

УДК 621.391.25

Юрій Романишин, Володимир Гудим

Національний університет "Львівська політехніка",  
кафедра електронних засобів інформаційно-комп'ютерних технологій**КОМПРЕСІЯ МОВНИХ СИГНАЛІВ НА ОСНОВІ  
ДИСКРЕТНИХ ХВИЛЬОВИХ ПЕРЕТВОРЕНЬ**

© Романишин Юрій, Гудим Володимир, 2001

Розглянуто застосування дискретних хвильових перетворень та психоакустичної моделі слухового сприйняття для компресії мовної інформації, проведений аналіз впливу вибору різних використовуваних хвильових функцій, тривалості часового вікна, кількості рівнів розкладу сигналу на параметри компресії.

**An applying of discrete wavelet transform and psychoacoustics model of audio perception for speech information compression is considered. The influence analysis of different wavelet functions, time window and number of signal decomposition levels on compression parameters is worked out.**

Компресія мовних сигналів з можливістю їх подальшого якісного відтворення є однією з важливих задач перетворення мовних сигналів для їх зберігання та передачі. За відомими оцінками у мовному сигналі є лише близько 10-15% корисної інформації, решта даних є надлишковими і їх можна усунути з сигналу без його суттєвого спотворення з точки зору слухового сприйняття мови. У зв'язку з цим для компресії мовних сигналів, на відміну від компресії довільних інформаційних даних, звичайно використовують методи втратної компресії, при яких у сигналі залишається лише необхідна корисна інформація. Для виділення цієї корисної інформації використовуються різні методи перетворення сигналів, більшість з яких базується на розділенні апіорі відомого частотного діапазону сигналу на окремі смуги, деякі з яких мало впливають на слухову відтворюваність сигналу і їх можна усунути. На відміну від звичайного частотного смугового кодування та компресії, які використовуються у стандартах MPEG, застосування дискретних хвильових перетворень з ортогональними базисними функціями, серед яких можна виділити такі, які близькі до фрагментів сигналу, обумовлює потенційні можливості кращої компресії.

Дискретне хвильове перетворення сигналу  $S[i]$  ( $i = \overline{1, m}$ ,  $m$  – кількість відліків сигналу) здійснюється з використанням масштабуючої функції  $\varphi(t)$ , яка при кожному з масштабів  $2^j$  задовольняє умову ортонормованості щодо зсувів у часі на  $2^{-j}k$  та  $2^{-j}m$  ( $j, k, m \in Z$ ):

$$\int_{-\infty}^{\infty} 2^{j/2} \varphi(2^j t - k) 2^{j/2} \varphi(2^j t - m) dt = \delta_{km}, \quad (1)$$

де  $\delta_{km}$  – символ Кронекера.

Крім того, функція  $\varphi(t)$  задовольняє умову нормування

$$\int_{-\infty}^{\infty} \varphi(t) dt = 1. \quad (2)$$

З масштабуючою функцією  $\varphi(t)$  зв'язана хвильова функція  $\psi(t)$ , дискретні відліки якої визначаються через відліки функції  $\varphi(t)$  співвідношенням

$$\varphi[i] = (-1)^i \psi[n+1-i]; \quad i = \overline{1, n}, \quad (3)$$

де кількість відліків  $n$  визначається функціями  $\varphi(t)$  та  $\psi(t)$ .

Дискретні відліки  $\tilde{\varphi}[i] = \varphi[n+1-i]$  та  $\tilde{\psi}[i] = \psi[n+1-i]$  є дискретними імпульсними характеристиками цифрових фільтрів відповідно нижніх та верхніх частот. Розклад сигналу за заданими дискретними функціями  $\varphi[i]$  та  $\psi[i]$  здійснюється у відповідності зі схемою, зображеною на рис. 1.

