

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Кваліфікаційна наукова
праця на правах рукопису

Лавриненко Олександр Юрійович

УДК 621.391

ДИСЕРТАЦІЯ

МЕТОДИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ СЕМАНТИЧНОГО
КОДУВАННЯ МОВНИХ СИГНАЛІВ

05.12.02 – «Телекомунікаційні системи та мережі»

Подається на здобуття наукового ступеня
кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей,
результатів і текстів інших авторів мають посилання на відповідне джерело

Науковий керівник: Конахович Георгій Філімонович
доктор технічних наук, професор, професор кафедри
телекомунікаційних та радіоелектронних систем
Національного авіаційного університету

Київ – 2021

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 формулюється 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 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 кепстральних коефіцієнтів, але з використанням базових принципів адаптивного спектрального аналізу за допомогою емпіричного вейвлет-перетворення, що підвищує ефективність даного методу. Доведено, що семантична складова мовного сигналу закодована в акустичному коливанні вельми неефективним чином, однак виділити її і перекодувати оптимальним методом є досить нетривіальною задачею на рішення якої і направлено дане наукове дослідження. Справа полягає в тому, що для семантичного розпізнавання нам не потрібні акустичні характеристики мови, тобто можна обійтися розпізнаванням фонем мовного повідомлення, що в свою чергу суттєво зменшить надмірність мови, а також об'єм переданих

даних через канал зв'язку. Запропоновано підхід, для кількісного виміру семантичної інформації, яка знаходиться в мовних сигналах та показано, що маючи можливість визначити кількість фонемної і акустичної інформації, на сьогоднішній день немає остаточного теоретичного обґрунтування проблеми пов'язаної з доведенням можливості кількісного виміру семантичної інформації. Автор вбачає головний сенс даного дослідження в спробі об'єднати методи спектрального аналізу з теорією інформації для системної оцінки семантичної складової мовних сигналів, де спектральний аналіз відповідає за оцінку якості інформації, а теорія інформації за оцінку кількісної її сторони. Саме на основі їх глибинного взаємозв'язку вдасться надати мовній інформації смислове забарвлення, що є досить важливим в епоху інтелектуального аналізу даних. Доведено адаптивність емпіричного вейвлет-перетворення за допомогою сегментації спектра Фур'є, де компактні носії вейвлет-фільтрів Мейера напряму залежать від того, де знаходиться потрібна нам семантична інформація в спектрі мовного сигналу. Проведено адаптивну порогову обробку мовного сигналу з розрахунком оптимальних значень порогів функції вейвлет-фільтрації, для відсіювання коефіцієнтів які характеризують амплітуди і частоти спектра Гільберта малої потужності. Проведено вибір оптимальної функції порогової обробки за допомогою оцінки параметрів λ_1 , λ_2 , порогової функції вейвлет-фільтрації, що дозволяє знайти оптимальні пороги λ_{1op} , λ_{2opt} , з мінімальним середньоквадратическим відхиленням $\Delta(\lambda_1, \lambda_2)$, що дозволить підвищити ефективність визначення семантичних ознак мовного сигналу. Проведені експериментально-наукові дослідження в програмному середовищі MATLAB R2020b показали, 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. В наступному експерименті 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 для порівняння двох реалізацій семантичних ознак голосових команд знайдених на основі емпіричного вейвлет-перетворення та спектрального аналізу Гілберта. З отриманих результатів дослідження 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 емпіричного вейвлет-перетворення та спектрального аналізу Гілберта. Це забезпечує можливість семантичної ідентифікації мовних сигналів, а точніше експериментально підтверджує твердження, що мовні сигнали дійсно мають глибинні семантичні закономірності, визначенню та оцінці яких присвячене дане наукове дослідження. В даному експерименті було оцінено середні значення КС, ШПБ, КК, ВСШ, ПВСШ і СКП для ста реалізацій семантичних ознак голосових команд знайдених на основі емпіричного вейвлет-перетворення та спектрального аналізу Гілберта. З наведених вище результатів дослідження 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. Проведено

порівняльний аналіз перетворення Фур'є, косинусного перетворення та вейвлет-перетворення в якості спектрального аналізу цифрових мовних сигналів в алгоритмах стиснення, де з результатів явно видно що вейвлет-перетворення на базі вейвлет-функцій Sym12 та Db12 в 2.8 рази більше стискає досліджуваний сигнал аніж перетворення Фур'є, та в 2.2 рази більше аніж косинусне перетворення при тій же самій якості КК – 0.95. Показано, що при збільшенні числа нульових моментів зростає довжина фільтра (фільтри: Dmey – 102 коефіцієнти, Coif5 – 30 коефіцієнтів, Sym12 та Db12 – 24 коефіцієнти), що призводить до зростання обчислювальної складності. Доведено, що оптимальним методом спектрального аналізу для цифрових мовних сигналів в задачах їх стиснення є вейвлет-перетворення на базі вейвлет-функцій Sym12 та Db12 з результатами: КК – 0.9639 (0.9639), ВСШ – 11.4959 дБ (11.4894 дБ), ПВСШ – 31.2966 дБ (31.2901 дБ), СКП – 0.2662 (0.2664). 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.

Ключові слова: мовні сигнали, семантичне кодування, семантичні ознаки мовних сигналів, мел-частотні кепстральні коефіцієнти, адаптивний спектральний аналіз, емпіричне вейвлет-перетворення, адаптивні вейвлет-фільтри Мейера, функції внутрішніх емпіричних мод, спектральний аналіз Гільберта, розпізнавання і синтез мови, голосове управління.

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. It is proved that the semantic component of the speech signal is encoded in the acoustic oscillation in a very inefficient way, however, to highlight it and recode it with the optimal method is a rather non-trivial task on the decision of which this scientific research is directed. The fact is that we do not need acoustic characteristics of speech for semantic recognition, that is, you can recognizing the phonemes of the speech message, which in turn will significantly reduce the redundancy of speech, as well as the amount of data transmitted through the communication channel. An approach for quantitative measurement of semantic information contained in speech signals is proposed and it is shown that having the ability to determine the amount of phonemic and acoustic information, to date there is no final theoretical justification for the problem associated with proving the possibility of quantitative measurement of semantic information. The author sees the main meaning of this study in an attempt to combine the methods of spectral analysis with information theory for a systematic assessment of the semantic component of speech signals, where spectral analysis is responsible for assessing the quality of information, and information theory for assessing its quantitative side. It is on the basis of their deep relationship that it will be possible to give speech information a semantic coloring, which is quite important in the era of data mining. The adaptability of the empirical wavelet transform is proved by segmentation of the Fourier spectrum, where compact wavelet filters Meyer directly depend on where we need semantic information in the spectrum of

the speech signal. Adaptive threshold processing of the speech signal with calculation of optimal values of thresholds of the wavelet filtering function is carried out, for to remove the coefficients that characterize the amplitude and frequencies of the Hilbert spectrum of low power. The choice of the optimal function of threshold processing by means of parameter estimation is carried out λ_1, λ_2 , threshold function of wavelet filtering, which allows you to find the optimal thresholds $\lambda_{1opt}, \lambda_{2opt}$, with a minimum standard deviation $\Delta(\lambda_1, \lambda_2)$, which will increase the efficiency of determining the semantic features of the speech signal. 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 experiment evaluated the compression ratio (CR), bit rate (BR), correlation coefficient (CC), signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR) and root mean square error (RMSE), as well as the probability of error-free recognition of semantic features, which are the main indicators of the effectiveness of the proposed method for comparing two implementations of semantic features of voice commands found on the basis of empirical wavelet transform and Hilbert spectral analysis. The results of the study clearly show that the optimal solution for the given criteria of semantic coding efficiency for voice commands is: «Up» – CR = 333, BR = 192 bit/s, CC = 0.95, SNR = 21 dB, PSNR = 38 dB, RMSE = 0.04; «Down» – CR = 333, BR = 192 bit/s, CC = 0.94, SNR = 23 dB, PSNR = 41 dB, RMSE = 0.09; «Right» – CR = 333, BR = 192 bit/s, CC = 0.97, SNR = 26 dB, PSNR = 43 dB, RMSE = 0.06; «Left» – CR = 333, BR = 192 bit/s, CC = 0.91, SNR = 22 dB, PSNR = 38 dB, RMSE = 0.1. Which shows quite high results, preserving the semantic features of voice commands found on the basis of empirical wavelet transform and spectral analysis of Hilbert. This provides the possibility of semantic identification of speech signals, or rather

