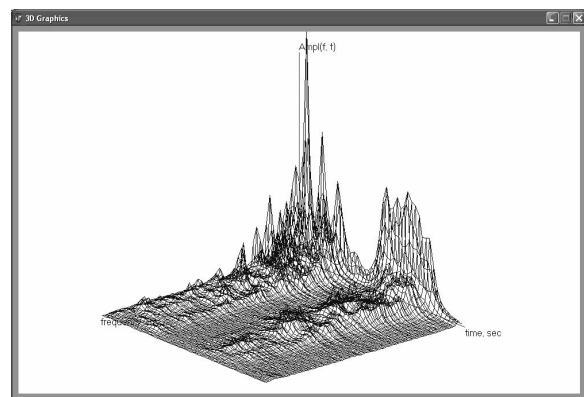
Для прикладу наведемо на рис. 3 а), б) СЧП слова "odin" та опис СЧП слова "odin" у класі дзвіноподібних функцій.



а)



б)



в)

Рис. 4. Опис СЧП у класі дзвіноподібних функцій для: а) складу "о"; б) складу "дин"; в) слова "odin"

Алгоритм синтезу ЕТП у класі дзвіноподібних функцій. Нехай кожна послідовність спектрально-часових параметрів j -ї МО словника визначена в прямокутній області $D^j: [\omega_0, \omega_M] \times [t_0, t_{N_j}]$ і подана у вигляді параметрів дзвіноподібних функцій $Z_{t(i)}(t_l)$, $Z_{\omega(i)}(\omega_k)$ ($i=1..L_j$, $k=1..M$, $l=1..N_j$). Нехай одним із алгоритмів пошуку знайдена комбінація з R ТП МО словника для еталонної ТП $X_S^*(\omega, t)$. Загальна область визначення для еталонної ТП з урахуванням діапазонів визначення кожної МО комбінації $D: [\omega_0, \omega_M] \times [t_0, t_N]$, де $t_N = t_{N_1} + t_{N_2} + \dots + t_{N_R}$.

Тоді опис СЧП для еталонної ТП $\tilde{X}_S^*(\omega, t)$ можна подати у вигляді суперпозиції R гладких функцій $Y_{S_m}^*(\omega, t)$ ($m=1..R$), які є аналітичним описом спектрально-часового подання МО словника, таким чином

$$\tilde{X}_S^*(\omega_k, t_l) = \sum_{m=1}^R Y_{S_m}^*(\omega_k, t_l),$$

$$\tilde{X}_S^*(\omega_k, t_l) = \sum_{m=1}^R \left(\sum_{i=1}^{L_m} Z_{\omega(i)}(\omega_k) \cdot Z_{t(i)}(t_l) \right), \quad (11)$$

де R – кількість складів в еталонній ТП $X_S^*(\omega_k, t_l)$, L_m – кількість параметрів дзвіноподібних функцій $Z_{t(i)}(t_l)$, $Z_{\omega(i)}(\omega_k)$ ($i=1..L_m$) для відповідного m -го складу комбінації для еталонної ТП.

На рис. 4 а), б), в) наведений приклад синтезу ЕТП згідно алгоритму.

Експериментальні дослідження. Першим етапом побудови системи розпізнавання на основі технології сегментно-складового синтезу є побудова словників МО (складів). Для цього попередньо був проаналізований та оброблений тестовий набір слів (цифри: 0-100, 200, 300, 400, 500, 600, 700, 800, 900, 1000). Кількість двосимвольних складів-еталонів – 68, трисимвольних складів-еталонів – 95. Множина ТП $\{Y_{S_k}\}$ для дво- та трисимвольних складів-еталонів $\{SL_{2k}\}$, $\{SL_{3k}\} \subset \{SL_k\}$ розбита на словники: $\{SL_q^2\}$ – двосегментні склади (52); $\{SL_r^3\}$ – трисегментні склади (72); $\{SL_s^4\}$ – чотирисегментні склади (39). В процесі навчання системи розпізнавання накопичуються еталони складів у вигляді: (а) вихідного спектрально-смугового та спектрально-часового подання; (б) параметрів моделі сплайн-опису спектрально-смугового подання; (в) параметрів моделі опису спектрально-часового подання у класі дзвіноподібних функцій.