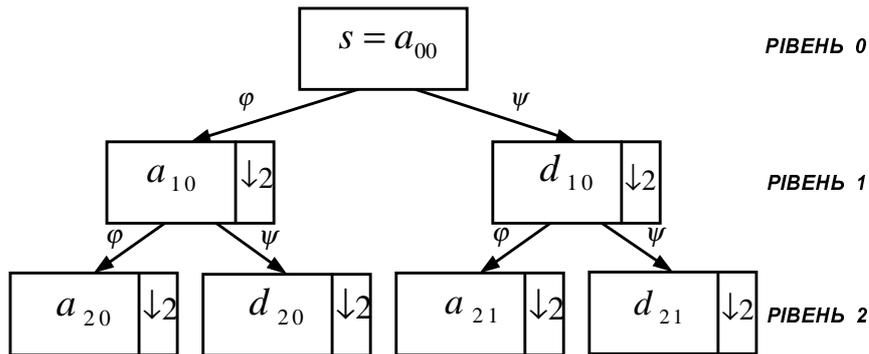


Рис. 1. Багаторівнева декомпозиція сигналу

Сигнал послідовно розкладається на деяку кількість рівнів. На кожному рівні сигнал формується з сукупності підрівнів, яким відповідають коефіцієнти апроксимації  $a_{jr}$  та коефіцієнти деталей  $d_{jr}$  ( $j$  – номер рівня;  $r$  – номер пари підрівнів). Кожний з підрівнів можна розкласти на два підрівні на нижчому рівні. Коефіцієнти  $a_{jr}$  утворюються внаслідок цифрової фільтрації сигналу на вищому рівні фільтром нижніх частот з імпульсною характеристикою  $\tilde{\varphi}[i]$ , а коефіцієнти  $d_{jr}$  – фільтром верхніх частот з характеристикою  $\tilde{\psi}[i]$  з подальшим проріджуванням ( $\downarrow 2$ ). Ці коефіцієнти визначаються за рекурентними співвідношеннями:

$$a_{j+1,2r}[k] = \sqrt{2} \sum_{i=\max(1;2k+1-n)}^{\min(n;2k)} a_{jr}[i] \varphi[i+n-2k]; \quad (4)$$

$$d_{j+1,2r}[k] = \sqrt{2} \sum_{i=\max(1;2k+1-n)}^{\min(n;2k)} a_{jr}[i] \psi[i+n-2k]; \quad (5)$$

$$a_{j+1,2r+1}[k] = \sqrt{2} \sum_{i=\max(1;2k+1-n)}^{\min(n;2k)} d_{jr}[i] \varphi[i+n-2k]; \quad (6)$$

$$d_{j+1,2r+1}[k] = \sqrt{2} \sum_{i=\max(1;2k+1-n)}^{\min(n;2k)} d_{jr}[i] \psi[i+n-2k]; \quad (7)$$

$$j = 0; r = 0; \quad j = 1, 2, \dots; r = 0, 1, \dots, 2^{j-1} - 1.$$

Формули відтворення коефіцієнтів апроксимації та деталей вищого рівня за коефіцієнтами нижчого рівня мають вигляд

$$a_{jr}[2k-1] = \sqrt{2} \sum_{i=k}^{k+n/2-1} (a_{j+1,2r}[i]\varphi[n+1-2i] + d_{j+1,2r}[i]\psi[n+1-2i]); \quad (8)$$

$$a_{jr}[2k] = \sqrt{2} \sum_{i=k}^{k+n/2-1} (a_{j+1,2r}[i]\varphi[n+2-2i] + d_{j+1,2r}[i]\psi[n+2-2i]); \quad (9)$$

$$d_{jr}[2k-1] = \sqrt{2} \sum_{i=k}^{k+n/2-1} (a_{j+1,2r+1}[i]\varphi[n+1-2i] + d_{j+1,2r+1}[i]\psi[n+1-2i]); \quad (10)$$

$$d_{jr}[2k] = \sqrt{2} \sum_{i=k}^{k+n/2-1} (a_{j+1,2r+1}[i]\varphi[n+2-2i] + d_{j+1,2r+1}[i]\psi[n+2-2i]); \quad (11)$$

$k = \overline{1, m_j/2}$  при  $m_j$  парному.

Для обчислення коефіцієнтів апроксимації та деталей і відтворення сигналів за їх розкладами використані відповідно функції DWT та IDWT математичного пакета MATLAB [1].

Для компресії сигналу проводиться аналіз значень коефіцієнтів апроксимації та деталей на деякому рівні розкладу та відкидаються ті підрівні, значення всіх коефіцієнтів на яких не перевищує заданого порогового значення.