experimentally confirms the statement, that speech signals do have deep semantic regularities, the definition and evaluation of which is devoted to this research. In this experiment, the average values were estimated CR, BR, CC, SNR, PSNR and RMSE for one hundred realizations of semantic features of voice commands found on the basis of empirical wavelet transform and Hilbert spectral analysis. From the above results of the study are clearly visible, that with the increase of implementations the indicators of semantic coding efficiency remain at a high level, where the semantic component of speech signals retains its semantic patterns, which makes this method resistant to nonstationary and nonlinear processes. A comparative analysis of the Fourier transform, cosine transform and wavelet transform is performed as a spectral analysis of digital speech signals in compression algorithms, where the results clearly show that the wavelet transform is based on wavelet functions Sym12 and Db12 and 2.8 times more compression of the studied signal than the Fourier transform, and 2.2 times more than the cosine transformation with the same quality of the CC – 0.95. It is shown that as the number of zero moments increases, the length of the filter increases (filters: Dmey – 102 coefficients, Coif5 – 30 coefficients, Sym12 and Db12 – 24 coefficients), which leads to an increase in computational complexity. It is proved that the optimal method of spectral analysis for digital speech signals in their compression problems is wavelet transform based on wavelet functions Sym12 and Db12 with results: CC – 0.9639 (0.9639), SNR – 11.4959 dB (11.4894 dB), PSNR – 31.2966 dB (31.2901 dB), RMSE – 0.2662 (0.2664). 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, semantic features of speech signals, mel-frequency cepstral coefficients, adaptive spectral analysis, empirical wavelet transform, adaptive wavelet-filters Meyer, functions of internal empirical modes, Hilbert spectral analysis, speech recognition and synthesis, voice control.

Список публікацій здобувача

- [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. Bakhtiarov, "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. Bakhtiarov, "Comparative analysis of speech recognition algorithms in UAV voice control system," *Science-Based Technologies*, vol. 38, no. 2, pp. 137-142, 2018.

[10] O. Yu. Lavrynenko, 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. Конахович, 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. Конахович, 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. Конахович, Р. С. Одарченко, Д. И. Бахтияров, "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. Конахович, "Система защищённого голосового управления беспилотным летательным апаратом," *Авиационно-космическая техника и технология*, Т. 148, № 4, С. 4-11, 2018.

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

[16] Г. Ф. Конахович, 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, № у 2016 12187, заявл.
01.12.2016, опубл. 25.05.2017, бюл. № 10, 5 с.

[17] О. Ю. Лавриненко, Г. Ф. Конахович, 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, № у 2017 03401, заявл. 10.04.2017, опубл. 25.09.2017,
бюл. № 18, 6 с.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	16
ВСТУП	17
РОЗДІЛ 1. АНАЛІЗ СУЧАСНИХ МЕТОДІВ СЕМАНТИЧНОГО КОДУВАННЯ МОВНИХ СИГНАЛІВ	23
1.1. Взаємодія з технічними об'єктами за допомогою мовних сигналів... 23	
1.2. Мова і мовний сигнал.....	25
1.3. Кількісна оцінка мовної інформації.....	29
1.4. Метод семантичного кодування мовних сигналів на основі мел- частотних кепстральних коефіцієнтів	33
1.5. Формалізація проблеми і наукової гіпотези дослідження.....	38
Висновки до першого розділу	39
Список використаних джерел в першому розділі	40
РОЗДІЛ 2. ЕМПІРИЧНЕ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ В ЗАДАЧАХ СЕМАНТИЧНОГО КОДУВАННЯ МОВНИХ СИГНАЛІВ.....	44
2.1. Кількісна оцінка семантичної мовної інформації	44
2.2. Дискретне вейвлет-перетворення	50
2.2. Адаптивні вейвлет-фільтри Мейєра	58
2.3. Сегментація спектра Фур'є.....	61
2.4. Структура кадру вейвлет-фільтра Мейєра.....	63
2.5. Емпіричне вейвлет-перетворення	65
Висновки до другого розділу	66
Список використаних джерел у другому розділі	67

РОЗДІЛ 3. МЕТОД СЕМАНТИЧНОГО КОДУВАННЯ МОВНИХ СИГНАЛІВ НА ОСНОВІ ЕМПІРИЧНОГО ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ.....	70
3.1. Структурна схема запропонованого методу і первинна обробка мовного сигналу.....	70
3.2. Емпіричне вейвлет-перетворення сегментованого спектра Фур'є.....	74
3.3. Спектральний аналіз Гільберта	76
3.4. Порогова обробка спектра Гільберта	79
3.5. Вибір оптимальної порогової функції.....	81
Висновки до третього розділу	86
Список використаних джерел в третьому розділі	87
РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ СЕМАНТИЧНОГО КОДУВАННЯ МОВНИХ СИГНАЛІВ.....	90
4.1. Показники ефективності семантичного кодування мови.....	90
4.2. Порівняльний аналіз ПФ, КП та ВП як спектрального аналізу цифрових мовних сигналів	92
4.3. Результати експериментального дослідження методу семантичного кодування мовних сигналів на основі ЕВП	101
Висновки до четвертого розділу	107
Список використаних джерел в четвертому розділі	107
ВИСНОВКИ.....	112
ДОДАТОК А. МОДЕЛЮВАННЯ КІЛЬКІСНОЇ ОЦІНКИ СЕМАНТИЧНОЇ МОВНОЇ ІНФОРМАЦІЇ.....	114
ДОДАТОК Б. ПАТЕНТИ	136
ДОДАТОК В. ДОКУМЕНТИ ЩОДО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ.....	138

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

МЧКК	– мел-частотні кепстральні коефіцієнти
СК	– семантичне кодування
ЕВП	– емпіричне вейвлет-перетворення
ВП	– вейвлет-перетворення
ПФ	– перетворення Фур'є
ПГХ	– перетворення Гільберта-Хуанга
КП	– косинусне перетворення
ПГ	– перетворення Гільберта
ЕСГ	– енергетичний спектр Гільберта
ВЕМ	– внутрішні емпіричні моди
КА	– кепстральний аналіз
КМА	– кратномасштабний аналіз
ВК	– вейвлет-коефіцієнти
ВФМ	– вейвлет-фільтри Мейера
КС	– коефіцієнт стиснення
ПЗКЗ	– пропускна здатність каналу зв'язку
ШПБ	– швидкість передачі бітів
ВСШ	– відношення сигнал/шум
ПВСШ	– пікове відношення сигнал/шум
КК	– коефіцієнт кореляції
СКП	– середньоквадратична помилка
МО	– математичне очікування
КН	– коефіцієнт надмірності
СА	– спектральний аналіз
ЦОС	– цифрова обробка сигналів
ТКС	– телекомунікаційні системи

ВСТУП

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 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.

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 Мейера, знаходження функції внутрішніх емпіричних мод, кепстральний аналіз, перетворення Фур'є, кратномасштабний аналіз, перетворення Гільберта-Хуанга, косинусне перетворення, 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, частотна декореляція, усереднення спектральних коефіцієнтів, короткочасний аналіз, порогова обробка вейвлет-коефіцієнтів для знаходження семантичних ознак мовних сигналів);

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

4. теорії інформації та кодування (оцінювання кількості інформації, ентропії джерела, ефективності кодування, інформаційної місткості, 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 кепстральних 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.

РОЗДІЛ 1

АНАЛІЗ СУЧАСНИХ МЕТОДІВ СЕМАНТИЧНОГО КОДУВАННЯ МОВНИХ СИГНАЛІВ

1.1. Взаємодія з технічними об'єктами за допомогою мовних сигналів

Системи мовної взаємодії з технічними об'єктами як правило мають два режими роботи: режим навчання і режим розпізнавання (тестування). Ці режими входять в структурну схему системи мовної взаємодії з технічними об'єктами (Рис. 1.1), завдання, якої полягає в первинній обробці мовного сигналу і виділення ознак розпізнавання (відліки ознак фонем мовного сигналу). Подальша поведінка системи залежить від режиму роботи. Якщо система знаходиться в режимі навчання, отримані на етапі виділення ознаки розпізнавання, зберігаються в базу еталонних ознак розпізнавання. При знаходженні системи в стані розпізнавання, набір ознак розпізнавання фонемі голосової команди суб'єкта управління послідовно порівнюється з усіма еталонними наборами ознак розпізнавання з бази еталонних ознак розпізнавання. Задача функції прийняття рішення визначити найкращий результат порівняння по одному із заданих критеріїв і видати результат розпізнавання [1-3].

