

УДК 681.518:004.93.1

О.А. ЯКУШЕВ

Сумський державний університет, Україна

## ПОЧАТКОВЕ ОБРОБЛЕННЯ СИГНАЛУ ПРИ РОЗПІЗНАВАННІ ГОЛОСОВИХ КОМАНД

Розглядається процес формування вхідного математичного опису при розпізнаванні голосових команд у рамках інформаційно-екстремальної інтелектуальної технології, що ґрунтується на максимізації інформаційної спроможності системи шляхом введення в процесі її навчання додаткових інформаційних обмежень та підвищення достовірності автоматичної класифікації в адаптації математичного опису на вході з метою досягнення на екзаміні достовірності розпізнавання, близької до асимптотичної. У рамках інформаційно-екстремальної інтелектуальної технології розглянемо оптимізацію такого важливого параметру функціонування початкової обробки сигналу як вікно згладжування для віконного перетворення Фур'є з логарифмічною шкалою стиснення частот.

**Ключові слова:** ІЕІ-технологія, віконне перетворення Фур'є, вікно згладжування, критерій функціональної ефективності, навчальна матриця, класифікатор.

### Вступ

Важливою частиною функціонування систем розпізнавання голосових образів є початкова обробка сигналу. Тому оптимізація та формування початкової обробки сигналу дозволяє підвищити ефективність та пристосованість системи.

На практиці початкова обробка сигналу для системи розпізнавання голосових команд полягає в отриманні частотних характеристик акустичного сигналу.

Відомо два найбільш застосовані методи для вирішення проблеми отримання спектру це перетворення Фур'є [1] та вейвлет перетворення [2]. Перетворення Фур'є застосовується для отримання спектру з сигналу оснований на періодичних коливаннях, вейвлет перетворення на отримання спектру з сигналу який немає періодики, або змішаного сигналу.

У статті розглянуто формування вхідного математичного опису для системи розпізнавання голосових команд за інформаційно-екстремальною інтелектуальною технологією (ІЕІ-технологія), що ґрунтується на максимізації інформаційної спроможності системи шляхом введення в процесі навчання додаткових інформаційних обмежень [3].

Однією з основних модулів систем розпізнавання акустичних сигналів є модуль початкової обробки сигналу, який безпосередньо формує вибірковою множини. Тому основною із задач є побудова оптимального механізму початкової обробки. Розглянемо постановку задачі стосовно гіперсферичного класифікатора, який будується в процесі навчан-

ня системи шляхом оптимізації контейнерів класів розпізнавання, що відновлюються в радіальному базисі простору ознак.

### Постановка задачі

Нехай дано алфавіт класів розпізнавання  $\{X_m^0 \mid m = \overline{1, M}\}$  і вектор параметрів функціонування системи розпізнавання  $g_m = \{d_m, x_m, \delta, \tau\}$ , де  $d_m$  – радіус контейнера  $K_m^0 \in X_m^0$ ;  $x_m$  – двійковий еталонний вектор-реалізація класу  $X_m^0$ ,  $x_m \in \Omega_B$ ;  $\delta$  – параметр контрольного поля допусків  $\delta \leq \delta_H / 2$ , де  $\delta_H$  – нормоване (експлуатаційне) поле допусків;  $Q$  – оператор формування навчальної матриці.

Треба визначити оператор  $Q$  таким чином щоб значення критерію наблизалося до свого максимального асимптотичного значення:

$$E^* = \max_Q E_m, \quad (1)$$

де  $E_m$  – критерій функціональної ефективності (КФЕ) [3] навчання системи розпізнавати реалізації класу  $X_m^0$ .

### Результати моделювання

Спектральне представлення акустичного сигналу загальним перетворенням Фур'є має ряд недоліків, які призвели до появи віконного перетворення Фур'є:

1) обмеженість інформативного аналізу сигналу на досить великій часовій ділянці;

2) поява ефекту Гіббса [4] на крайових та екстремальних значеннях сигналу.

