

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ**

**Лавриненко Олександр Юрійович**

УДК 621.391

**МЕТОДИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ  
СЕМАНТИЧНОГО КОДУВАННЯ МОВНИХ СИГНАЛІВ**

05.12.02 – «Телекомунікаційні системи та мережі»

**Автореферат**  
дисертації на здобуття наукового ступеня  
кандидата технічних наук

Київ – 2021

Дисертацією є рукопис.

Робота виконана в Національному авіаційному університеті Міністерства освіти і науки України.

**Науковий керівник:**

доктор технічних наук, професор  
**Конахович Георгій Філімонович,**  
Національний авіаційний університет,  
професор кафедри телекомунікаційних  
та радіоелектронних систем.

**Офіційні опоненти:**

доктор технічних наук, професор  
**Сайко Володимир Григорович,**  
Київський національний університет  
імені Тараса Шевченка, професор кафедри  
прикладних інформаційних систем;

доктор технічних наук, професор  
**Климаш Михайло Миколайович,**  
Національний університет «Львівська  
політехніка», завідувач кафедри  
телекомунікацій.

Захист відбудеться «27» серпня 2021 р. о 10<sup>00</sup> на засіданні спеціалізованої вченої ради Д 26.062.19 у Національному авіаційному університеті за адресою: Україна, 03058, м. Київ, пр-т. Любомира Гузара, 1, ауд. 6.205.

З дисертацією можна ознайомитись у Науково-технічній бібліотеці Національного авіаційного університету за адресою: 03058, м. Київ, пр-т. Любомира Гузара, 1.

Автореферат розісланий «26» липня 2021 р.

Учений секретар  
спеціалізованої вченої ради  
д.т.н., доцент



Р.С. Одарченко

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність дослідження.** На сьогоднішній день у зв'язку з активним розвитком технологій розпізнавання і синтезу мови, голосового управління технічними об'єктами, низькошвидкісного кодування мовної інформації, голосового перекладу з іноземних мов і т.д., вирішення задачі семантичного кодування мовних сигналів має важливе науково-технічне значення, тому що саме від нього залежить ефективність функціонування даного типу систем. А якщо врахувати наростаючу тенденцію дистанційної взаємодії людей і роботизованої техніки за допомогою даних технологій, то безсумнівно піднімається головна проблема в телекомунікаційних системах, а саме підвищення пропускну здатності каналу передачі семантичних мовних даних за рахунок ефективного їх кодування, тобто формулюється питання підвищення ефективності семантичного кодування, а саме – з якою мінімальною швидкістю можливо розувати семантичні ознаки мовних сигналів із заданою ймовірністю безпомилкового їх розпізнавання? Саме на це питання буде дана відповідь у даному науковому дослідженні, що є актуальною науково-технічною задачею.

**Зв'язок роботи з науковими програмами, планами, темами.** Результати дисертаційної роботи відображені в рамках держбюджетних науково-дослідних робіт Національного авіаційного університету за темами: «Технологія створення, експлуатації та експертизи комплексних систем захисту інформації» (шифр № 674 - ДБ10, № держреєстрації 0110U000225), «Створення і дослідження нових систем захищеного авіаційного радіозв'язку в рамках Концепції CNS / ATM ICAO» (шифр № 874 - ДБ13, № держреєстрації 0113U000093) на замовлення Міністерства освіти і науки України, в яких здобувач брав участь в якості безпосереднього відповідального виконавця. Також отримані результати використовуються в навчальному процесі кафедри телекомунікаційних та радіоелектронних систем Національного авіаційного університету в рамках навчальної дисципліни – «Методи цифрової обробки мовних сигналів».

**Мета і задачі дослідження.** Метою дисертаційної роботи є розроблення методів підвищення ефективності семантичного кодування мовних сигналів.

Для досягнення поставленої мети **необхідно вирішити наступні наукові задачі:**

1. дослідити відомий метод підвищення ефективності семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів;
2. обґрунтувати ефективність використання адаптивного емпіричного вейвлет-перетворення в задачах кратномасштабного аналізу та семантичного кодування мовних сигналів;
3. розробити метод семантичного кодування мовних сигналів на основі адаптивного емпіричного вейвлет-перетворення з подальшим застосуванням спектрального аналізу Гільберта;
4. провести оцінку підвищення ефективності розробленого методу семантичного кодування мовних сигналів на відміну від існуючого методу.

**Об'єктом дослідження** є процеси семантичного кодування мовних сигналів.

**Предметом дослідження** є методи підвищення ефективності семантичного кодування мовних сигналів.

**Методи дослідження.** Проведені дослідження базуються на сучасних методах:

1. спектрального аналізу (емпіричне вейвлет-перетворення, побудова адаптивних вейвлет-фільтрів Мейєра, знаходження функції внутрішніх емпіричних мод, кепстральний аналіз, перетворення Гільберта для знаходження семантичних ознак мовних сигналів);

2. цифрової обробки сигналів (сегментація спектра Фур'є, обробка банком трикутних мел-частотних фільтрів, логарифмування енергії спектра Фур'є, порогова обробка вейвлет-коефіцієнтів для знаходження семантичних ознак мовних сигналів);

3. теорії електричного зв'язку (оцінювання коефіцієнта стиснення, швидкості передачі бітів, відношення сигнал/шум та пікового відношення сигнал/шум семантичних ознак мовних сигналів);

4. теорії інформації та кодування (оцінювання кількості інформації, ентропії джерела, ефективності кодування, коефіцієнта надмірності та швидкість кодування семантичних ознак мовних сигналів);

5. теорії ймовірностей і математичної статистики (оцінювання коефіцієнта кореляції, математичного очікування, дисперсії, середньоквадратичної помилки та ймовірності безпомилкового розпізнавання семантичних ознак мовних сигналів).

**Наукова новизна одержаних результатів** полягає в наступному:

1. *вперше розроблено* метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення, який відрізняється від існуючих методів побудовою множини адаптивних смугових вейвлет-фільтрів Мейера з подальшим застосуванням спектрального аналізу Гільберта для знаходження миттєвих амплітуд і частот функцій внутрішніх емпіричних мод, що дозволить визначити семантичні ознаки мовних сигналів та підвищити ефективність їх кодування;

2. *вперше запропоновано* використовувати метод адаптивного емпіричного вейвлет-перетворення в задачах кратномасштабного аналізу та семантичного кодування мовних сигналів, що дозволить підвищити ефективність спектрального аналізу за рахунок розкладання високочастотного мовного коливання на його низькочастотні складові, а саме внутрішні емпіричні моди;

3. *отримав подальший розвиток* метод семантичного кодування мовних сигналів на основі мел-частотних спектральних коефіцієнтів, але з використанням базових принципів адаптивного спектрального аналізу за допомогою емпіричного вейвлет-перетворення, що підвищує ефективність даного методу.

*Наукова новизна одержаних результатів підпадає під 16, 17 і 18 пункти паспорта спеціальності 05.12.02 – «Телекомунікаційні системи та мережі».*

**Практичне значення одержаних результатів** полягає в наступному:

1. розроблено метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення, що дозволяє знизити швидкість кодування від 320 до 192 біт/с та необхідну смугу пропускання від 40 до 24 Гц з ймовірністю безпомилкового розпізнавання близько 0.96 (96%) і відношенням сигнал/шум 48 дБ, згідно чого його ефективність підвищується в 1.6 рази на відміну від існуючого методу;

2. розроблено алгоритм семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення та його програмна реалізація мовою програмування MATLAB R2020b.

Отримані в дисертаційній роботі результати можуть бути використані для побудови систем дистанційної взаємодії людей і роботизованої техніки за допомогою мовних технологій, таких як розпізнавання і синтез мови, голосове управління технічними об'єктами, низькошвидкісне кодування мовної інформації, голосовий переклад з іноземних мов і т.д. Результати дослідження впроваджені в науково-технічну діяльність Навчально-науково-виробничого комплексу «Інформаційно-комунікаційні системи» та навчально-науковому процесі кафедри телекомунікаційних та радіоелектронних систем факультету Аеронавігації, електроніки та телекомунікацій Національного авіаційного університету, що підтверджено відповідними актами впровадження.