Згідно з алгоритмом сегментно-складового синтезу, сегментована ТП $X_S(\omega, t)$ поданого мовного сигналу, розглядається як сукупність дво-, три- і чотирисегментних складів, для якої одним із алгоритмів пошуку знаходиться комбінація з ТП МО словника [11, 12] і відповідно будується результат розпізнавання.

Проведені дослідження по розпізнаванню мовленнєвих команд, які належать до тестового набору слів, за допомогою алгоритма сегментно-складового синтезу: А – алгоритм використовує модель опису ТП МО словника поліномами 3-го порядку (2); В – алго-

ритм використовує модель опису ТП МО словника у класі дзвіноподібних функцій (10); С – базовий алгоритм сегментно-складового синтезу.

Середній час T_{cp} , який необхідний для розпізнавання поданої реалізації слова (середня довжина реалізації 7 сегментів) та відсоток похибок E_{gr} , які отримані на комп'ютері типу IBM PC з процесором Intel Celeron 700 МГц наведені у наступній таблиці.

Алгоритм / Критерій	А	В	С
T_{cp} , сек	0,6	0,9	0,1
E_{gr} , %	2,0	1,7	3,3

З таблиці видно, що найменше значення похибки E_{gr} отримане для алгоритму В з прийнятним часом розпізнавання. Отримані результати показують доцільність використання опису МО словника у класі дзвіноподібних функцій в алгоритмі розпізнавання на основі сегментно-складового синтезу.

Модель опису ТП МО словника поліномами 3-го порядку (2) має наступні переваги: невелика кількість лінійних параметрів, що дозволяє будувати словники МО, які займають значно менші ресурси пам'яті; знаходження параметрів моделі засновано на застосуванні швидкодіючих алгоритмів. Проте застосування цієї моделі опису в алгоритмі сегментно-складового синтезу погіршує результати розпізнавання, оскільки виникає необхідність модифікації ТП МО в процесі синтезу.

Дзвіноподібні функції достатньо універсальні для забезпечення можливості варіювання форми сигналу у широких межах із задовільною точністю, є гладкими функціями на всьому діапазоні визначення. Синтез ЕТП з ТП, послідовності параметрів яких описані за допомогою дзвіноподібних функцій, внаслідок їхніх властивостей, уникає проблем зклеювання суміжних ТП, за рахунок чого отримані кращі результати розпізнавання.

Висновки

У статті запропоновано новий підхід до розв'язання задачі розпізнавання мовлення для великих словників на основі сегментно-складового синтезу траєкторії параметрів в класі дзвіноподібних функцій. Розглянуто загальні проблеми обробки параметрів мовних сигналів в процесі синтезу послідовності параметрів, для якої досягнута найкраща близькість з траєкторією параметрів для поданого сигналу.

ЛІТЕРАТУРА

1. Ronzhin A.L., Yusupov R.M., Li I.V., Leontieva A.B. Survey of Russian Recognition Systems. // In Proc. of Int. Conf. SPECOM'2006, St. Petersburg, 2006, pp. 54-60.
2. Ронжин А.Л., Карпов А.А., Ли И.В. Система автоматического распознавания русской речи SIRIUS // Искусственный интеллект. – 2005. - № 3. – С. 590-601.
3. Bahl L.R., Bakis R., De Souza P.V., Mercer R.L. Obtaining candidate words by polling in a large vocabulary speech recognition system // In Proc. ICASSP'88 Int. Conf. Acoust., Speech and Signal Process., NY, 1988, pp. 489-492.