Головна проблема, що виникає при розробці такого роду систем, полягає в варіативній вимові одного і того ж слова як різними людьми, так і одним і тим же людиною в різних ситуаціях. Крім того, на вхідний сигнал впливають численні фактори, такі як навколишній шум, відлуння і перешкоди в каналі зв'язку. Ускладнюється це і тим, що шум і спотворення заздалегідь невідомі, тобто система не може бути підлаштована під них до початку роботи. Так як, апріорі неможливо оцінити, які ознаки більш підходять для розпізнавання мови, головне завдання даної наукової роботи виділити в мовному сигналі ознаки, релевантні для задачі семантичного розпізнавання, тобто інформацію, що представляє семантичну складову голосу суб'єкта [4].

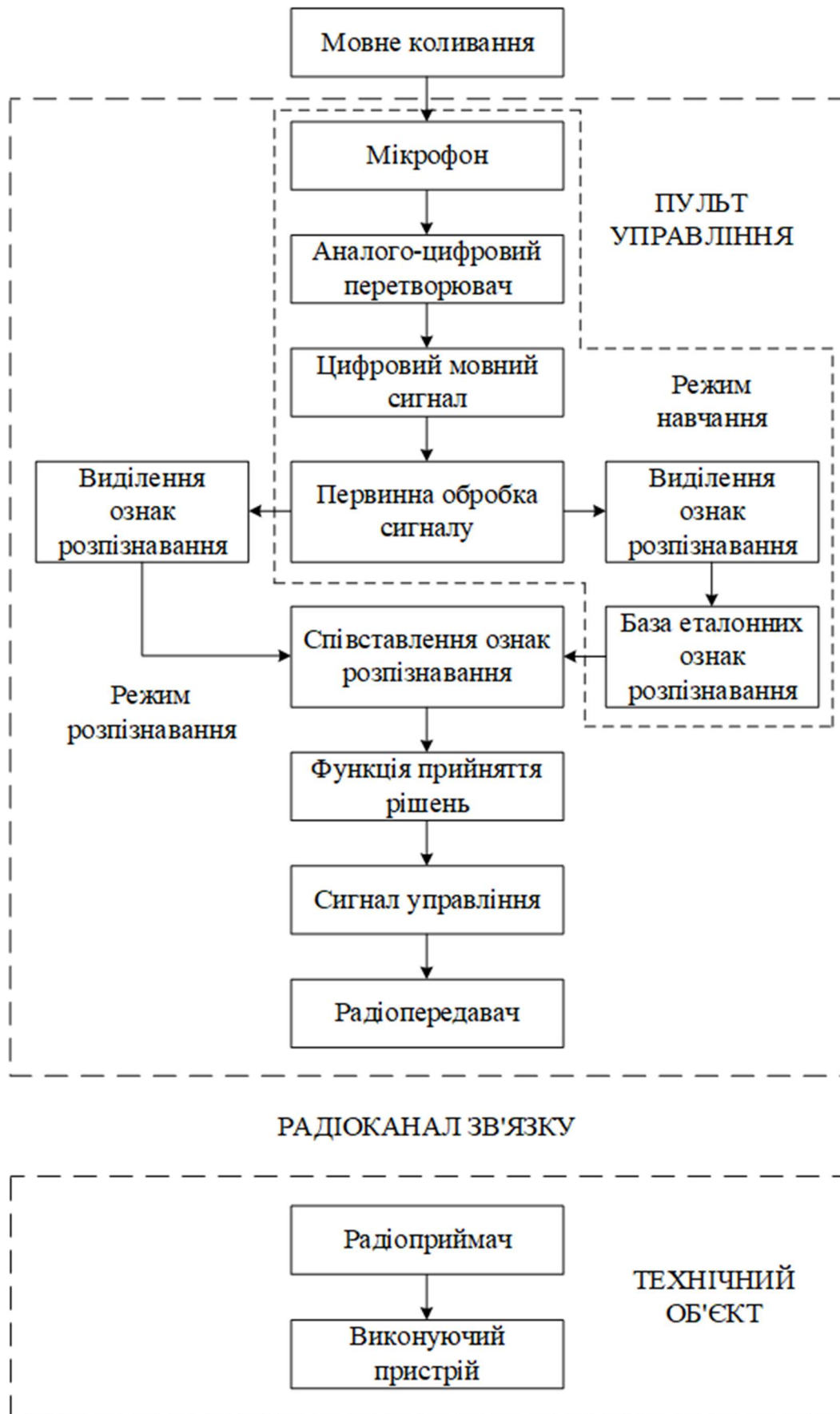


Рис. 1.1. Загальна структурна схема системи мовної взаємодії з технічними об'єктами

Виділені ознаки будуть використані для формування бази еталонних семантичних ознак розпізнавання мови або для порівняння з вже зареєстрованими ознаками розпізнавання (МЧКК) в базі. Процес визначення найбільш підходящих ознак розпізнавання фонем мовного сигналу полягає в проведенні порівняльного аналізу сучасних методів з подальшою експериментальною оцінкою [5-7].

При розробці системи автоматичної мовної взаємодії з технічними об'єктами, що є найбільш складною всеосяжною підсистемою мовного діалогу суб'єкта взаємодії з технічними об'єктами, використовуються різні методи обробки мовної інформації, такі як перетворення Фур'є, лінійне передбачення мови, вейвлет-перетворення, косинусне перетворення і т.д. Виділення найкращих параметричних характеристик семантичної складової мовних сигналів є важливим завданням при розробці будь-якої системи мовної взаємодії з технічними об'єктами. Воно суттєво впливає на якість розпізнавання мови, а саме підвищує стійкість системи до помилкової ідентифікації фонем, а значить підвищує ефективність системи мовної взаємодії з технічними об'єктами [8-10].

У більшості робіт для розпізнавання фонем використовуються параметри у вигляді коефіцієнтів кепстра, який обчислюється по огинаючій спектра, отриманого через перетворення Фур'є, за допомогою гребінки фільтрів, або по передаточній функції мовного тракту, знайденої методом лінійного передбачення. Крім того до коефіцієнтів кепстра використовуються також їх перші і другі різниці за часом. Перевага такого підходу полягає в обчислювальній простоті, а також в тому, що в кепстрі відображаються семантичні характеристики мовного тракту.

1.2. Мова і мовний сигнал

Мова призначена для спілкування. Можливості мови з цієї точки зору можна характеризувати по-різному. Один з кількісних підходів заснований на теорії інформації, розробленої К. Шенноном. Відповідно до цієї теорії мову

можна описати її інформаційним змістом, або інформацією. Інший спосіб опису мови полягає в представленні її у вигляді сигналу, тобто акустичного коливання. Мовне спілкування починається з того, що в мозку диктора виникає в абстрактній формі деяке повідомлення. В процесі мовотворення це повідомлення перетворюється в акустичне мовне коливання. Повідомлення, що передається за допомогою мовного сигналу - дискретно, тобто може бути представлено у вигляді послідовності символів з кінцевого їх числа. Звукові символи, з яких складається мовний сигнал, називаються фонемами. При нормальному темпі мовлення паузи з'являються між уривками фраз. Як правило, слова вимовляються разом, хоча слухачий сприймає слова окремо. При уповільненому темпі мови, паузи можуть робитися між словами і навіть їх частинами. Прийменники, сполучники звучать завжди разом з наступним словом. Один і той же звук мови різні люди вимовляють по-різному. Вимова звуків мови залежить від наголосу, сусідніх звуків і т.п. Але при всьому різноманітті в їх вимові звуки є фізичними реалізаціями (проголошенням) обмеженого числа узагальнених звуків мови (фонем). Фонема - це те, що людина повинна вимовити, а звук - то, що людина фактично вимовляє. Фонема по відношенню до звуку мови відіграє ту ж роль, що і зразкова буква по відношенню до її рукописної форми в конкретному написанні. В українській мові налічується 38 основні та 3 невизначені фонemi. Звуки мови діляться на дзвінки і глухі. Дзвінки звуки утворюються за участю голосових зв'язок, в цьому випадку знаходяться в напруженому стані.