**Особистий внесок здобувача.** Основні положення і наукові результати дисертаційної роботи, які виносяться на захист, отримані автором самостійно та опубліковані в 17-ти наукових працях. У роботах, які опубліковані в співавторстві, особисто здобувачем зроблено наступний науковий внесок: [1], [4], [14] – розроблено метод захищеного голосового радіоуправління функціями БПЛА на основі кепстрального аналізу, а саме мел-частотних кепстральних коефіцієнтів, які формують множину стегано-семантичних ознак; [2], [7], [12] – досліджено основні методи захисту радіоканалу управління БПЛА від несанкціонованого доступу та розроблено систему захищеного радіоуправління БПЛА на основі блочного шифрування ГОСТ 28147-2009, а також побудовано її комп'ютерну модель; [3], [5] – досліджено можливість перехоплення мовної інформації за рахунок побічних електромагнітних випромінювань та виявлення пристроїв несанкціонованого прослуховування; [6], [8], [11], [13] – досліджено сучасні методи спектрального аналізу та розроблено алгоритм стиснення мовних сигналів на основі дискретного вейвлет-перетворення із застосуванням ентропійного арифметичного кодування для зниження смуги пропускання радіоканалу зв'язку голосового управління БПЛА; [9], [10], [15] – досліджено методи знаходження інформативних ознак розпізнавання мови в системі голосового управління БПЛА, а саме мел-частотні кепстральні коефіцієнти та коефіцієнти лінійного передбачення та розроблено метод формування стегано-семантичних ознак; [16], [17] – розроблено формулу та опис патентів на корисну модель.

**Апробація результатів дисертації.** Основні положення та результати дисертаційної роботи доповідалися і обговорювалися на наступних наукових конференціях: 4-й міжнародній IEEE конференції «Methods and Systems of Navigation and Motion Control» (Київ, 2016 р.), 5-й міжнародній IEEE конференції «Methods and Systems of Navigation and Motion Control» (Київ, 2018 р.), 5-й міжнародній IEEE конференції «Actual Problems of Unmanned Aerial Vehicles Developments» (Київ, 2019 р.).

**Публікації.** Основні наукові положення та результати дисертації опубліковано у 17-ти наукових працях, з яких 2 патенти на корисну модель, 10 статей у періодичних фахових виданнях України, 1 стаття у періодичних наукових виданнях держав Європейського Союзу, а також 4 тези доповідей на міжнародних IEEE конференціях, що індексуються наукометричними базами даних Scopus та Web of Science.

**Структура дисертації та її обсяг.** Дисертація складається з анотації, вступу, чотирьох розділів, загальних висновків, додатків, списку використаних джерел, що в загальному обсязі становить 139 сторінок, зокрема 105 сторінок основного тексту, 38 рисунків, 5 таблиць, 26 сторінок додатків та 85 найменувань використаних джерел.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** подано загальну характеристику роботи, обґрунтовано актуальність, сформульовано мету і задачі досліджень, відображено наукову новизну і практичну цінність отриманих результатів, наведено дані щодо їх апробації та впровадження.

У **першому розділі** досліджено відомий метод підвищення ефективності семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів (МЧКК), який полягає в знаходженні середніх значень коефіцієнтів дискретного косинусного перетворення (ДКП)

$$c[n] = \sum_{m=0}^{N_f-1} E[m] \cos \left( \frac{\pi \left( m + \frac{1}{2} \right)}{N_f} \right), \quad n = 0, \dots, N_f - 1, \quad (1)$$

прологарифмованої енергії спектра

$$E[m] = \ln \left( \sum_{k=0}^{N-1} |X[k]|^2 H_m[k] \right), \quad m = 0, \dots, N_f - 1, \quad (2)$$

дискретного перетворення Фур'є (ДПФ)

$$X[k] = \sum_{n=0}^{N-1} x[n] w[n] e^{-\frac{2\pi j}{N} kn}, \quad k = 0, \dots, N-1, \quad (3)$$

обробленого трикутним фільтром

$$H_m[k] = \begin{cases} 0, & k < f[m-1] \\ \frac{(k - f[m-1])}{(f[m] - f[m-1])}, & f[m-1] \leq k < f[m] \\ \frac{(f[m+1] - k)}{(f[m+1] - f[m])}, & f[m] \leq k \leq f[m+1] \\ 0, & k > f[m+1] \end{cases} \quad (4)$$

де,  $f[m] = \left( \frac{N_f}{F_s} \right) M^{-1} \left( M(F_{\min}) + m \frac{M(F_{\max} - F_{\min})}{N_f + 1} \right)$  в мел-шкалі  $M = 1127.01048 \ln(1 + F/700)$ .

*Проблема полягає в тому, що представлений метод семантичного кодування мовних сигналів на основі МЧКК не дотримується умови адаптивності*

$$\bigcup_{n=1}^N \Lambda_n = [0, \pi], \quad (5)$$

де  $\Lambda_n = [\omega_{n-1}, \omega_n]$  – сегменти спектра Фур'є  $[0, \pi]$  досліджуваного мовного сигналу, який розбивається на  $N$  суміжних сегментів з границями  $\omega_n$  (де  $\omega_0 = 0$  і  $\omega_N = \pi$ ).

*Під ефективністю кодування розуміється зниження швидкості передачі інформації із заданою ймовірністю безпомилкового розпізнавання семантичних ознак мовних сигналів, що дозволить значно знизити необхідну смугу пропускання, тим самим підвищуючи пропускну здатність каналу зв'язку.*

*Сформулюємо основну наукову гіпотезу дослідження, яка полягає в тому що підвищити ефективність семантичного кодування мовних сигналів можливо за рахунок використання адаптивного емпіричного вейвлет-перетворення з подальшим застосуванням спектрального аналізу Гільберта.*

**Другий розділ** присвячений обґрунтуванню ефективності використання адаптивного емпіричного вейвлет-перетворення в задачах кратномасштабного аналізу та семантичного кодування мовних сигналів.

В даному науковому дослідженні пропонується застосувати сучасний метод емпіричного вейвлет-перетворення на основі побудови сімейства адаптивних вейвлет-функцій для підвищення ефективності спектрального аналізу мовних сигналів, та подальшого семантичного кодування. Якщо взяти за основу особливості частотного спектра Фур'є, то поставлена задача еквівалентна побудові набору смугових вейвлет-фільтрів. Один із способів досягнення адаптивності – це враховувати, що компактні носії вейвлет-фільтрів напряму залежать від того, де знаходиться потрібна нам семантична інформація в спектрі мовного сигналу, тобто більші амплітуди спектра Фур'є несуть в собі більш важливу інформацію для відновлення функції, а значить, і для якісної оцінки семантичної складової мовного сигналу, а малі амплітуди – менш важливу. Дійсно, властивості функції внутрішньої емпіричної моди, викладені Н. Хуангом, еквівалентні

твердженню, що спектр цієї функції має компактний носій і зосереджений навколо певної частоти (в залежності від сигналу). Для наочності теоретичного викладення сутності даного методу ми розглянемо тільки дійсні періодичні сигнали (їх спектр симетричний щодо частоти  $\omega = 0$ ), а отже, простіше побудувати доказову базу, але наступні міркування можна з легкістю застосовувати і на мовні сигнали, що ми в подальшому і зробимо, побудувавши різні вейвлет-фільтри на позитивних та негативних частотах, відповідно. В процесі викладення матеріалу ми будемо розглядати нормований спектр Фур'є, який має  $2\pi$  періодичність, з метою дотримання критеріїв Шеннона, та обмежимося частотою  $\omega \in [0, \pi]$ .

Почнемо з припущення, що частотний спектр Фур'є  $[0, \pi]$  розбивається на  $N$  суміжних сегментів (пізніше ми обговоримо, як отримати таке розбиття). Визначимо  $\omega_n$  як границі між кожним сегментом (де  $\omega_0 = 0$  і  $\omega_N = \pi$ ), див. Рис. 1.