Застосування віконного перетворення Фур'є [4] дозволяє позбутися цих недоліків:

$$F(t, w) = \int_{-\infty}^{\infty} f(\tau)w(\tau-t)e^{-i\omega t}d\tau, \quad (2)$$

де  $w(\tau-t)$  – деяка віконна функція, яка дозволяє позбутися ефекту Гіббса, згладити бокові лепестків сигналу та зменшити відсоток шуму. Серед віконних функцій [5] можна виділити такі:

1) вікно Ханна:

$$w(t) = 0,5 \cos\left(\frac{2\pi t}{N}\right), \quad (3)$$

де  $t$  – час зняття інформації,  $N$  – розмір вікна перетворення Фур'є;

2) вікно Хеммінга:

$$w(t) = 0,53836 - 0,46164 \cos\left(\frac{2\pi t}{N}\right); \quad (4)$$

3) вікно Кайзера:

$$w(t) = \frac{I_0\left(\beta\sqrt{1-\left(\frac{2t-N-1}{N-1}\right)^2}\right)}{|I_0(\beta)|}, \quad (5)$$

де  $I_0$  – модифікована функція Бесселя першого роду нульового порядку,  $\beta$  – коефіцієнт визначаючий долю енергії, зосередженої в головному лепестку віконної функції;

4) вікно Блекмана:

$$w(t) = 0,42 - 0,5 \cos\left(\frac{2\pi t}{N}\right) + 0,08 \cos\left(\frac{4\pi t}{N}\right). \quad (6)$$

У табл. 1 наведено результати порівняльного аналізу віконних функцій для 8 голосових команд при реалізації алгоритму навчання системи розпізнавання з паралельною оптимізацією контрольних допусків [3].

Таблиця 1

Порівняння віконних функцій під час роботи паралельного алгоритму оптимізації

	(Більше; Менше)	(Лівише; Правише)	(Включити; Виключити)	(Старт; Стоп)	Середнє значення
вікно Ханна	7,65	7,65	3,71	5,33	6,08
вікно Хеммінга	7,65	7,65	3,21	4,35	5,72
вікно Кайзера	7,65	7,65	3,71	3,71	5,68
вікно Блекмана	7,65	7,65	4,35	5,33	6,25

У табл. 1 значення в клітинках є значення КФЕ після роботи класифікатора, який застосовує паралельний алгоритм оптимізації [3]. У першому рядку наведені команди для класифікаторів. Наведені у клітинах значення це значення КФЕ при розпізнаванні вище наведених команд. Аналізуючи табл. 1 бачимо що середнє значення КФЕ досягається при застосуванні вікна Блекмана.

При отриманні вихідних даних спектру треба враховувати, що для формування ознак треба акцентувати на інформативні частоти, тому для акцентування на більш інформативні частоти використовується mel-стиснення [5]:

$$H_m = 1125 \cdot \log(0,0016H_0 + 1), \quad (7)$$

де  $H_0$  – частота існуючого спектру, Гц,  $H_m$  – частота в новому зжатому частотному просторі.

Загальний спектр, який отримується під час пересування вікна, показано на рис. 1.

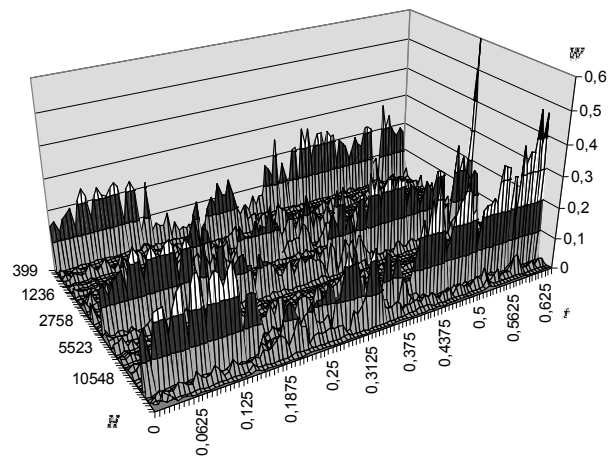


Рис. 1. Частотний спектр акустичного сигналу при застосуванні віконного перетворення Фур'є

На рис. 1 прийняті такі позначення:  $H$  – частота,  $t$  – час зняття інформації віконним перетворенням Фур'є,  $W$  – енергія частоти. З метою формування словника ознак для наведеного на рис. 1 спектру для кожної частоти визначено максимальне глобальне значення енергії частоти за часом перетворення Фур'є, узагальнений спектр наведено на рис. 2. Параметри спектру (рис. 2) визначають узагальнені ознаки розпізнавання, які відбивають характеристики амплітуди та фази акустичного сигналу.

Функціональна ефективність запропонованого способу оброблення акустичного сигналу перевірилась у процесі навчання системи розпізнавання голосових команд «включити» та «виключити» у рамках ІЕІ-технології [3]. При цьому алгоритм навчання було реалізовано з паралельною оптимізацією системи контрольних допусків на ознаки розпізнавання при застосуванні вікна Блекмана [5].