Під напором повітря, що йде з легенів, вони періодично розсуваються, в результаті чого створюється переривчастий потік повітря. Імпульси потоку повітря, що створюються голосовими зв'язками, з достатньою точністю можуть вважатися періодичними. Відповідний період повторення імпульсів називають періодом основного тону голосу T_0 , а зворотну величину $1/T_0$ - частотою основного тону. Якщо зв'язки тонкі і сильно напружені, то період виходить коротким і частота основного тону високою; для товстих, слабо

напружених зв'язок - низькою. Частота основного тону для всіх голосів лежить в межах 70 - 450 Гц. При вимовленні мови вона безперервно змінюється відповідно до наголосу, підкреслення звуків і слів, а також з проявом емоцій (питання, вигук, здивування і т.д.). Зміна частоти основного тону називається інтонацією. У кожної людини свій діапазон зміни основного тону (зазвичай він буває трохи більше октави) і своя інтонація. Остання має велике значення для впізнавання мовця. Основний тон, інтонація, усний почерк і тембр голосу служать для впізнання людини, і ступінь достовірності така ж висока, як за відбитками пальців. Імпульси основного тону мають пилоподібну форму, і тому при їх періодичному повторенні виходить дискретний спектр з великим числом гармонік (до 40), частоти яких кратні частоті основного тону. Обвідна спектра основного тону має спад в бік високих частот з крутизною близько 6 дБ/окт, тому для чоловічого голосу рівень складових на частоті 3000 Гц нижче їх рівня на частоті 100 Гц приблизно на 30 дБ. При вимовленні глухих звуків зв'язки знаходяться в розслабленому стані, потік повітря з легенів вільно проходить в порожнину рота. Зустрічаючи на своєму шляху різні перешкоди у вигляді мови, зубів, губ, він утворює завихрення, що створюють шум із суцільним спектром. При проголошенні звуків мови язик, губи, зуби, нижня щелепа, голосові зв'язки повинні знаходитися для кожної фонемі в строго певному положенні або русі. Ці рухи називають артикуляцією органів мови. При цьому в мовотворчому тракті створюються резонансні порожнини, певні для даної фонемі, а для злитого звучання фонем в мові - певні переходи від однієї форми тракту до іншого.

При вимовленні звуків мови через мовний тракт проходить або тональний імпульсний сигнал, або шумовий, або той і інший разом. Мовний тракт являє собою складний акустичний фільтр з рядом резонансів, що створюються порожнинами рота, носа і носоглотки, тобто за допомогою артикуляційних органів мови. Внаслідок цього рівномірний тональний або шумовий спектр перетворюється в спектр з рядом максимумів і мінімумів.

Максимуми спектра називають формантами, а нульові провали - антиформантами. Для кожної фонемної огинаючої спектра має індивідуальну і цілком певну форму. При вимовленні мови спектр її безперервно змінюється, і утворюються формантні переходи. Частотний діапазон мовлення знаходиться в межах 70 - 7000 Гц. Дзвінкі звуки мови, особливо голосні, мають високий рівень інтенсивності, глухі - найнижчий. Гучність мови безперервно змінюється, особливо різко при проголошенні вибухових звуків. Динамічний діапазон рівнів мови знаходиться в межах 35 - 45 дБ. Голосні звуки мови мають в середньому тривалість близько 0,15 с, приголосні - близько 0,08 (звук п - близько 30 мс). Звуки мови неоднаково інформативні. Так, голосні звуки містять малу інформацію про сенс мови, а глухі приголосні найбільш інформативні (наприклад, в слові «посилка» послідовність «о, и, а» нічого не говорить, а «п, с, лк» дає майже однозначну відповідь про сенс). Тому розбірливість мови знижується при дії шумів, в першу чергу через маскування глухих звуків. Утворення звуків мови відбувається шляхом подачі команд до м'язів артикуляційних органів мови від мовного центру мозку. Загальний потік повідомлень від нього становить в середньому не більше 100 біт/с. Вся інша інформація в мовному сигналі називається супутньою. Майже вся інформація про звуки мови укладена в спектральній обвідній мови і її часовій зміні, частково - в переходах від тонального спектра до шумового і назад, за якими дізнаються про зміну дзвінких звуків на глухі і назад. Всі ці зміни відбуваються повільно (в темпі мови). Для відтворення мови досить передавати відомості про форму обвідної спектра мови та її тимчасову зміну в темпі зміни звуків мови, а також про зміну основного тону мови і переходів тон-шум. Розглянуті закономірності побудови мови формують складний багаточастотний сигнал, який потрібно належним чином обробити для виділення інформаційної частини. Для цього застосовують різні методи і перетворення.

1.3. Кількісна оцінка мовної інформації

Мовне спілкування починається з того, що в мозку диктора виникає в абстрактній формі деяке повідомлення. В процесі мовотворення це повідомлення перетворюється в акустичне мовне коливання. Інформація, що міститься в повідомленні, представлена в акустичному коливанні вельми складним чином. Повідомлення спочатку перетворюється в послідовності нервових імпульсів, які управляють артикуляторним апаратом (тобто переміщенням язика, губ, голосових зв'язок і т.д.). В результаті впливу нервових імпульсів артикуляторний апарат приходить в рух, результатом якого є акустичне мовне коливання, що несе інформацію про вихідне повідомлення [11].

Під семантикою мови розуміються типізовані формантні рисунки спектра мовного сигналу, які відповідають певній фонемі досліджуваної мови. Так, як семантична складова мовного сигналу знаходиться саме в частотній області, то досліджувані спектральні перетворення є важливим аспектом в доведенні ефективності семантичного кодування мови та кількісної оцінки семантичної інформації яка знаходиться в мовних сигналах [12-14].

Повідомлення, що передається за допомогою мовного сигналу, є дискретним, тобто може бути представлено у вигляді послідовності символів з кінцевого їх числа. Символи, з яких складається мовний сигнал, називаються фонемами. Найменшою елементарною одиницею мови прийнято вважати фонему, яку визначають, як членороздільний звук, що служить для розрізнення смислових одиниць мови. У кожній мові є властива їй множина фонем N , зазвичай від 30 до 50. Наприклад, в українській мові – $N = 38$ фонем. Таким чином, мовне повідомлення можна представити, як дискретну послідовність фонем [15-16].

Щоб мати можливість дати об'єктивну кількісну оцінку мовному повідомленню, вводиться поняття кількості інформації в мовному повідомленні. Таким чином, отримуємо обґрунтоване твердження, що якщо

фонема відповідної мови вибираються за умови рівної ймовірності їх появи, тобто $P = 1/N$, де N – число всіх можливих фонем, то середня кількість інформації на одну фонему української мови складатиме $H = \log_2 N = \log_2 38 = 5.2$ [біт/фонему]. Розрахована таким чином величина називається ентропією, що визначає середню кількість інформації, що припадає на один символ дискретного повідомлення. В даному випадку ми маємо справу виключно з фонемами мовних повідомлень [17-19].

Відомо, що фізичні обмеження на переміщення елементів артикуляторного апарату дозволяють людині вимовляти в середньому за хвилину 80 - 130 слів або близько 10 фонем в секунду. Приймаючи середню швидкість вимовлення рівній $W = n/\tau = 10/1 = 10$ [фонем/с], отримаємо, що швидкість передачі фонемної інформації мовного повідомлення українською мовою складе $C = W \times H = 10 \times 5.2 = 52$ [біт/с]. Іншими словами, при нормальному темпі вимовлення, письмовий еквівалент мовного повідомлення на різних світових мовах становить приблизно 60 біт/с. Ця цифра приблизно характеризує інформаційний зміст мови, що з'являється в її лінгвістичній структурі. Отримане значення добре узгоджується з результатами психоакустичних дослідів, які встановили, що людина здатна обробляти інформацію, що надходить по слуховому каналу, зі швидкістю до 50 біт/с. Ця оцінка, однак, не враховує таких факторів, як індивідуальність і емоційний стан диктора, швидкість вимовлення, гучність голосу і т.д [20-22].

Але якщо звернутися безпосередньо до акустичних характеристик мови, то тут інформаційна картина буде іншою. Миттєвий спектр мовного сигналу займає смугу частот приблизно від 100 до 8000 Гц, а динамічний діапазон амплітуд складає величину порядку 48 дБ. Звукові коливання характеризуються не тільки амплітудно-часовими і частотно-часовими параметрами, а й фазовими співвідношеннями. Якщо все це врахувати, то для запису повного набору звукових ознак, що міститься в одному слові, вимовленому за одну секунду, в формі, що вживається в обчислювальній

математиці, потрібно приблизно кілька десятків тисяч двійкових символів. Таким чином, мовний сигнал має величезну інформаційну надмірність [23].