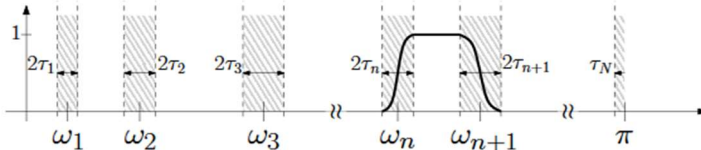


Рис. 1. Розділення спектру Фур'є

Кожен сегмент спектру позначається як  $\Lambda_n = [\omega_{n-1}, \omega_n]$ , тоді очевидно, що  $\bigcup_{n=1}^N \Lambda_n = [0, \pi]$ . У центрі кожного  $\omega_n$ , ми визначаємо перехідну фазу (сірі заштриховані області на Рис. 1)  $T_n$  шириною  $2\tau_n$ . Емпіричні вейвлет-функції визначаються як смугові фільтри для кожного сегмента спектру  $\Lambda_n$ . Для цього ми використовуємо ідею, яка застосовується при побудові вейвлет-функцій Літвуда-Пейли і Мейера. Тоді  $\forall n > 0$ , ми визначаємо емпіричну функцію масштабування і емпіричні вейвлет-функції за допомогою рівнянь (6) і (7) відповідно.

$$\hat{\phi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \leq \omega_n - \tau_n \\ \cos \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\tau_n} (|\omega| - \omega_n + \tau_n) \right) \right] & \text{if } \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

i

$$\hat{\psi}_n(\omega) = \begin{cases} 1 & \text{if } \omega_n + \tau_n \leq |\omega| \leq \omega_{n+1} - \tau_{n+1} \\ \cos \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\tau_{n+1}} (|\omega| - \omega_{n+1} + \tau_{n+1}) \right) \right] & \text{if } \omega_{n+1} - \tau_{n+1} \leq |\omega| \leq \omega_{n+1} + \tau_{n+1} \\ \sin \left[ \frac{\pi}{2} \beta \left( \frac{1}{2\tau_n} (|\omega| - \omega_n + \tau_n) \right) \right] & \text{if } \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Функція  $\beta(x)$  є довільно взятою  $C^k([0,1])$  функцією така, що

$$\beta(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x \geq 1 \end{cases} \text{ и } \beta(x) + \beta(1-x) = 1 \quad \forall x \in [0,1]. \quad (8)$$

Багато різних функцій дотримуються цим властивостям, найбільш часто використовується в літературі наступна функція

$$\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3). \quad (9)$$

Що стосується вибору  $\tau_n$ , то можливо кілька варіантів. Найпростіший – вибрати  $\tau_n$  пропорційно до  $\omega_n$ :  $\tau_n = \gamma\omega_n$  де  $0 < \gamma < 1$ . Отже,  $\forall n > 0$ , рівняння (6) і (7) спрощуються до рівнянь (10) і (11) відповідно

$$\hat{\phi}_n(\omega) = \begin{cases} 1 & \text{if } |\omega| \leq (1-\gamma)\omega_n \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_n}(|\omega| - (1-\gamma)\omega_n)\right)\right] & \text{if } (1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

і

$$\hat{\psi}_n(\omega) = \begin{cases} 1 & \text{if } (1+\gamma)\omega_n \leq |\omega| \leq (1-\gamma)\omega_{n+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_{n+1}}(|\omega| - (1-\gamma)\omega_{n+1})\right)\right] & \text{if } (1-\gamma)\omega_{n+1} \leq |\omega| \leq (1+\gamma)\omega_{n+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\gamma\omega_n}(|\omega| - (1-\gamma)\omega_n)\right)\right] & \text{if } (1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

Приклад емпіричної функції масштабування  $\hat{\phi}_n$  для  $\nu_n=1$ ,  $\gamma=0,5$  та емпіричної вейвлет-функції  $\hat{\psi}_n$  для  $\nu_n=1$ ,  $\nu_{n+1}=2,5$ ,  $\gamma=0,2$  в частотній області наведено на Рис. 2.

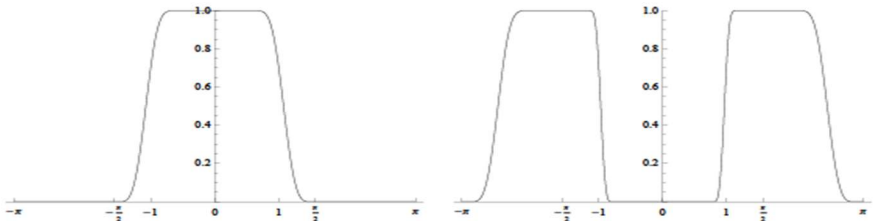


Рис. 2. Зліва: перетворення Фур'є функції масштабування для  $\nu_n=1$ ,  $\gamma=0,5$ .

Справа: перетворення Фур'є вейвлет-функції для  $\nu_n=1$ ,  $\nu_{n+1}=2,5$ ,  $\gamma=0,2$ .

Якісна сегментація спектра Фур'є мовного сигналу, є першочерговою задачею в процесі його семантичного кодування на основі емпіричного вейвлет-перетворення,



оскільки цей крок забезпечує адаптивність запропонованого методу до аналізованого сигналу, що дозволяє більш якісно дослідити частотний спектр. З огляду на вищесказане ми прагнемо розділити спектр Фур'є на різні сегменти, які відповідають функціям внутрішніх емпіричних мод, а також, є центровані навколо певної частоти і компактного носія.

На даному етапі ми припускаємо, що номер сегмента  $N$  відомий і заданий (нижче ми запропонуємо метод оцінки оптимальної кількості сегментів). З цього слідує, що потрібно всього  $N + 1$  границь, але ми вже спочатку маємо в своєму розпорядженні  $0$  і  $\pi$ , тобто 2 границі, отже, нам потрібно знайти  $N - 1$  додаткових границь. Щоб знайти такі границі, потрібно спочатку виявити локальні максимуми в частотному спектрі та відсортувати їх в порядку спадання ( $0$  і  $\pi$  не враховуються). Припустимо, що алгоритм знайшов  $M$  максимумів.

Можуть виникнути два випадки:

1)  $M \geq N$ : алгоритм знайшов достатньо максимумів, щоб визначити бажану кількість сегментів, тоді ми зберігаємо тільки перші  $N - 1$  максимумів;

2)  $M < N$ : сигнал має менше внутрішніх емпіричних мод, ніж очікувалося, тоді ми зберігаємо кількість всіх виявлених максимумів і скидаємо  $N$  на відповідне значення.

Тепер, маючи множину знайдених максимумів, а також  $0$  і  $\pi$ , ми визначаємо границі  $\omega_n$  кожного сегмента як центр між двома послідовними максимумами.

Якщо для простих сигналів є можливість дослідним шляхом визначити оптимальну кількість емпіричних мод (сегментів частотного спектру)  $N$ , то в загальному випадку, як правило ми маємо справу зі складними сигналами, такими як мовні сигнали і т.д., де завжди апріорна інформація про моди досліджуваного сигналу недоступна. У таких випадках повинна бути можливість автоматично оцінити необхідну кількість мод. Загалом, така задача є досить складною, нижче ми наводимо простий спосіб оцінки  $N$ , але для забезпечення більшої ефективності методу необхідно провести поглиблений аналіз цього питання.

Наступне твердження показує, що при відповідному виборі параметра  $\gamma$ , можна отримати щільну структуру кадру.

Твердження 1. Якщо  $\gamma < \min_n \left( \frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n} \right)$ , тоді множина  $\left\{ \phi_1(t), \{\psi_n(t)\}_{n=1}^N \right\}$  є

щільною структурою кадру  $L^2(\mathfrak{R})$ .

Доведення. Ми дотримуємося ідеї побудови вейвлет-функції Мейєра.

Множина  $\left\{ \phi_1(t), \{\psi_n(t)\}_{n=1}^N \right\}$  є щільною структурою кадру, якщо

$$\sum_{k=-\infty}^{+\infty} \left( \left| \hat{\phi}_1(\omega + 2k\pi) \right|^2 + \sum_{n=1}^N \left| \hat{\psi}_n(\omega + 2k\pi) \right|^2 \right) = 1. \quad (11)$$

Згідно з періодичністю  $2\pi$  (див. Рис. 3), достатньо зосередитися на інтервалі  $[0, 2\pi]$

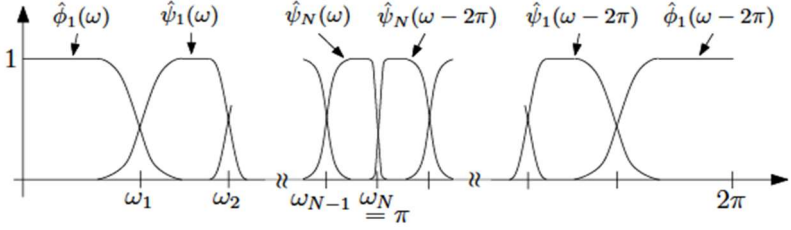


Рис. 3. Періодичність банку фільтрів

Дотримуючись попередніх визначень, ми можемо записати наступний вираз

$$[0, 2\pi] = \bigcup_{n=1}^N \Lambda_n \cup \bigcup_{n=1}^N \Lambda_{\sigma(n)}, \quad (12)$$