Якість компресії характеризують два основні числові параметри – коефіцієнт компресії та коефіцієнт кореляції між початковим та відтвореним після компресії сигналом. Під коефіцієнтом компресії будемо розуміти  $z = N_0/N$ , де  $N_0$  – початкова кількість відліків сигналу;  $N$  – кількість відліків у компресованому сигналі. Коефіцієнт кореляції визначається співвідношенням

$$R = \frac{\sum_{i=1}^m S[i]\tilde{S}[i]}{\sqrt{\sum_{i=1}^m S^2[i]\sum_{i=1}^m \tilde{S}^2[i]}}, \quad (12)$$

де  $S[i]$  та  $\tilde{S}[i]$  – відліки відповідно початкового та відтвореного після компресії сигналу.

Крім загального критерію (12) порівняння сигналів, при компресії звукових сигналів використовуються особливості слухового сприйняття звуку людиною, які виражаються психоакустичною моделлю, що описує частотну залежність порога сприйняття звуку людиною і виражається формулою [2, 3]

$$T(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4, \quad (13)$$

де  $f$  – частота в Гц;  $T(f)$  – відношення порога чутливості за енергією на частоті  $f$  до порога на частоті 2 кГц в дБ.

На рис. 2 зображена ця частотна залежність порога чутливості для амплітуди сигналу. Хоча ця модель використовується при розкладі сигналу на частотні смуги в стандартах компресії MPEG, ця ж модель пристосована для компресії на основі хвильового перетворення. Оскільки кожний підрівень хвильового перетворення отри-мується внаслідок цифрової фільтрації певної смуги частот, обчислюється середнє значення порогової функції психоакустичної моделі (13) для відповідного підрівня, яке перемножується з кожним значенням

коефіцієнтів апроксимації та деталей розкладу сигналу. У результаті реконструкції хвилових коефіцієнтів отримується сигнал, який має меншу кількість частотних складових (коефіцієнтів розкладу), але достатньо чітко сприймається людським вухом. На рис. 3 відображено заданий мовний сигнал та реконструйований на основі хвилових перетворень з використанням психоакустичної моделі при кількості рівнів розкладу 6. Коефіцієнт кореляції між цими сигналами дорівнює 0,9356.

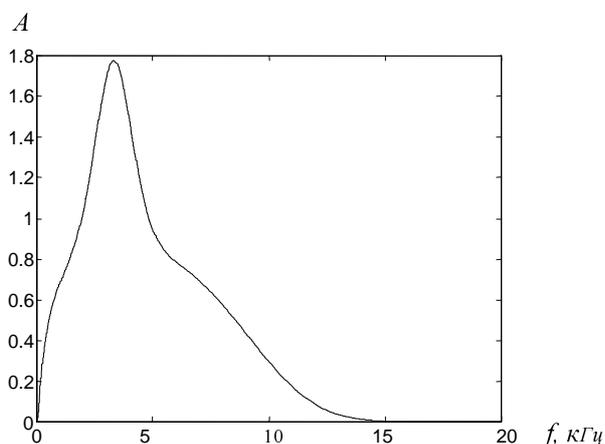


Рис. 2. Частотна залежність амплітудного порога чутливості

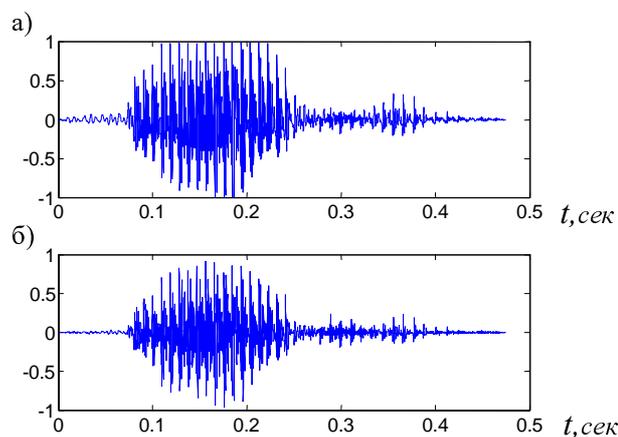


Рис. 3. Мовний сигнал:  
а – початковий; б – після розкладу та відтворення з використанням психоакустичної моделі

Однією з особливостей DWT-розкладу є збільшення кількості дискретних відліків у перетвореному сигналі порівняно з початковим. Такого збільшення немає, наприклад, при дискретному перетворенні Фур'є (з врахуванням комплексно-спряженості відповідних значень спектра). Загальна кількість відліків у сигналі після DWT-перетворення становить  $p = 2[(m + n - 1) / 2]$ , де  $m$  – кількість відліків у вхідному сигналі;  $n$  – кількість відліків хвилової функції;  $[\ ]$  – ціла частина числа. Таке збільшення є несуттєвим при  $m \gg n$ , але в протилежному випадку відносне збільшення кількості відліків не дозволяє отримати великий коефіцієнт компресії. На рис. 4 показана залежність загальної кількості додаткових значень при розкладі сигналу від номера рівня розкладу з використанням масштабуючої функції db3 ( $n = 6$ ).