4. Meisel W.S., Anikst M.T., Pirzadeh S.S., Schumacher J.E., Soares M.C., Trawick D.J. The SSI large-vocabulary speaker-independent continuous speech recognition system // In Proc. ICASSP'91 Int. Conf. Acoust., Speech and Signal Process., Toronto, 1991, pp. 337-340.
5. Кодзасов С.В., Кривнова О.Ф. Общая фонетика. – М.: РГГУ, 2001. – 592 с.
6. Карпов О.Н. Технология построения устройств распознавания речи. – Д., 2001. – 184 с.
7. Kopeček I. Speech synthesis based on the composed syllable segments. // In Proc. of the First Workshop on Text, Speech and Dialogue – TSD'98, 1998, pp. 259-262.
8. Vasylyeva N., Sazhok M. Text selection for training procedures under phoneme units variety. // In Proc. SPECOM' 2005, St. Petersburg, 2005. – P. 629-631.
9. Рассел С., Норвиг П. Искусственный интеллект: современный подход – М., 2006. – 1408 с.
10. Шаповалова С.И. Оптимизация решения задач поиска на прологе // Искусственный интеллект. 2000. № 3. С. 121–127.
11. Карпов О.Н., Савенкова О.А. Распознавание речи на основе сегментно-слогового синтеза в терминах пространства состояний // Искусственный интеллект. – 2006. - №3. – С. 532-536.
12. Карпов О.Н., Савенкова О.А. Эвристический алгоритм поиска оптимальных решений сегментно-слогового синтеза // Искусственный интеллект. – 2007. - №4. – С. 378-385.
13. Карпов О.Н., Савенкова О.А. Некоторые эксперименты по повышению надежности распознавания слов заданного словаря // Системные технологии. Региональный межвуз. сборник науч. тр. Выпуск 6 (35), Днепропетровск, 2004, с. 60-66.
14. Карпов О.Н. Вычислительные схемы представления функций многих переменных в классах функций меньшего числа переменных: Аналитическое описание поверхностей и спектров речевых сигналов. Монография. – Д.: Изд-во Днепропетр. ун-та, 2003. – 120 с.

пост. 15.01.08.

Поповнення зі згладжуванням послідовностей відліків функцій двох змінних на основі сплайнів

П.О.ПРИСТАВКА

Національний авіаційний університет

Для чотирикратного поповнення зі згладжуванням послідовностей відліків гладких функцій двох змінних подано лінійні оператори, отримані за використанням двовимірних локальних поліноміальних сплайнів на основі *B*-сплайнів, що є близькими до інтерполяційних у середньому.

Для четырёхкратного пополнения со сглаживанием последовательностей отсчётов гладких функций двух переменных поданы линейные операторы, полученные с использованием двумерных локальных полиномиальных сплайнов на основе *B*-сплайнов, близких к интерполяционным в среднем.

This article is the solution of practical research of the polynomial splines of two variables based on the *B*-splines that, on average, are related to the interpolator. These splines allow us to get simple calculating schemes which are convenient for the practical application.

Постановка проблеми. Розвиток інформаційних технологій стимулює удосконалення математичного апарату опрацювання дискретних даних. Порівняно громіздкі класичні методи апроксимації гладких функцій за подібними даними все менше задовольняють розробників програмного та апаратного забезпечення розв'язку задач стиснення та відтворення зображень, опрацювання сигналів, тощо. Так при обробці растрових зображень з випадковою вадою серед інших, може виникати завдання збільшення розміру вихідного зображення з вимогою придушення білого шуму. Зазвичай, подібна операція супроводжується обчисленнями на великих обсягах даних, представленими результатом оцифрування, що містять файли графічних форматів. Наприклад, якісний цифровий знімок для друку фотографії формату А4-А3 містить приблизно 4 мегапіксели, а враховуючи, що більшість зображень виконано у кольоровому просторі sRGB, останню цифру слід потрої-

ти за рахунок розкладу на складові червоного, зеленого та синього. При проведенні аерофотозйомки кількість пікселів цифрованого зображення на порядок може перевищувати вказаний. Тому, зважаючи на зазначене, при масштабуванні растру актуальною є задача вибору математичного забезпечення обробки, яке відповідало б вимогам адекватної апроксимації та швидкодії обчислень одночасно.

Аналіз досліджень та постановка задачі. Останні кілька десятиріч розвиток отримали методи, що базуються на обчислювальному аспекті, зокрема вейвлети та процедури, основані на бінарному поповненні послідовностей відліків гладких функцій. Щодо до останніх, можна зазначити можливість їх одержання на підставі методів сплайн-апроксимації, зокрема, з використанням поліноміальних сплайнів, визначених на локальних носіях, для яких і обчислювальний апарат і дослідження проведено досить розлого.