де  $\Lambda_{\sigma(n)}$  є копією  $\Lambda_n$  але центрованою на  $2\pi - \nu_n$  замість  $\nu_n$ . По-перше, це легко побачити згідно виразів (13) і (14), тобто що для

$$\omega \in \left( \bigcup_{n=1}^N \Lambda_n \right) \cup \left( \bigcup_{n=1}^N \Lambda_{\sigma(n)} \right), \quad (13)$$

ми маємо

$$|\hat{\phi}_1(\omega)|^2 + |\hat{\phi}_1(\omega - 2\pi)|^2 + \sum_{n=1}^N \left( |\hat{\psi}_n(\omega)|^2 + |\hat{\psi}_n(\omega - 2\pi)|^2 \right) = 1. \quad (14)$$

Потім залишається подивитися на перехідні ділянки. Через властивості  $\beta$ , цей результат також дотримується в  $T_n$ , якщо послідовні  $T_n$  не перекриваються:

$$\tau_n + \tau_{n+1} < \omega_{n+1} - \omega_n, \quad (15)$$

$$\Leftrightarrow \gamma \omega_n + \gamma \omega_{n+1} < \omega_{n+1} - \omega_n, \quad (16)$$

$$\Leftrightarrow \gamma < \frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n}. \quad (17)$$

Умова (17) має виконуватися для всіх  $n$ , а також, для найменшого  $T_n$ , що є рівнозначним, внаслідок, ми отримаємо необхідний результат, якщо

$$\gamma < \min_n \left( \frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n} \right). \quad (18)$$

На Рис. 4 в якості прикладу наведено банк емпіричних вейвлет-фільтрів на основі множини  $\omega_n \in \{0, 1.5, 2, 2.8, \pi\}$  з  $\gamma = 0.05$  (згідно теорії  $\gamma < 0.057$ ).

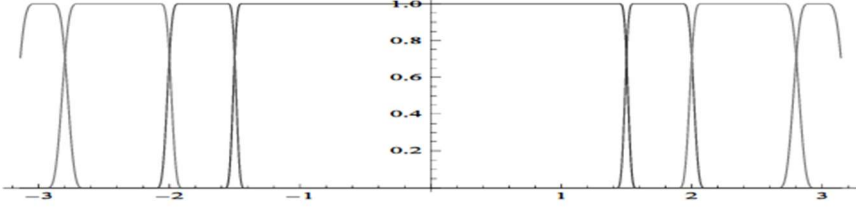


Рис. 4. Приклад розділення спектра Фур'є банкам емпіричних вейвлет-фільтрів

З попереднього підрозділу нам стало відомо, як побудувати множину кадрів емпіричних вейвлет-функцій щільної структури. Тепер ми можемо визначити емпіричне вейвлет-перетворення (ЕВП),  $W_f^\varepsilon(n, t)$ , таким же чином, як і для класичного вейвлет-перетворення. Тоді коефіцієнти деталізації задаються скалярними добутками з емпіричними вейвлет-функціями:

$$W_f^\varepsilon(n, t) = \langle f, \psi_n \rangle = \int f(\tau) \overline{\psi_n(\tau - t)} d\tau = \left( \hat{f}(\omega) \overline{\hat{\psi}_n(\omega)} \right)^\vee, \quad (19)$$

а коефіцієнти апроксимації (позначимо їх наступним чином –  $W_f^\varepsilon(0, t)$ ) скалярним добутком з функцією масштабування:

$$W_f^\varepsilon(0, t) = \langle f, \phi_1 \rangle = \int f(\tau) \overline{\phi_1(\tau - t)} d\tau = \left( \hat{f}(\omega) \overline{\hat{\phi}_1(\omega)} \right)^\vee, \quad (20)$$

де  $\hat{\psi}_n(\omega)$  і  $\hat{\phi}_1(\omega)$  визначаються рівняннями 11 і 10 відповідно.

Тоді реконструкція (обернене ЕВП) вихідного мовного сигналу  $f(t)$  по вейвлет-коефіцієнтам деталізації та апроксимації задається наступним виразом

$$\begin{aligned} f(t) &= W_f^\varepsilon(0, t) \cdot \phi_1(t) + \sum_{n=1}^N W_f^\varepsilon(n, t) \cdot \psi_n(t) \\ &= \left( \hat{W}_f^\varepsilon(0, \omega) \cdot \hat{\phi}_1(\omega) + \sum_{n=1}^N \hat{W}_f^\varepsilon(n, \omega) \cdot \hat{\psi}_n(\omega) \right)^\vee. \end{aligned} \quad (21)$$

Викладені вище твердження доводять ефективність використання емпіричного вейвлет-перетворення в задачах спектрального аналізу мовних сигналів, що дозволить підвищити ефективність їх семантичного кодування за рахунок збереження адаптивності до досліджуваного сигналу.

У **третьому розділі** розроблено метод семантичного кодування мовних сигналів на основі адаптивного емпіричного вейвлет-перетворення з подальшим застосуванням спектрального аналізу Гільберта.

Згідно розробленого методу (див. Рис. 5) на його вхід подається мовний сигнал частотний діапазон, якого вельми обмежений і розташовується в інтервалі від 300 до 3400 Гц. З даного факту випливає, що шляхом моделювання смугового фільтра можна відкинути частотні складові, які знаходяться за межами цього діапазону і відповідно, не несуть семантичного навантаження.



Рис. 5. Метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення та спектрального аналізу Гільберта

Як відомо мовні сигнали є нестационарними сигналами складної форми, параметри і характеристики яких, як правило, швидко змінюються протягом часу. Сформований підхід до процедури обробки мовних сигналів полягає у використанні короткочасного аналізу. Тобто сигнал розбивається на часові кадри фіксованого розміру, на яких параметри сигналу не змінюються. Для того щоб отримати набір семантичних ознак однакової довжини, потрібно розбити мовної сигнал на рівні кадри, а потім виконувати перетворення, вважаючи, що сигнал на такому відрізьку приблизно стаціонарний (див. Рис. 6).

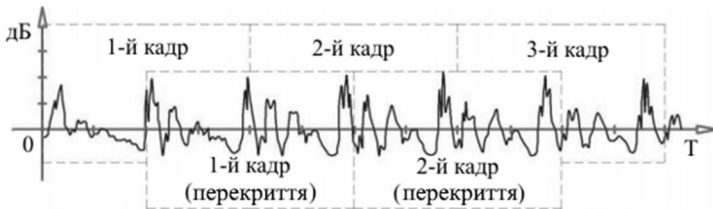


Рис. 6. Поділ мовного сигналу на кадри

Для мовного сигналу розмір кадру зазвичай вибирається в межах 10-20 мс. Для більш точного представлення сигналу між кадрами роблять перекриття, рівну половині довжини кадру. Перекриття кадрів використовується для запобігання втрати семантичної інформації про сигнал на границі кадрів. Чим менше перекриття, тим меншою розмірністю в результаті буде мати набір ознак, характерний для даної ділянки мовного сигналу. Потім до кожного кадру застосовується алгоритм виділення семантичної складової.

Виходячи із вищесказаного, мовний сигнал що пройшов попередню обробку розбивається на  $K$  кадрів по  $N$  відліків, що перетинаються на  $1/2$  довжини кадру.

На вхід блоку дискретного перетворення Фурє (ДПФ) подається послідовність відліків длянки мовного сигналу ( $K$ -го кадру), досліджуваного на даній ітерації,  $x_0, \dots, x_{N-1}$ . До даної послідовності застосовується вагова функція і потім ДПФ.

Вагова функція використовується для зменшення спотворень в Фурє аналізі, викликаних скінченністю вибірки.

На практиці в якості вагової функції часто використовується вікно Хеммінга, яке має наступний вигляд:

$$w[n] = 0.53836 - 0.46164 \cdot \cos\left(2\pi \frac{n}{N-1}\right), \quad n = 0, \dots, N-1, \quad (22)$$

де,  $N$  – довжина вікна, виражена у відліках.

Тоді ДПФ зваженого мовного сигналу можна записати у вигляді формули:

$$X[k] = \sum_{n=0}^{N-1} x[n] w[n] e^{-\frac{2\pi j}{N} kn}, \quad k = 0, \dots, N-1. \quad (23)$$

Значення індексів  $k$  відповідають частотам:

$$f[k] = \frac{F_s}{N} k, \quad k = 0, \dots, N/2, \quad (24)$$

де,  $F_s$  – частота дискретизації мовного сигналу.