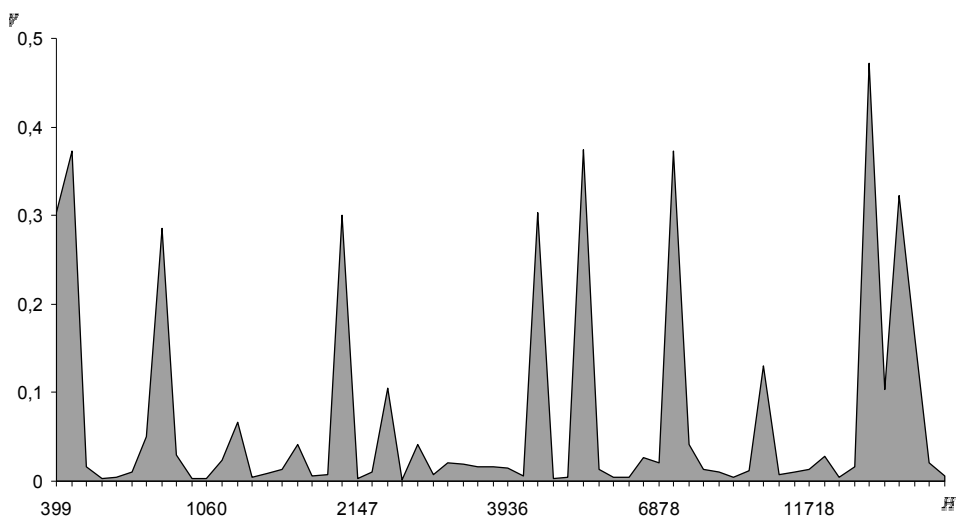


Рис. 2. Узагальнений спектр акустичного сигналу

При цьому як критерій оптимізації параметрів навчання системи розпізнавання використовувався інформаційний критерій Кульбака [3]:

$$E = (D_{1,1}^{(k)} - \beta) \log_2 \left( \frac{1 + (D_{1,1}^{(k)} - \beta) + 10^{-\tau}}{1 - (D_{1,1}^{(k)} - \beta) + 10^{-\tau}} \right), \quad (8)$$

де  $D_{1,1}^{(k)}$  – достовірність першого роду;  $\beta$  – помилка другого роду; де  $\tau$  – число цифр у мантисі значення критерію Кульбака.

Графік залежності КФЕ Кульбака від системи контрольних допусків наведено на рис. 3, де  $E$  – критерій функціональної ефективності який є функціоналами від точнісних характеристик класифікатора, а  $\delta$  – системи контрольних допусків [3]. Також темною ділянкою графіка позначено робочу (допустиму) область визначення функції, в якій  $D_{1,1}^{(k)}$  і  $D_{1,2}^{(k)}$  більше 0,5.

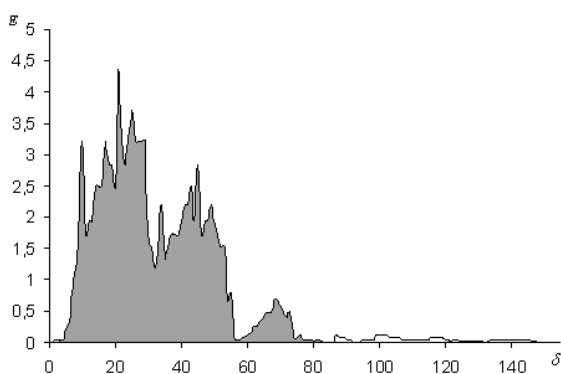


Рис. 3. Графік залежності КФЕ Кульбака від системи контрольних допусків

Алгоритм оптимізації проводився за першим класом, який характеризує команду «включити» та складається з 30 реалізацій та 60 ознак. Аналіз рис. 3 показує, що оптимальне значення критерію функціональної ефективності  $E$  дорівнює 4,35.

Більш детальну інформація про оптимізацію контрольних допусків наведено у табл. 2, де темною ділянкою позначено робочу (допустиму) область визначення функції критерію.

Таблиця 2

Залежність КФЕ від поля контрольних допусків

$\delta$	$E$	$D_1$	$\beta$	$d_1$	$d_c$
0	0,0000	1,0000	1,0000	1	0
1	0,0000	0,5333	0,5333	3	0
2	0,0287	0,6667	0,5667	6	0
3	0,0511	0,5667	0,4333	8	1
4	0,0287	0,6000	0,5000	10	2
5	0,2081	0,6667	0,4000	12	2
...	...	...	...	...	...
18	2,8193	0,9667	0,1333	19	20
19	2,8193	0,9000	0,0667	18	20
20	2,4860	0,9000	0,1000	19	22
<b>21</b>	<b>4,3529</b>	<b>1,0000</b>	<b>0,0667</b>	<b>18</b>	<b>22</b>
22	3,2160	0,9667	0,1000	18	20
23	2,8193	0,9000	0,0667	16	19
...	...	...	...	...	...
80	0,0127	0,5333	0,4667	3	2
81	0,0287	0,5667	0,4667	3	2
82	0,0127	0,5667	0,5000	3	2
83	0,0032	0,5667	0,5333	3	1