Наведемо аргументоване обґрунтування даного твердження. Як відомо фонемна та акустична інформація мовного повідомлення знаходиться в смузі частот від 300 Гц до 3400 Гц і з динамічним діапазоном D не менше 48 дБ, а значить скориставшись теоремою відліків для неперервних сигналів з обмеженими (фінітними) спектрами, отримаємо частоту дискретизації $F_d = 2f_b = 2 \times 3400 = 6800$ [Гц]. Беручи до уваги умову, що $F_d \geq 2f_b$, на практиці частота дискретизації часто приймає значення $F_d = 8000$ [Гц]. Тоді загальне число значень відліків ν для сигналу тривалістю T складе $\nu = F_d T = 8000 \times 1 = 8000$ [відліків], де $T = 1$ [с], а швидкість передачі ν відліків мовного повідомлення тривалістю τ секунд складатиме $W_\nu = \nu/\tau = 8000/1 = 8000$ [відліків/с], де $\tau = 1$ [с], тобто ця величина представляє, очевидно, розмірність простору, відповідну базі сигналу [24-26]. Так як динамічний діапазон D мови повинен становити не менше 48 дБ, то тоді кількість біт що відводяться на один відлік мовного повідомлення буде становити не менше $k = 8$ [біт/відлік], що відповідає числу можливих рівнів квантування $L_\nu = 256$ і динамічному діапазону $D = 48.2$ [дБ]. В такому випадку швидкість передачі мовної інформації C_ν (акустичні характеристики мови) складе $C_\nu = W_\nu \times k = 8000 \times 8 = 64000$ [біт/с].

Таким чином, швидкість передачі фонемної інформації (лінгвістична семантика) мовного повідомлення $C = 60$ [біт/с], а швидкість передачі інформації акустичних характеристик мови $C_\nu = 64000$ [біт/с], що показує надмірність більше як в 1000 разів [27]. Це доводить що семантична складова мовного сигналу закодована в акустичному коливанні вельми неефективним чином, однак виділити її і перекодувати оптимальним методом є досить нетривіальною задачею на рішення якої і направлено дане наукове дослідження. Справа полягає в тому, що для семантичного розпізнавання нам не потрібні акустичні характеристики мови, тобто можна обійтися

розпізнаванням фонем мовного повідомлення, що в свою чергу суттєво зменшить надмірність мови, а також об'єм переданих даних через канал зв'язку.

Проблема наукового дослідження полягає в тому, що маючи можливість визначити і виміряти кількість фонемної і акустичної інформації, яка знаходиться в мовних сигналах згідно вищевикладеного матеріалу, на сьогоднішній день немає остаточного теоретичного обґрунтування проблеми пов'язаної з семантичним кодуванням мови, а саме доведення можливості кількісного виміру семантики прихованої в глибинних закономірностях мовного сигналу. Пов'язано це в більшій мірі з тим що мовний сигнал за своєю суттю є нестационарним і нелінійним процесом, а тому дослідження такого роду функцій на предмет глибинних семантичних складових (миттєвих частот та амплітуд) є проблематичним, оскільки існуючі методи семантичного кодування мови на основі спектральних аналізів таких як, перетворення Фур'є, вейвлет-перетворення та косинусне перетворення використовують апріорні базисні функції на всіх ітераціях розкладання, що не дозволяє довести оптимальність кодування при даній умові, так як помилка введена самим базисом буде накопичуватися в амплітудно-частотному формантному рисунку характерному даному спектральному перетворенню.

Тоді оптимальність семантичного кодування мовних сигналів і визначити кількісну міру семантичної інформації можливо тільки в тому випадку якщо буде дотримана адаптивність базисної функції до досліджуваного сигналу на кожній ітерації спектрального розкладання в певний базисний ряд з подальшим визначенням миттєвої частоти і амплітуди формантного рисунка спектра мовного сигналу.

Досліджувані сучасні методи семантичного кодування мовних сигналів не дотримуються сформульованому твердженню, з цієї причини було вперше запропоновано використовувати метод адаптивного емпіричного вейвлет-перетворення з подальшим спектральним аналізом Гільберта для визначення

семантичних ознак мовних сигналів і їх інформаційного кількісного виміру. Розроблений метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення з подальшим застосуванням спектрального аналізу Гільберта дотримується умов адаптивності в повній мірі, за рахунок чого буде теоретично доведено оптимальність даного методу і отриманий вигравш у якості підвищення ефективності семантичного кодування на відміну від існуючих методів.

1.4. Метод семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів

Для виділення семантичних ознак мовних сигналів вдаються до кепстрального аналізу, а саме, розраховуються МЧКК, що включає в себе наступні кроки (див. Рис. 1.2).

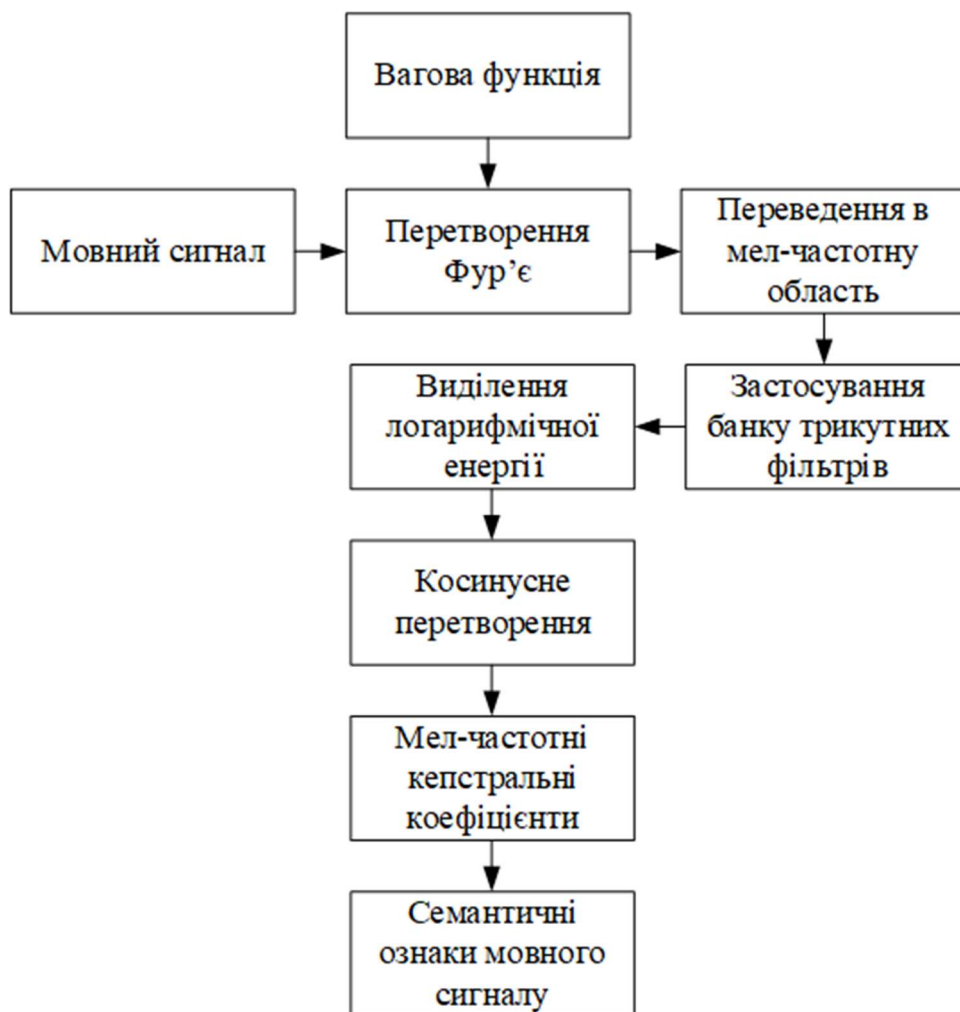


Рис. 1.2. Метод обчислення МЧКК

На вхід блоку дискретного перетворення Фур'є (ДПФ) подається послідовність відліків ділянки 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 Фур'є аналізі, викликаних скінченністю вибірки [34].