Ми дотримуємося тієї ідеї, що найбільш важлива інформація щодо оцінки семантики мови зберігається в максимальних амплітудах (максимуми) спектра Фур'є вихідного сигналу (що відповідають центру кожного з  $N$  сегментів Фур'є), які значно перевищують інші, існуючі в спектрі максимуми. Визначимо множину  $M$  знайдених максимумів амплітуд спектра Фур'є через  $\{M_i\}_{i=1}^M$ . Припустимо, що ця множина відсортована в порядку спадання значень ( $M_1 \geq M_2 \geq \dots \geq M_M$ ) і нормована відповідно до  $[0;1]$ . У цьому випадку, вищезгадана ідея рівносильна збереженню всіх амплітудних максимумів частотного спектра, які перевищують певну величину різниці між більшим і меншим максимумами. Можна формалізувати це наступним чином: потрібно зберегти всі амплітудні максимуми спектра Фур'є, які більше заданого порогу  $M_M + \alpha(M_1 - M_M)$ , де

$\alpha$  відповідає відносному співвідношенню амплітуд. Завдання полягає в тому, щоб підібрати таке  $\alpha$ , яке б призводило до компромісу між занадто частим виявленням так званих «хибних максимумів», що не несуть важливої інформації і якісним поділом спектра Фур'є на сегменти, які відповідають емпіричним модам мовного сигналу. В результаті значення порога безпосередньо впливатиме на кількість знайдених максимумів і смуг сегментації спектра Фур'є, а останні відповідно на кількість мод, на які розкладається досліджуваний мовний сигнал.

Слідуючи цьому формалізму, емпірична мода  $f_k$ , визначена Н. Хуангом, як кінцева сума  $N+1$  функцій внутрішнього режиму  $f_k(t)$  з амплітудною і частотною модуляціями, які можна записати у вигляді

$$f_k(t) = F_k(t) \cos(\phi_k(t)), \quad (25)$$

де  $F_k(t), \varphi_k^i(t) > 0 \forall t$ , та є такими, що

$$f(t) = \sum_{k=0}^N f_k(t), \quad (26)$$

задається формулами

$$f_0(t) = W_f^{\varepsilon}(0, t) \cdot \phi_1(t), \quad (27)$$

$$f_k(t) = W_f^{\varepsilon}(k, t) \cdot \psi_k(t). \quad (28)$$

Як уже згадувалося вище, ЕВП представляє собою інструмент для частотно-часового аналізу нестационарних і нелінійних сигналів, якими і є мовні сигнали. Один із способів виразити нестационарний характер мовних даних – це визначення миттєвої частоти та амплітуди досліджуваного сигналу. Перетворення Гільберта (ПГ) сигналу  $x(t)$  задається наступним виразом

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau, \quad (29)$$

де  $P$  – головне значення Коші сингулярного інтеграла.

За допомогою ПГ сигналу  $x(t)$  можна одержати аналітичний сигнал

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)}, \quad (30)$$

де  $i = (-1)^{1/2}$ .

Тоді  $a(t)$  можна виразити як

$$a(t) = \sqrt{(x^2 + y^2)}, \quad (31)$$

де  $a(t)$  – миттєва амплітуда.

Миттєва фазова функція може бути виражена наступним чином

$$\theta(t) = \arctan \frac{y}{x}. \quad (32)$$

Миттєва частота визначається виразом

$$\omega = \frac{d\theta}{dt}. \quad (33)$$

Застосовуючи ПГ до окремих компонентів внутрішніх емпіричних мод, вихідні дані можуть бути виражені як дійсна частина наступного рівняння

$$x(t) = \operatorname{Re} \left\{ \sum_{j=1}^n a_j(t) \exp \left[ i \int \omega_j(t) dt \right] \right\}. \quad (34)$$

Рівняння (34) визначає дійсну частину амплітуди (31) і частоти (33) кожного компонента внутрішніх емпіричних мод як функцію часу. Аналіз сигналів в частотно-часовій області може бути виражений у вигляді енергетичного спектра Гільберта (ЕСГ) або амплітудного спектра Гільберта (АСГ). ЕСГ визначається як розподіл щільності енергії, а також, аналогічно визначається АСГ як розподіл щільності амплітуди в частотно-часовому просторі.

Спектр щільності енергії Гільберта (СЩЕГ) визначається як

$$S_{i,j} = H(t_i, \omega_j) = \frac{1}{\Delta t \cdot \Delta \omega} H \left[ \sum_{k=1}^n a_k^2(t) \right]. \quad (35)$$

Роздільна здатність спектра Гільберта задається інтервалами однакового розміру  $\Delta t \cdot \Delta \omega$ . Кожен інтервал представляє собою значення  $a^2(t)$  в даний момент часу і частоти.

Це перетворення має властивість компактності енергії: більший енергії відповідає менша кількість інформації. Таким чином, ми маємо дуже невеликий набір семантичних ознак, який при кодуванні успішно замінює тисячі відліків мовного сигналу, які саме і відповідають семантичній формі мовного сигналу.

**Четвертий розділ** присвячено експериментальним дослідженням розробленого методу семантичного кодування мовних сигналів на основі адаптивного емпіричного вейвлет-перетворення з подальшим використанням спектрального аналізу Гільберта та знаходження миттєвої частоти і амплітуди мовного сигналу.

В роботі було досліджено розроблений метод семантичного кодування мовних сигналів на основі ЕВП і проведено моделювання в програмному пакеті MATLAB, зокрема було здійснено оцінювання коефіцієнта стиснення (КС), швидкості передачі бітів (ШПБ), коефіцієнта кореляції (КК), відношення сигнал/шум (ВСШ), пікового відношення сигнал/шум (ПВСШ) і середньоквадратичної помилки (СКП), а також ймовірності безпомилкового розпізнавання семантичних ознак, які виступають головними показниками ефективності запропонованого методу.

Таблиця 1

Результати оцінки ефективності розробленого методу семантичного кодування мовних сигналів на прикладі голосових команд

Навчання	Тестування			
	«Вгору»	«Вниз»	«Вправо»	«Вліво»
Голосові команди				
«Вгору»	КС = 333 ШПБ = 192 біт/с КК = 0.95 ВСШ = 21 дБ ПВСШ = 38 дБ СКП = 0.04	КС = 333 ШПБ = 192 біт/с КК = 0.32 ВСШ = 6 дБ ПВСШ = 11 дБ СКП = 0.38	КС = 333 ШПБ = 192 біт/с КК = 0.12 ВСШ = 4 дБ ПВСШ = 9 дБ СКП = 0.52	КС = 333 ШПБ = 192 біт/с КК = 0.41 ВСШ = 7 дБ ПВСШ = 15 дБ СКП = 0.45
«Вниз»	КС = 333 ШПБ = 192 біт/с КК = 0.37 ВСШ = 8 дБ ПВСШ = 14 дБ СКП = 0.33	КС = 333 ШПБ = 192 біт/с КК = 0.94 ВСШ = 23 дБ ПВСШ = 41 дБ СКП = 0.09	КС = 333 ШПБ = 192 біт/с КК = 0.22 ВСШ = 3 дБ ПВСШ = 7 дБ СКП = 0.38	КС = 333 ШПБ = 192 біт/с КК = 0.34 ВСШ = 6 дБ ПВСШ = 13 дБ СКП = 0.43
«Вправо»	КС = 333 ШПБ = 192 біт/с КК = 0.14 ВСШ = 9 дБ ПВСШ = 16 дБ СКП = 0.52	КС = 333 ШПБ = 192 біт/с КК = 0.27 ВСШ = 7 дБ ПВСШ = 11 дБ СКП = 0.39	КС = 333 ШПБ = 192 біт/с КК = 0.97 ВСШ = 26 дБ ПВСШ = 43 дБ СКП = 0.06	КС = 333 ШПБ = 192 біт/с КК = 0.18 ВСШ = 5 дБ ПВСШ = 14 дБ СКП = 0.67
«Вліво»	КС = 333 ШПБ = 192 біт/с КК = 0.45 ВСШ = 3 дБ ПВСШ = 9 дБ СКП = 0.47	КС = 333 ШПБ = 192 біт/с КК = 0.38 ВСШ = 4 дБ ПВСШ = 10 дБ СКП = 0.42	КС = 333 ШПБ = 192 біт/с КК = 0.24 ВСШ = 8 дБ ПВСШ = 14 дБ СКП = 0.27	КС = 333 ШПБ = 192 біт/с КК = 0.91 ВСШ = 22 дБ ПВСШ = 38 дБ СКП = 0.1

Результати науково-експериментальних досліджень підвищення ефективності семантичного кодування мовних сигналів викладені нижче (Табл. 1). В даному експерименті було оцінено КС, ШПБ, КК, ВСШ, ПВСШ і СКП для двох реалізацій семантичних ознак голосових команд знайдених на основі ЕВП та спектрального аналізу Гілберта.