У табл. 2 за значення  $\beta$  прийняте значення помилки другого роду, яке дорівнює  $\beta = 1 - D_{1,2}^{(k)}$ ,  $d_1$  – радіус контейнера першого класу,  $d_c$  – відстань між центрами першого та другого класів. Аналіз табл. 2 показує, що за оптимальне значення параметра  $\delta$  слід вибрати  $\delta^* = 21$ , тому що значення  $E$  для цього параметру є максимальним. Оптимальний параметр поля допусків дорівнює  $\delta^* = 21$ , при цьо-

му КФЕ  $E = 4,3529$  не досягає свого максимального значення, оскільки  $D_1 = 1$  і  $\beta = 0,07$ . Таким чином, не побудовано безпомилковий за навчальною матрицею класифікатор, але значення критерію достатньо високе та значення першої та другої достовірності близькі до 1, що дозволяє стверджувати, що при подальшій оптимізації параметрів функціонування системи значення критерію буде наближеним до максимального значення.

### Висновки

Початкове оброблення сигналів із застосування віконного перетворення Фур'є забезпечує високу достовірність розпізнавання голосових команд.

Застосування одержаних результатів може бути перспективним при розробці інформаційного та алгоритмічного забезпечення інтелектуальних сис-

тем розпізнавання голосових команд, наприклад в медицині, робототехніці тощо.

### Література

1. Kammler D.W. *First Course in Fourier Analysis* / D.W. Kammler. – NJ: Prentice Hall, 2007. – 798 p.
2. Debnath L. *Wavelet Transforms and Their Applications* / L. Debnath. – H.: Boston book, 2002. – 565 p.
3. Довбиш А.С. *Основи проектування інтелектуальних систем: навчальний посібник* / А.С. Довбиш. – Суми: СумДУ, 2009. – 171 с.
4. Марпл С. *Цифровой спектральный анализ и его приложения* / С. Марпл. – М.: Мир, 1990. – 584 с.
5. Шелухин О. *Цифровая обработка и передача речи* / О. Шелухин, Н. Лукьянцев. – М.: Радио и связь. 2000. – 456 с.

Надійшла до редакції 30.03.2010

**Рецензент:** д-р техн. наук, проф., зав. каф. інформатики О.Ю. Соколов, Національний аерокосмічний університет ім. М.С. Жуковського «Харківський авіаційний інститут», Харків.

## НАЧАЛЬНАЯ ОБРАБОТКА СИГНАЛА ПРИ РАСПОЗНАВАНИИ ГОЛОСОВЫХ КОМАНД

*А.А. Якушев*

Рассматривается процесс формирования входного математического описания при распознавании голосовых команд в рамках информационно-экстремальной интеллектуальной технологии, что базируется на максимизации информационной способности системы путём введения в процесс её обучения дополнительных информационных ограничений и повышение вероятности автоматической классификации в адаптации математического описания на входе с целью достижения на экзамене вероятности распознавания, близкой к асимптотической. В основных концепциях информационно-экстремальной интеллектуальной технологии рассмотрим оптимизацию важного параметра функционирования начальной обработки сигнала как окно сглаживания для оконного преобразования Фурье с логарифмической шкалой сжатия частот.

**Ключевые слова:** ИЕИ-технология, оконное преобразование Фурье, окно сглаживания, критерий функциональной эффективности, обучающая матрица, классификатор.

## INITIAL SIGNAL PROCESSING AT VOCAL COMMANDS RECOGNITION

*A.A. Yakushev*

The process of forming entrance mathematical description vocal commands recognition system is examined within the bounds of information-extreme intelligence technology which is based on informative possibility maximization of the system by introduction in the process of its learning additional informative limitations and the increase of automatic classification probability in mathematical description adaptation on an entrance with the purpose achievement on examination of recognition probability, near to asymptotic. In basic conceptions of information-extreme intelligence technology will consider optimization of important parameter functioning of the initial signal processing as window of smoothing out window transformation of Fourier with the logarithmic scale of frequencies compression.

**Key words:** IEI-technology, short-time Fourier transform, window function, criterion of functional efficiency, teaching matrix, classifier.

**Якушев Александр Анатольевич** – аспирант кафедры информатики, Сумский государственный университет, Сумы, Украина, e-mail: info@id.sumdu.edu.ua.