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

$$w[n] = 0.53836 - 0.46164 \times \cos\left(2\pi \frac{n}{N-1}\right), \quad n = 0, \dots, N-1, \quad (1.1)$$

де, N – довжина вікна, виражена у відліках.

Тоді ДПФ мовного сигналу з попередньою обробкою ваговою функцією можна записати у вигляді формули:

$$X[k] = \sum_{n=0}^{N-1} x[n] w[n] e^{-\frac{2\pi j}{N} kn}, \quad k = 0, \dots, N-1. \quad (1.2)$$

Значення індексів k відповідають частотам:

$$f[k] = \frac{F_s}{N} k, \quad k = 0, \dots, N/2, \quad (1.3)$$

де, F_s – частота дискретизації мовного сигналу [35-37].

Отримане представлення мовного сигналу в частотній області розбивають на діапазони за допомогою банку (гребінки) трикутних фільтрів. Помноживши функцію на фільтр, ми усереднюємо її на деякій ділянці.

Кожен трикутний фільтр знаходить зважене середнє тих амплітудних спектральних значень, що відповідають частотам в межах між нижньою і верхньою частотою для даного фільтра. Якщо амплітуда відповідає точно середній частоті смуги, то вона множиться на коефіцієнт який дорівнює одиниці.

При пересуванні відповідної амплітудному значенню частоти від середини до нижньої або верхньої границі, коефіцієнт зменшується від одиниці до нуля. Одержані добутки амплітуд на коефіцієнти додаються і діляться на число амплітудних значень. В результаті знаходимо зважене

середнє для даної смуги частот. Границі фільтрів розраховують в мел-шкалі. Дана шкала є результатом досліджень по здатності людського вуха до сприйняття звуків на різних частотах [38].

Мел-шкала моделює частотну чутливість людського слуху (див. Рис. 1.3). Подібні одиниці виміру часто використовують при вирішенні задач розпізнавання мови, так як вони дозволяють наблизитися до механізмів людського сприйняття, яке поки що лідирує серед відомих систем розпізнавання мови.

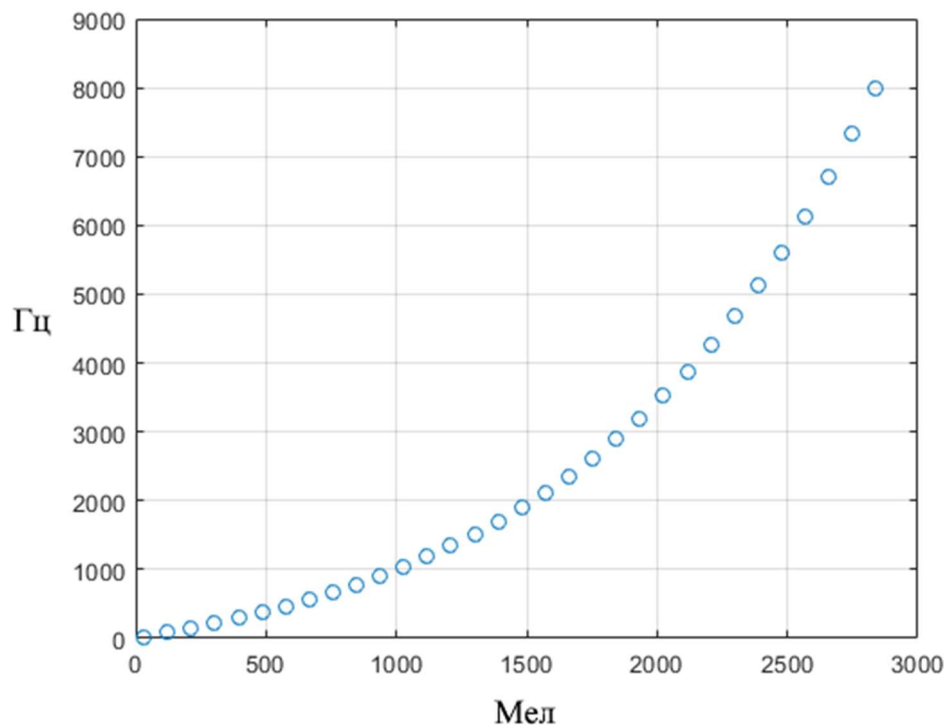


Рис. 1.3. Залежність частот в Гц від частот в мел-шкалі

Переведення в мел-частотну область здійснюється за формулою:

$$M = 1127.01048 \times \ln(1 + F/700). \quad (1.4)$$

Обернене перетворення в герци виражається за формулою:

$$F = 700 \times (e^{M/1127.01048} - 1). \quad (1.5)$$

Формула для поділу осі частот на трикутні фільтри матиме наступний вигляд:

$$f[m] = \left(\frac{N_f}{F_s}\right) M^{-1} \left(M(F_{min}) + m \frac{M(F_{max} - F_{min})}{N_f + 1} \right), \quad (1.6)$$

де, N_f – кількість мел-фільтрів (зазвичай використовують близько 24 фільтрів), F_s – частота дискретизації, $M(F)$ – функція переведення частоти в герцах в частоту в мелах, розглянута раніше, $M(F_{max} - F_{min})$ – досліджуваний діапазон частот в мелах, який розбивається на N_f рівномірно розподілених діапазонів які перекриваються, де обчислюються відповідні границі в області лінійних частот.

Складемо гребінку трикутних фільтрів за такою формулою:

$$H_m[k] = \begin{cases} 0, & k < f[m - 1] \\ \frac{(k - f[m - 1])}{(f[m] - f[m - 1])}, & f[m - 1] \leq k < f[m] \\ \frac{(f[m + 1] - k)}{(f[m + 1] - f[m])}, & f[m] \leq k \leq f[m + 1] \\ 0, & k > f[m + 1] \end{cases} \quad (1.7)$$

де, $H_m[k]$ – вагові коефіцієнти отриманих фільтрів.

Застосування фільтру полягає в попарному перемноженні його значень зі значеннями спектра. Оскільки фільтрів у нас N_f , коефіцієнтів буде стільки ж. Фільтри застосовуються до квадрату модулів коефіцієнтів ДПФ, тобто, нам потрібно застосувати мел-фільтри не до значень спектра, а до його енергії. Для цього потрібно обчислити енергію для кожного вікна, після чого прологарифмувати отримані результати. Вважається, що таким чином знижується чутливість коефіцієнтів до шумів [39].

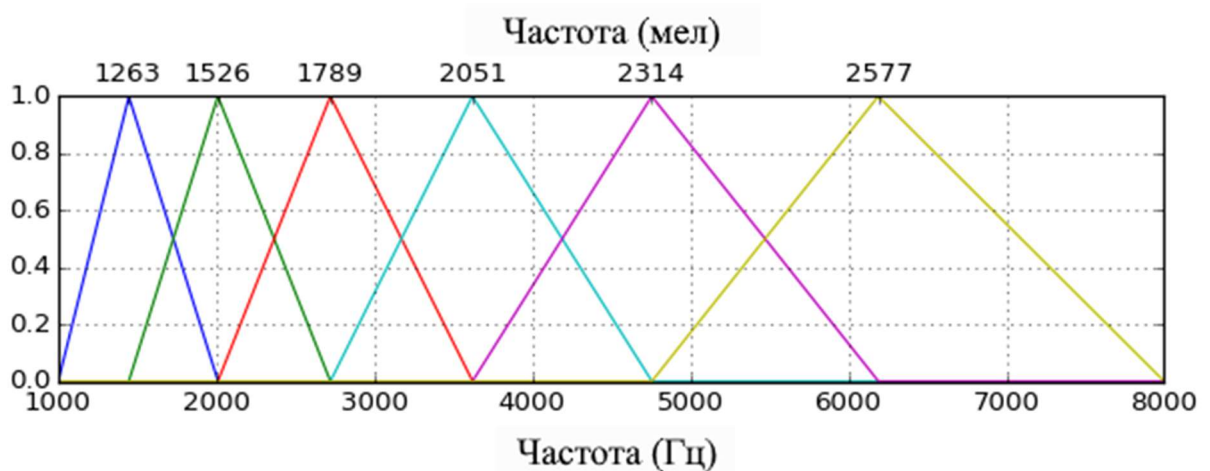


Рис. 1.4. Гребінка трикутних мел-фільтрів

Отримані значення енергії спектра логарифмуються і представляються наступним чином:

$$E[m] = \ln \left(\sum_{k=0}^{N-1} |X[k]|^2 H_m[k] \right), \quad m = 0, \dots, N_f - 1. \quad (1.8)$$

Заключним етапом в обчисленні МЧКК для зменшення кількості вихідних відліків і їх декореляції є ДКП, яке задається наступною формулою:

$$c[n] = \sum_{m=0}^{N_f-1} E[m] \cos \left(\frac{\pi n \left(m + \frac{1}{2} \right)}{N_f} \right), \quad n = 0, \dots, N_f - 1. \quad (1.9)$$

Це перетворення має властивість компактності енергії: більшій енергії відповідає менша кількість інформації. Отриманий набір значень називається МЧКК. Коефіцієнт $c[0]$ не використовується, тому що представляє собою енергію мовного сигналу [40-41]. Таким чином, ми маємо дуже невеликий набір значень, який при розпізнаванні успішно замінює тисячі відліків мовного сигналу. Зазвичай зберігаються тільки перші кілька елементів (від 8 до 16), за якими в подальшому проводиться ідентифікація голосових команд управління. На Рис. 1.5 зображені вичислені по досліджуваному методу МЧКК семантичні ознаки голосових команд: «Вгору», «Вниз», «Вправо», «Вліво».