На Рис. 7 зображені семантичні ознаки голосових команд на основі ЕВП та спектрального аналізу Гілберта.

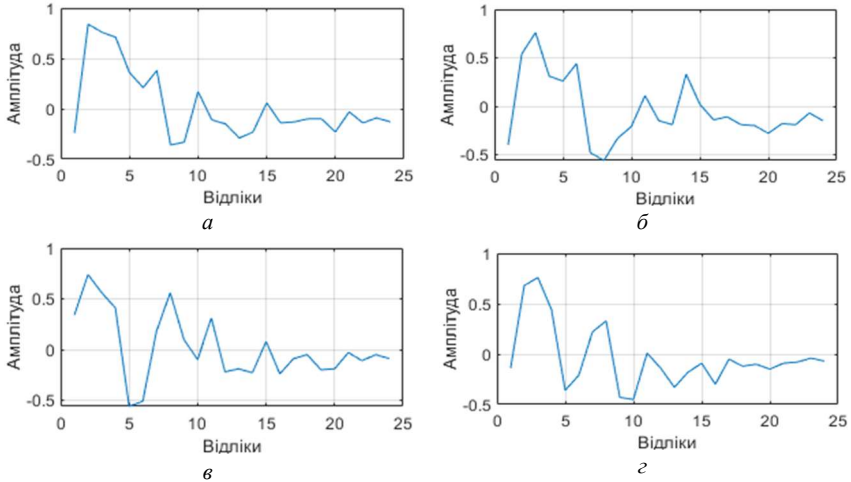


Рис. 7. Семантичні ознаки на основі ЕВП голосових команд: «Вгору» (а), «Вниз» (б), «Вправо» (в), «Вліво» (г)

З наведених вище результатів дослідження (Табл. 1) наглядно видно, що оптимальним рішенням за заданими критеріями ефективності семантичного кодування для голосових команд є: «Вгору» – КС = 333, ШПБ = 192 біт/с, КК = 0.95, ВСШ = 21 дБ, ПВСШ = 38 дБ, СКП = 0.04; «Вниз» – КС = 333, ШПБ = 192 біт/с, КК = 0.94, ВСШ = 23 дБ, ПВСШ = 41 дБ, СКП = 0.09; «Вправо» – КС = 333, ШПБ = 192 біт/с, КК = 0.97, ВСШ = 26 дБ, ПВСШ = 43 дБ, СКП = 0.06; «Вліво» – КС = 333, ШПБ = 192 біт/с, КК = 0.91, ВСШ = 22 дБ, ПВСШ = 38 дБ, СКП = 0.1. Що показує досить високі результати, зберігаючи семантичні ознаки голосових команд знайдених на основі ЕВП та спектрального аналізу Гілберта. Це забезпечує можливість семантичної ідентифікації мовних сигналів.

Результати наступного науково-експериментального дослідження оцінки підвищення ефективності семантичного кодування мовних сигналів за допомогою розробленого методу викладені нижче (Табл. 2). В даному експерименті було оцінено середні значення КС, ШПБ, КК, ВСШ, ПВСШ і СКП для ста реалізацій семантичних ознак голосових команд знайдених на основі ЕВП та спектрального аналізу Гілберта.

З наведених вище результатів дослідження (Табл. 2) наглядно видно, що зі збільшенням реалізацій показники ефективності семантичного кодування залишаються на високому рівні, де семантична складова мовних сигналів зберігає свої семантичні закономірності, що робить даний метод стійким до нестационарних та нелінійних процесів. Також даний факт підтверджується на Рис 8, де зображено сто реалізацій семантичних ознак знайдених на основі ЕВП та спектрального аналізу Гілберта голосових команд: «Вгору», «Вниз», «Вправо», «Вліво».



Таблиця 2

Результати оцінки ефективності розробленого методу семантичного кодування мовних сигналів на прикладі голосових команд

Навчання	Тестування			
Голосові команди	«Вгору»	«Вниз»	«Вправо»	«Вліво»
«Вгору»	КС = 333 ШПБ = 192 біт/с КК = 0.92 ВСШ = 19 дБ ПВСШ = 34 дБ СКП = 0.07	КС = 333 ШПБ = 192 біт/с КК = 0.29 ВСШ = 5 дБ ПВСШ = 9 дБ СКП = 0.44	КС = 333 ШПБ = 192 біт/с КК = 0.09 ВСШ = 3 дБ ПВСШ = 7 дБ СКП = 0.56	КС = 333 ШПБ = 192 біт/с КК = 0.32 ВСШ = 5 дБ ПВСШ = 12 дБ СКП = 0.51
«Вниз»	КС = 333 ШПБ = 192 біт/с КК = 0.35 ВСШ = 6 дБ ПВСШ = 13 дБ СКП = 0.44	КС = 333 ШПБ = 192 біт/с КК = 0.93 ВСШ = 21 дБ ПВСШ = 37 дБ СКП = 0.12	КС = 333 ШПБ = 192 біт/с КК = 0.09 ВСШ = 2 дБ ПВСШ = 6 дБ СКП = 0.46	КС = 333 ШПБ = 192 біт/с КК = 0.23 ВСШ = 4 дБ ПВСШ = 7 дБ СКП = 0.44
«Вправо»	КС = 333 ШПБ = 192 біт/с КК = 0.11 ВСШ = 5 дБ ПВСШ = 14 дБ СКП = 0.60	КС = 333 ШПБ = 192 біт/с КК = 0.21 ВСШ = 5 дБ ПВСШ = 9 дБ СКП = 0.47	КС = 333 ШПБ = 192 біт/с КК = 0.95 ВСШ = 22 дБ ПВСШ = 39 дБ СКП = 0.08	КС = 333 ШПБ = 192 біт/с КК = 0.15 ВСШ = 4 дБ ПВСШ = 9 дБ СКП = 0.74
«Вліво»	КС = 333 ШПБ = 192 біт/с КК = 0.24 ВСШ = 2 дБ ПВСШ = 5 дБ СКП = 0.56	КС = 333 ШПБ = 192 біт/с КК = 0.28 ВСШ = 3 дБ ПВСШ = 7 дБ СКП = 0.49	КС = 333 ШПБ = 192 біт/с КК = 0.20 ВСШ = 5 дБ ПВСШ = 9 дБ СКП = 0.38	КС = 333 ШПБ = 192 біт/с КК = 0.90 ВСШ = 19 дБ ПВСШ = 33 дБ СКП = 0.14

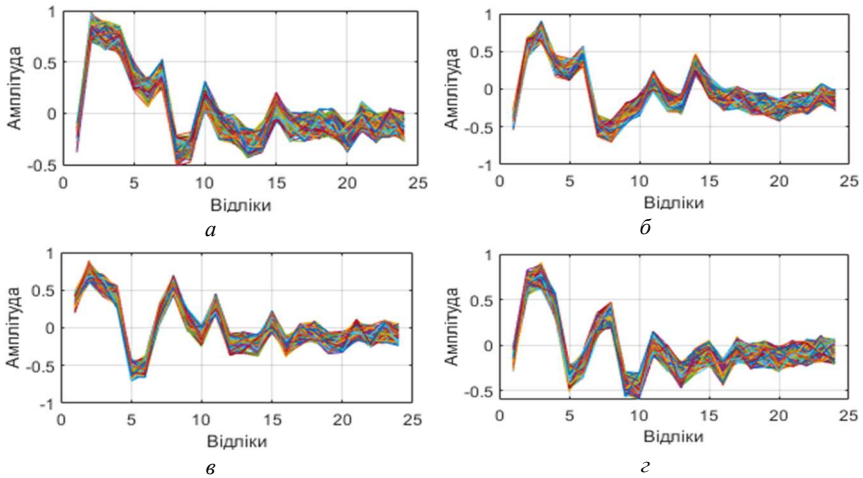


Рис. 8. Сто реалізацій семантичних ознак на основі ЕВП голосових команд: «Вгору» (а), «Вниз» (б), «Вправо» (в), «Вліво» (г)