Суттєве значення для параметрів компресії мовного сигналу має вибір хвилової функції для розкладу сигналу. На відміну від спектрального аналізу з фіксованим гармонічним базисом, хвиловий аналіз дає широкі можливості для використання різних базисів, але при цьому виникає задача вибору кращого базису серед певного класу базисів. Очевидно, що ці базиси необхідно якось порівняти між собою з точки зору ефективності розкладу за ними. Ефективність розкладу тим вища, чим більш чітко виділяються підрівні з великими значеннями коефіцієнтів на певному рівні розкладу. Одним з використовуваних параметрів для оцінки ефективності розкладу є ентропія сигналу  $H$ , яка визначається співвідношенням

$$H(S) = - \sum_{i=1}^n a[i] \log a[i], \quad (14)$$

де  $a[i] = |s[i]| / \sum_{i=1}^n |s[i]|$  – нормовані абсолютні значення відліків сигналу.

Для аналізу за ентропією було вибрано сімейство дискретних хвильових функцій db1-db15, які охоплюють широкий клас сигналів. Найменше значення ентропії сукупності коефіцієнтів апроксимації та деталей на 5-му рівні розкладу при використанні функції db10 ( $H(S)=0.0426$ ). Однак недоліком використання функцій db(i) високого порядку є збільшення кількості дискретних відліків функції та відповідне збільшення кількості додаткових значень при розкладі.

У зв'язку з властивістю хвильових перетворень збільшувати кількість коефіцієнтів виникає також задача вибору тривалості часового вікна та рівня розкладу. При стисканні сигналу з вікнами короткої тривалості додаткова інформація, зумовлена властивостями хвильового перетворення, зменшує коефіцієнт компресії. При використанні великих вікон вплив цього ефекту зменшується, однак, отримується деяка надлишкова інформація на підрівнях, що зумовлена розширенням спектра частот у вікні. На рис. 5 показана залежність коефіцієнта компресії від тривалості часового вікна розкладу сигналу.