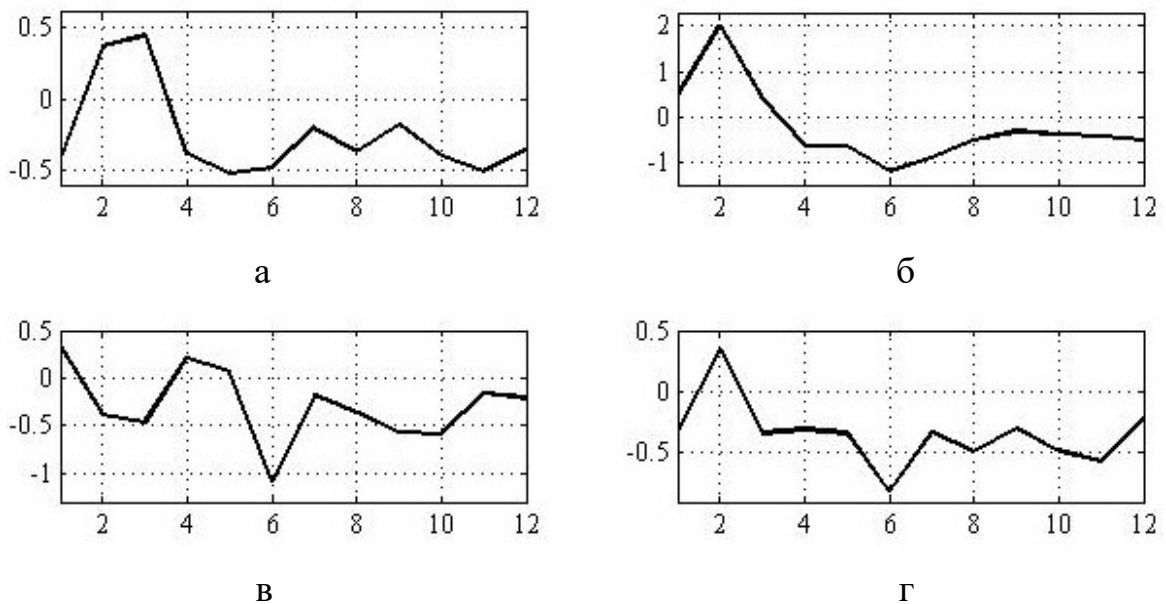


Рис. 1.5. МЧКК голосових команд:

а – «Вгору», б – «Вниз», в – «Вправо», г – «Вліво»

1.5. Формалізація проблеми і наукової гіпотези дослідження

У роботі 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)

$$c[n] = \sum_{m=0}^{N_f-1} E[m] \cos\left(\frac{\pi n \left(m + \frac{1}{2}\right)}{N_f}\right), \quad n = 0, \dots, N_f - 1. \quad (1.10)$$

прологарифмованої енергії спектра

$$E[m] = \ln\left(\sum_{k=0}^{N-1} |X[k]|^2 H_m[k]\right), \quad m = 0, \dots, N_f - 1. \quad (1.11)$$

дискретного перетворення Фур'є (ДПФ)

$$X[k] = \sum_{n=0}^{N-1} x[n] w[n] e^{-\frac{2\pi j}{N}kn}, \quad k = 0, \dots, N - 1. \quad (1.12)$$

обробленого трикутним фільтром

$$H_m[k] = \begin{cases} 0, & k < f[m-1] \\ \frac{(k - f[m-1])}{(f[m] - f[m-1])}, & f[m-1] \leq k < f[m] \\ \frac{(f[m+1] - k)}{(f[m+1] - f[m])}, & f[m] \leq k \leq f[m+1] \\ 0, & k > f[m+1] \end{cases} \quad (1.13)$$

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 $M = 1127.01048 \times \ln(1 + F/700)$.

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

$$\bigcup_{n=1}^N \Lambda_n = [0, \pi], \quad (1.14)$$

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

Висновки до першого розділу

1. Показано, що маючи можливість визначити кількість фонемної і акустичної інформації, яка знаходиться в мовних сигналах, на сьогоднішній день немає остаточного теоретичного обґрунтування проблеми пов'язаної з доведенням можливості кількісного виміру семантичної інформації.

2. Досліджено відомий метод 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.

3. Доведено, що представлений метод семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів не дотримується умови адаптивності.

4. 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.

Список використаних джерел в першому розділі

- [1] B. Milner, J. Darch, “Robust acoustic speech feature prediction from noisy mel-frequency cepstral coefficients,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 2, pp. 338-347, 2011.
- [2] D. L. Donoho, M. Vetterli, R. A. DeVore, I. Daubechies, “Data compression and harmonic analysis,” *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2435-2476, 1998.
- [3] Л. Рабинер, Б. Гоулд, *Теория и применение цифровой обработки сигналов*. Москва: Мир, 1978.
- [4] A. Bruce, D. Donoho, H.-Y. Gao, “Wavelet analysis (for signal processing),” *IEEE Spectrum*, vol. 33, no. 10, pp. 26-35, 1996.
- [5] B. Sklar, *Digital Communications: Fundamentals and Applications*. Upper Saddle River, NJ: Prentice Hall, 2001.
- [6] К. Шеннон, *Работы по теории информации и кибернетике*. Москва: Иностранная литература, 1963.
- [7] A. V. Oppenheim, R. Schaffer, J. R. Buck, *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice Hall, 1999.
- [8] S. Furui, “Cepstral analysis technique for automatic speaker verification,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 2, pp. 254-272, 1981.
- [9] Е. В. Егорова, М. Х. Аксяитов, А. Н. Рыбаков, *Методы повышения эффективности вейвлет-преобразований при обработке, сжатии и восстановления радиотехнических сигналов*. Тамбов: Консалтинговая компания Юком, 2019.
- [10] D. Imseng, P. Motlicek, P. N. Garner, H. Bourlard, “Impact of deep MLP architecture on different acoustic modeling techniques for under-resourced speech recognition,” *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding*, Olomouc, Dec. 8-12, 2013, pp. 332-337.

- [11] М. А. Сапожков, В. Г. Михайлов, *Вокодерная связь*. Москва: Радио и связь, 1983.
- [12] Т. К. Винцюк, *Анализ, распознавание и интерпретация речевых сигналов*. Киев: Наукова думка, 1987.
- [13] L. R. Rabiner, "Applications of voice processing to telecommunications," *Proceedings of the IEEE*, vol. 82, no. 2, pp. 199-228, 1994.
- [14] P. C. Loizou, *Speech Enhancement*. Boca Raton, FL: CRC Press, 2013.
- [15] Е. Н. Мясникова, *Объективное распознавание звуков речи*. Ленинград: Энергия, 1967.
- [16] A. V. Oppenheim, R. W. Schafer, "From frequency to quefrequency: a history of the cepstrum," *IEEE Signal Processing Magazine*, vol. 21, no. 5, pp. 95-106, 2004.
- [17] В. А. Игнатов, *Теория информации и передачи сигналов*. Москва: Советское радио, 1979.
- [18] А. Оппенгейм, *Применение цифровой обработки сигналов*. Москва: Мир, 1980.
- [19] S. Thomas, S. Ganapathy, H. Hermansky, "Phoneme recognition using spectral envelope and modulation frequency features," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Taipei, April 19-24, 2009, pp. 4453-4456.
- [20] А. Г. Зюко, Д. Д. Кловский, В. И. Коржик, М. В. Назаров, *Теория электрической связи*. Москва: Радио и связь, 1999.
- [21] K. Lee, D. P. W. Ellis, "Audio-based semantic concept classification for consumer video," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1406-1416, 2010.
- [22] В. Г. Михайлов, Л. В. Златоустова, *Измерение параметров речи*. Москва: Радио и связь, 1987.
- [23] J. Holmes, W. Holmes, *Speech Synthesis and Recognition*. London: CRC Press, 2001.