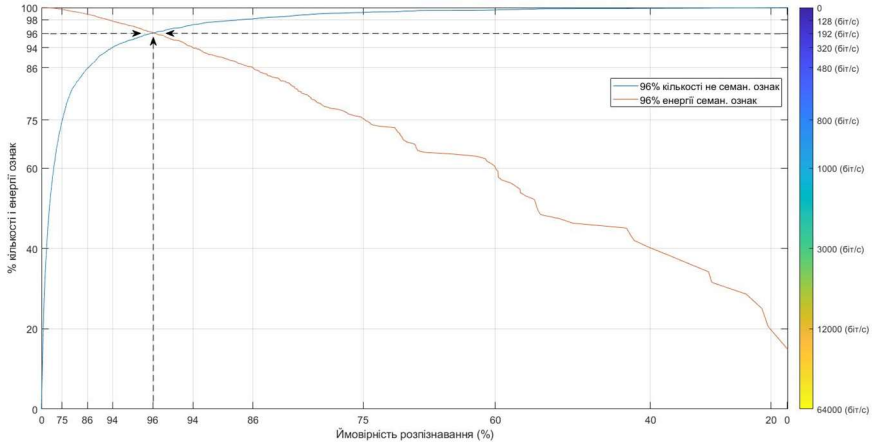


Рис. 9. Залежність ймовірності розпізнавання семантичних ознак від скорочення їх енергії і швидкості кодування за допомогою запропонованого методу

Експериментальне дослідження показало, що розроблений метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення дозволяє знизити швидкість кодування від 320 до 192 біт/с та необхідну смугу пропускання від 40 до 24 Гц з ймовірністю безпомилкового розпізнавання близько 0.96 (96%) і відношенням сигнал/шум 48 дБ, згідно чого його ефективність підвищується в 1.6 рази на відміну від існуючого методу, не перевищуючи при цьому пограничне значення затримки обробки і передачі даних в 300 мс, це дозволить системі функціонувати в режимі реального часу.

У **додатках** вміщено акти впровадження результатів дисертаційної роботи і лістинги (фрагменти кодів) розробленого у роботі програмного забезпечення.

## ВИСНОВКИ

Результатом виконаної роботи є розв'язання актуальної науково-практичної задачі розробки і дослідження нових ефективних методів семантичного кодування мовних сигналів.

У процесі виконання дисертаційної роботи отримані такі наукові результати:

1. Досліджено відомий метод семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів, який не дотримується умови адаптивності до досліджуваного сигналу, що являється суттєвим недоліком існуючого методу.

2. Розроблено метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення, який будує множини адаптивних смугових вейвлет-фільтрів Мейера з подальшим застосуванням спектрального аналізу Гільберта для знаходження миттєвих амплітуд і частот функцій внутрішніх емпіричних мод, що дозволить визначити семантичні ознаки мовних сигналів та підвищити ефективність їх кодування.

3. Запропоновано використовувати метод адаптивного емпіричного вейвлет-перетворення в задачах кратномасштабного аналізу та семантичного кодування мовних сигналів, що дозволить підвищити ефективність спектрального аналізу за рахунок розкладання високочастотного мовного коливання на його низькочастотні складові, а саме внутрішні емпіричні моди.

4. Досліджено метод семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів, але з використанням базових принципів адаптивного спектрального аналізу за допомогою емпіричного вейвлет-перетворення, що підвищує ефективність даного методу щонайменше в 1.3 рази.

5. Розроблений метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення дозволяє знизити швидкість кодування від 320 до 192 біт/с та необхідну смугу пропускання від 40 до 24 Гц з ймовірністю безпомилкового розпізнавання близько 0.96 (96%) і відношенням сигнал/шум 48 дБ, згідно чого його ефективність підвищується в 1.6 рази на відміну від існуючого методу, не перевищуючи при цьому пограничне значення затримки обробки і передачі даних в 300 мс, це дозволить системі функціонувати в режимі реального часу.

6. Розроблено алгоритм семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення та його програмна реалізація мовою програмування MATLAB R2020b.

### ПУБЛІКАЦІ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. O. Lavrynenko, G. Konakhovych, D. Bakhtiiarov, «Method of voice control functions of the UAV», *Proc. IEEE 4th Int. Conf. on Methods and Systems of Navigation and Motion Control (MSNMC)*, Kyiv, Oct. 18-20, 2016, pp. 47-50. (*Scopus, Web of Science*)

2. D. Bakhtiiarov, G. Konakhovych, O. Lavrynenko, «Protected system of radio control of unmanned aerial vehicle», *Proc. IEEE 4th Int. Conf. on Methods and Systems of Navigation and Motion Control (MSNMC)*, Kyiv, Oct. 18-20, 2016, pp. 196-199. (*Scopus, Web of Science*)

3. D.I. Bakhtiiarov, G.F. Konakhovych, O.Y. Lavrynenko, «An approach to modernization of the Hat and COST 231 model for improvement of electromagnetic compatibility in premises for navigation and motion control equipment», *Proc. IEEE 5th Int. Conf. on Methods and Systems of Navigation and Motion Control (MSNMC)*, Kyiv, Oct. 16-18, 2018, pp. 271-274. (*Scopus, Web of Science*)

4. O. Lavrynenko, A. Taranenko, I. Machalin, Ye. Gabrousenko, I. Terentyeva, D. Bakhtiiarov, «Protected voice control system of UAV», *Proc. IEEE 5th Int. Conf. on Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)*, Kyiv, Oct. 22-24, 2019, pp. 295-298. (*Scopus, Web of Science*)

5. D. Bakhtiiarov, O. Lavrynenko, N. Lishchynovska, I. Basiuk, T. Prykhodko, «Methods for assessment and forecasting of electromagnetic radiation levels in urban environments», *Informatics, Control, Measurement in Economy and Environmental Protection*, vol. 11, no. 1, pp. 24-27, 2021. (*Poland*)

6. G.F. Konakhovych, O.Y. Lavrynenko, V.V. Antonov, D.I. Bakhtiiarov, «A digital speech signal compression algorithm based on wavelet transform», *Electronics and Control Systems*, vol. 48, no. 2, pp. 30-36, 2016.

7. I.O. Kozliuk, D.I. Bakhtiiarov, O.Y. Lavrynenko, I.V. Tretiak, «Problems of unauthorized interference to the work of UAV and methods of its solving», *Science-Based Technologies*, vol. 30, no. 2, pp. 206-211, 2016.

8. O.Yu. Lavrynenko, G.F. Konakhovych, D.I. Bakhtiiarov, «Compression algorithm of voice control commands of UAV based on wavelet transform», *Electronics and Control Systems*, vol. 55, no. 1, pp. 17-22, 2018.

9. O. Lavrynenko, G. Konakhovych, D. Bakhtiiarov, «Comparative analysis of speech recognition algorithms in UAV voice control system», *Science-Based Technologies*, vol. 38, no. 2, pp. 137-142, 2018.

10. О.Ю. Лавриненко, Y.A. Kocherhin, G.F. Konakhovych, «Voice control command recognition system of UAV based on steganographic-cepstral analysis», *Electronics and Control Systems*, vol. 56, no. 2, pp. 11-17, 2018.

11. Г.Ф. Коначович, О.І. Давлет'янц, О.Ю. Лавриненко, Д.І. Бахтіяров, «Порівняльний аналіз перетворення Фур'є, косинусного перетворення та вейвлет-перетворення як спектрального аналізу цифрових мовних сигналів», *Наукоємні технології*, Т. 27, № 3, С. 210-220, 2015.

12. Г.Ф. Коначович, Д.І. Бахтіяров, О.Ю. Лавриненко, «Комп'ютерне моделювання захищеного каналу керування безпілотним літальним апаратом», *Наукоємні технології*, Т. 28, № 4, С. 283-290, 2015.

13. А.Ю. Лавриненко, Г.Ф. Коначович, Р.С. Одарченко, Д.І. Бахтіяров, «Алгоритм сжатия сигналов речевых команд управления функциями беспилотного летательного аппарата», *Авиационно-космическая техника и технология*, Т. 130, № 3, С. 51-61, 2016.