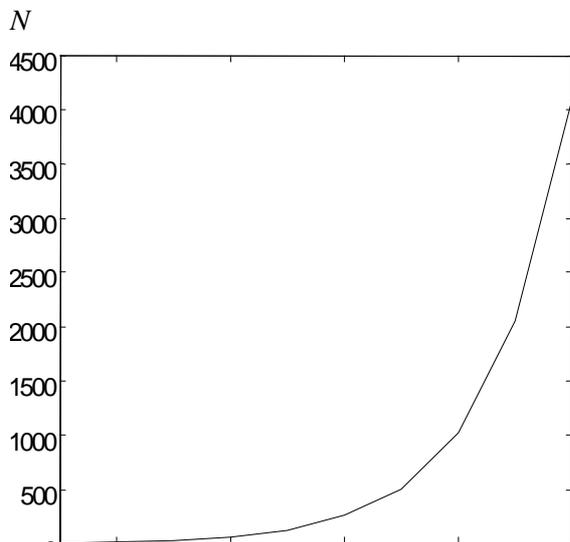


Рис. 4. Залежність кількості додаткових значень від рівня декомпозиції

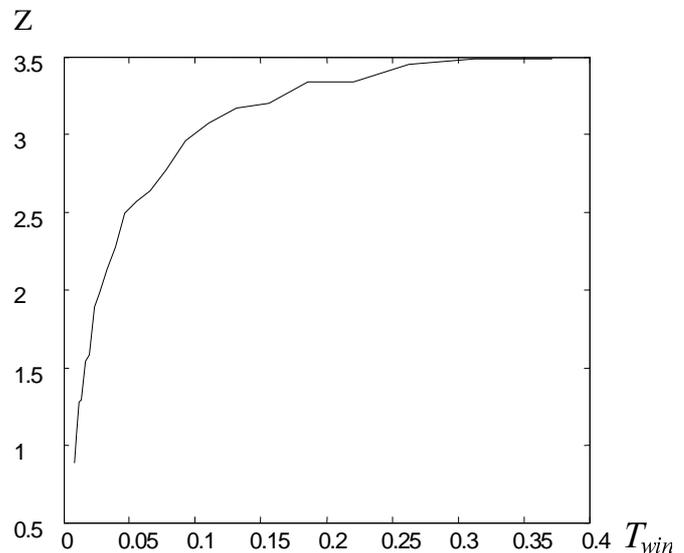


Рис. 5. Залежність коефіцієнта компресії від тривалості вікна

Нижче наведені деякі практичні результати застосування дискретних хвильових перетворень для компресії мовної інформації. Для дослідження був вибраний дискретний мовний сигнал  $s[i]$ ;  $i = \overline{1, m}$ , що відповідає слову “база” при частоті дискретизації 11.05 кГц та кількості бітів квантування 16. Тривалість сигналу становила 0.5 с при кількості дискретів  $m = 5525$ . Для компресії з сигналу послідовно виділялися часові вікна різної тривалості, сигнал розкладався на основі дискретних хвильових перетворень з різними масштабуючими функціями до 5-6-го рівнів. У кожному вікні після розкладу залишалися лише ті підрівні, максимальні значення коефіцієнтів розкладу на яких перевищували 25 % від максимального значення по всіх підрівнях, після чого сигнал реконструювався та визначалися його коефіцієнти кореляції та компресії.

У табл. 1 наведені значення коефіцієнтів кореляції та компресії для різних тривалостей часового вікна при розкладі сигналу на 6-му рівні та масштабуючій функції db3.

Таблиця 1

## Значення коефіцієнтів кореляції та компресії

Тривалість вікна, с	Коефіцієнт кореляції	Коефіцієнт компресії
0.00816	0.99550	0.889
0.01161	0.99466	1.275
0.01642	0.99560	1.546
0.02322	0.99477	1.890
0.03283	0.99625	2.134

Тривалість вікна, с	Коефіцієнт кореляції	Коефіцієнт компресії
0.06567	0.99597	2.636
0.09288	0.99444	2.960
0.13134	0.99574	3.169
0.18576	0.99409	3.338
0.26268	0.99560	3.448
0.37152	0.99520	3.486

З наведених значень можна зробити висновок, що коефіцієнт кореляції, на відміну від коефіцієнта компресії, слабо залежить від тривалості часового вікна.

У табл. 2 наведені числові параметри компресії для різних масштабуючих (хвильових) функцій при тривалості часового вікна 92 мс та розкладі сигналу на 5-му рівні.

Таблиця 2

## Числові параметри компресії

Хвильова функція	Коефіцієнт кореляції	Коефіцієнт компресії
db1	0.91574	7.314
db3	0.92349	6.636
db5	0.96496	6.111
db7	0.97288	5.693
db9	0.97952	5.365
db10	0.98018	5.216
sym2	0.93343	6.952
sym4	0.95695	6.357
sym6	0.96504	5.891
sym8	0.97247	5.528
coif1	0.92359	6.636
coif3	0.96804	5.365
coif5	0.97324	4.633

Хвильова функція	Коефіцієнт кореляції	Коефіцієнт компресії
bior2.2	0.92057	6.636
bior2.4	0.93047	6.111
bior2.6	0.95118	5.693
bior2.8	0.95040	5.365
bior3.1	0.84966	6.952
bior3.3	0.93540	6.357
bior3.5	0.93169	5.891
bior3.7	0.94279	5.528
bior3.9	0.93771	5.216
bior4.4	0.96289	6.111
bior5.5	0.96308	5.891
bior6.8	0.96439	5.365

Найбільше значення коефіцієнта компресії для функції db1 зумовлене тим, що лише при хвильовому розкладі за цією функцією не збільшується кількість коефіцієнтів, однак, при цьому отримується менший коефіцієнт кореляції.

1. Misiti M., Misiti Y., Oppenheim G., Poggi J.-M. *Wavelet Toolbox. For Use with MATLAB. User's Guide. Version 1. Computation. Visualization. Programming.* 1996. 2. Painter T., Spanias A. *A Review of Algorithms for Perceptual Coding of Digital Audio Signals* // <http://www.eas.asu.edu/~spanias/papers/dsp97.ps>. – 30 pp. 3. Painter T., Spanias A. *Perceptual Coding of Digital Audio* // <http://www.eas.asu.edu/~spanias/audiopaper1.pdf> – 60 pp.