14. А.Ю. Лавриненко, Ю.А. Кочергин, Г.Ф. Коначович, «Система защищённого голосового управления беспилотным летательным аппаратом», *Авиационно-космическая техника и технология*, Т. 148, № 4, С. 4-11, 2018.

15. А.Ю. Лавриненко, Ю.А. Кочергин, Г.Ф. Коначович, «Система распознавания стеганографически-преобразованных голосовых команд управления БПЛА», *Радиоэлектронные и компьютерные системы*, Т. 87, № 3, С. 20-28, 2018.

16. Г.Ф. Коначович, О.Ю. Лавриненко, Р.С. Одарченко, В.М. Чуприн, «Пристрій для голосового радіоуправління функціями безпілотного літального апарата з вбудованим захистом від несанкціонованого доступу», *Патент № 116514 України на корисну модель*, МПК G10L 13/00, заявник та патентовласник Національний авіаційний університет, № u 2016 12187, заявл. 01.12.2016, опубл. 25.05.2017, бюл. № 10, 5 с.

17. О.Ю. Лавриненко, Г.Ф. Коначович, Р.С. Одарченко, Д.І. Бахтіяров, «Пристрій захищеного голосового радіоуправління функціями безпілотного літального апарата», *Патент № 119441 України на корисну модель*, МПК G10L 13/00, заявник та патентовласник Національний авіаційний університет, № u 2017 03401, заявл. 10.04.2017, опубл. 25.09.2017, бюл. № 18, 6 с.

#### АНОТАЦІЯ

*Лавриненко О.Ю.* Методи підвищення ефективності семантичного кодування мовних сигналів. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.12.02 – «Телекомунікаційні системи та мережі». – Національний авіаційний університет, Київ, 2021.

Дисертаційна робота присвячена вирішенню актуальної науково-практичної проблеми в телекомунікаційних системах, а саме підвищення пропускної здатності каналу передачі семантичних мовних даних за рахунок ефективного їх кодування, тобто формується питання підвищення ефективності семантичного кодування, а саме – з якою мінімальною швидкістю можливо кодувати семантичні ознаки мовних сигналів із заданою ймовірністю безпомилкового їх розпізнавання? Саме на це питання буде дана відповідь у даному науковому дослідженні, що є актуальною науково-технічною задачею враховуючи зростаючу тенденцію дистанційної взаємодії людей і роботизованої техніки за допомогою мови, де безпомилковість функціонування даного типу систем безпосередньо залежить від ефективності семантичного кодування мовних сигналів. У роботі досліджено відомий метод підвищення ефективності семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів, який полягає в знаходженні середніх значень коефіцієнтів дискретного косинусного перетворення прологарифмованої енергії спектра дискретного перетворення Фур'є обробленого

трикутним фільтром в мел-шкалі. Проблема полягає в тому, що представлений метод семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів не дотримується умови адаптивності, тому було сформульовано основну наукову гіпотезу дослідження, яка полягає в тому що підвищити ефективність семантичного кодування мовних сигналів можливо за рахунок використання адаптивного емпіричного вейвлет-перетворення з подальшим застосуванням спектрального аналізу Гільберта. Під ефективністю кодування розуміється зниження швидкості передачі інформації із заданою ймовірністю безпомилкового розпізнавання семантичних ознак мовних сигналів, що дозволить значно знизити необхідну смугу пропускання, тим самим підвищуючи пропускну здатність каналу зв'язку. У процесі доведення сформульованої наукової гіпотези дослідження були отримані наступні результати: 1) вперше розроблено метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення, який відрізняється від існуючих методів побудовою множини адаптивних смугових вейвлет-фільтрів Мейєра з подальшим застосуванням спектрального аналізу Гільберта для знаходження миттєвих амплітуд і частот функцій внутрішніх емпіричних мод, що дозволить визначити семантичні ознаки мовних сигналів та підвищити ефективність їх кодування; 2) вперше запропоновано використовувати метод адаптивного емпіричного вейвлет-перетворення в задачах кратномасштабного аналізу та семантичного кодування мовних сигналів, що дозволить підвищити ефективність спектрального аналізу за рахунок розкладання височастотного мовного коливання на його низькочастотні складові, а саме внутрішні емпіричні моди; 3) отримав подальший розвиток метод семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів, але з використанням базових принципів адаптивного спектрального аналізу за допомогою емпіричного вейвлет-перетворення, що підвищує ефективність даного методу.

**Ключові слова:** мовні сигнали, семантичне кодування мовних сигналів, семантичні ознаки мовних сигналів, мел-частотні кепстральні коефіцієнти, адаптивний спектральний аналіз, кратномасштабний аналіз, емпіричне вейвлет-перетворення, адаптивні вейвлет-фільтри Мейєра, функції внутрішніх емпіричних мод, спектральний аналіз Гільберта, розпізнавання і синтез мови, голосове управління, низькошвидкісне кодування мови, голосовий переклад з іноземних мов.

#### ABSTRACT

*O. Lavrynenko. Methods of increasing the efficiency of semantic coding of speech signals.*  
– Manuscript.

Thesis for a degree of Candidate of Technical Science in specialty 05.12.02 – «Telecommunication Systems and Networks». – National Aviation University, Kyiv, 2021.

The thesis is devoted to the solution of the actual scientific and practical problem in telecommunication systems, namely increasing the bandwidth of the semantic speech data transmission channel due to their efficient coding, that is the question of increasing the efficiency of semantic coding is formulated, namely – at what minimum speed it is possible to encode semantic features of speech signals with the set probability of their error-free recognition? It is on this question will be answered in this research, which is an urgent scientific and technical task given the growing trend of remote human interaction and robotic technology through speech, where the accurateness of this type of system directly depends on the effectiveness of semantic coding of speech signals. In the thesis the well-known method of increasing the efficiency of semantic coding of speech signals based on mel-frequency cepstral coefficients is investigated, which consists in finding the average values of the coefficients of the discrete cosine transformation of the prologarithmic energy of the spectrum of the discrete Fourier transform treated by a triangular filter in the mel-scale. The problem is that the presented method of

semantic coding of speech signals based on mel-frequency cepstral coefficients does not meet the condition of adaptability, therefore the main scientific hypothesis of the study was formulated, which is that to increase the efficiency of semantic coding of speech signals is possible through the use of adaptive empirical wavelet transform followed by the use of Hilbert spectral analysis. Coding efficiency means a decrease in the rate of information transmission with a given probability of error-free recognition of semantic features of speech signals, which will significantly reduce the required passband, thereby increasing the bandwidth of the communication channel. In the process of proving the formulated scientific hypothesis of the study, the following results were obtained: 1) the first time the method of semantic coding of speech signals based on empirical wavelet transform is developed, which differs from existing methods by constructing a sets of adaptive bandpass wavelet-filters Meyer followed by the use of Hilbert spectral analysis for finding instantaneous amplitudes and frequencies of the functions of internal empirical modes, which will determine the semantic features of speech signals and increase the efficiency of their coding; 2) the first time it is proposed to use the method of adaptive empirical wavelet transform in problems of multiscale analysis and semantic coding of speech signals, which will increase the efficiency of spectral analysis due to the decomposition of high-frequency speech oscillations into its low-frequency components, namely internal empirical modes; 3) received further development the method of semantic coding of speech signals based on mel-frequency cepstral coefficients, but using the basic principles of adaptive spectral analysis with the application empirical wavelet transform, which increases the efficiency of this method. Conducted experimental research in the software environment MATLAB R2020b showed, that the developed method of semantic coding of speech signals based on empirical wavelet transform allows you to reduce the encoding speed from 320 to 192 bit/s and the required passband from 40 to 24 Hz with a probability of error-free recognition of about 0.96 (96%) and a signal-to-noise ratio of 48 dB, according to which its efficiency increases 1.6 times in contrast to the existing method. The results obtained in the thesis can be used to build systems for remote interaction of people and robotic equipment using speech technologies, such as speech recognition and synthesis, voice control of technical objects, low-speed encoding of speech information, voice translation from foreign languages, etc.

**Keywords:** speech signals, semantic coding of speech signals, semantic features of speech signals, mel-frequency cepstral coefficients, adaptive spectral analysis, multiscale analysis, empirical wavelet transform, adaptive wavelet-filters Meyer, functions of internal empirical modes, Hilbert spectral analysis, speech recognition and synthesis, voice control, low-speed speech encoding, voice translation from foreign languages.