- [24] О. Ю. Гусев, Г. Ф. Конахович, В. І. Корнієнко, Г. В. Кузнецов, О. Ю. Пузиренко, *Теорія електричного зв'язку*. Львів: Магнолія, 2006.
- [25] У. Ли и др., *Методы автоматического распознавания речи. Том 1*. Москва: Мир, 1983.
- [26] M. Grimaldi, F. Cummins, "Speaker identification using instantaneous frequencies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 6, pp. 1097-1111, 2008.
- [27] М. А. Сапожков, *Речевой сигнал в кибернетике и связи*. Москва: Государственное издательство литературы по вопросам связи и радио, 1963.
- [28] B. Milner, X. Shao, "Prediction of fundamental frequency and voicing from mel-frequency cepstral coefficients for unconstrained speech reconstruction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 24-33, 2007.
- [29] D. M. Howard, J. A. S. Angus, *Acoustics and Psychoacoustics*. New York, NY: Routledge, 2017.
- [30] Р. Лайонс, *Цифровая обработка сигналов*. Москва: Бином-Пресс, 2006.
- [31] M. Gavish, D. L. Donoho, "Optimal shrinkage of singular values," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2137-2152, 2017.
- [32] У. Ли и др., *Методы автоматического распознавания речи. Том 2*. Москва: Мир, 1983.
- [33] У. Томаси, *Электронные системы связи*. Москва: Техносфера, 2007.
- [34] H. Krim, D. Tucker, S. Mallat, D. Donoho, "On denoising and best signal representation," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2225-2238, 1999.
- [35] Г. Фант, *Акустическая теория речеобразования*. Москва: Наука, 1964.

- [36] L. E. Boucheron, P. L. De Leon, S. Sandoval, “Low bit-rate speech coding through quantization of mel-frequency cepstral coefficients,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 2, pp. 610-619, 2012.
- [37] Д. Л. Фланаган, *Анализ, синтез и восприятие речи*. Москва: Связь, 1968.
- [38] А. А. Харкевич, *Очерки общей теории связи*. Москва: Государственное издательство технико-теоретической литературы, 1955.
- [39] D. Imseng, H. Bourlard, M. Magimai.-Doss, J. Dines, “Language dependent universal phoneme posterior estimation for mixed language speech recognition,” *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, May 22-27, 2011, pp. 5012-5015.
- [40] D. L. Donoho, A. Javanmard, A. Montanari, “Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7434-7464, 2013.
- [41] Л. А. Чистович и др., *Физиология речи. Восприятие речи человеком*. Ленинград: Наука, 1976.

РОЗДІЛ 2

ЕМПІРИЧНЕ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ В ЗАДАЧАХ СЕМАНТИЧНОГО КОДУВАННЯ МОВНИХ СИГНАЛІВ

2.1. Кількісна оцінка семантичної мовної інформації

Кількість семантичної інформації, що міститься в мовних сигналах, визначається з таких міркувань.

Нехай дано спектр F мовного сигналу у вигляді дискретної множини частот f_n і відповідних їм амплітуд $|a_n|$ $F = \left\{ \begin{matrix} |a_1| & |a_2| & \dots & |a_n| \\ f_1 & f_2 & \dots & f_n \end{matrix} \right\}$, об'єм якого дорівнює $2n$ значень. Тоді задаємо множину рівнів квантування $Q = \{q_1 \ q_2 \ \dots \ q_L\}$, де L – кількість рівнів квантування амплітуд $|a_n|$ частотного спектра F . Тоді можна переписати множину F у вигляді $F = \left\{ \begin{matrix} q_1 & q_2 & \dots & q_n \\ f_1 & f_2 & \dots & f_n \end{matrix} \right\}$ з квантовими амплітудами q_n спектра.

Припустимо, що поява кожного значення частоти f_n з множини F в спектрі мовного сигналу, рівноймовірна і несе в собі однаковий енергетичний внесок q_n (останнє твердження наводиться для плавного введення в опис пропонованого методу, відповідно до теорії спектрального аналізу, це звичайно не так, на чому і буде надалі зроблений акцент).

Отже, визначимо кількість семантичної інформації I в спектрі F мовного сигналу

$$I = n \log_2 L \text{ [біт]}, \quad (2.1)$$

де n – кількість семантичних частот в спектрі F , L – кількість рівнів квантування спектральних амплітуд.

Для визначення середньої кількості інформації H (ентропія), що припадає на одне значення квантованої спектральної амплітуди q_n візьмемо логарифм за основою два від загального числа можливих рівнів квантування L

$$H = \log_2 L \text{ [біт/відлік]}, \quad (2.2)$$

за умови, що всі квантовані амплітуди q_n спектра F рівноймовірні, тобто $P = 1/L$ – ймовірність появи кожної амплітуди в спектрі мовного сигналу.

Дане визначення кількості інформації може бути застосовано лише до дискретних сигналів, причому таких, у яких амплітудно-частотні складові рівноймовірні і взаємно незалежні. Останнє твердження є досить сумнівним для мовних сигналів, а точніше хибним, що буде доведено далі. Це підтверджується хоча б тим фактом, що спектр мовного сигналу схильний до повної взаємної кореляції. За таких міркувань заснована теорія дискретизації згідно теореми відліків, яка також не враховує ймовірність появи частот спектра сигналу і відповідні їм амплітуди потужності, і в тих випадках, коли потрібно кількісно оцінити мовну функцію повністю, нам і не потрібно амплітудно-частотний розподіл ймовірностей, досить лише знати максимальну частоту фінітного спектра, але існують задачі, коли нам потрібно кількісно оцінити не всю складову мовного сигналу, а тільки її частину, таку, як наприклад семантика мови. Так як вся інформація про мовний сигнал, зокрема семантична, прихована саме в його амплітудно-частотних глибинних закономірностях, то вводиться поняття амплітудно-частотної невизначеності, або спектральної невизначеності, як загальний випадок.

Дане твердження вимагає пояснень, автор ґрунтується на тому теоретично доведеному факті, що, чим складніше аналізована функція часу, тим вище її амплітудно-частотна невизначеність, наприклад, функція синусоїдального закону розподілу має чітку амплітудно-частотну визначеність, так само, як прямокутний і трикутний імпульси і т.д., але в складних сигналах, якими є мовні сигнали (складність характеризується нестационарністю і нелінійністю мовного процесу) простежується чітка амплітудно-частотна невизначеність, оскільки в них приховані закономірності більш глибокого рівня. Виходячи з вищевикладеного можна припустити, що амплітудно-частотна невизначеність залежить безпосередньо від прихованих семантичних закономірностей досліджуваного сигналу, і чим більше ці

закономірності приховані, тим більше зростає невизначеність, а значить більш складним стає його аналіз, але в той же час прихована інформація представляє собою більш сильну цінність. Дане вищевикладене твердження не враховує теорема відліків, а значить неможливо оцінити кількість інформації мовних сигналів з якісної сторони. Розробка нових підходів кількісної оцінки якості мовних сигналів, є науково важливими, оскільки, ми переходимо не просто до аналізу мовних даних, а до їх інтелектуального аналізу і відповідної їм інформації з усіма внутрішніми прихованими закономірностями. Саме для цих задач застосовується розглянуте в другому розділі адаптивне емпіричне вейвлет-перетворення, в якості ефективного методу розкладання мовного сигналу на внутрішні емпіричні моди з подальшим спектральним аналізом Гілберта, з метою визначення глибинних семантичних закономірностей мовних сигналів в амплітудно-частотній області. Потрібно мати на увазі для більш чіткого розуміння нижчевикладених суджень, що процес вимірювання семантичної інформації розглядається в трьох вимірному просторі, а саме, час-амплітуда-частота, які існують паралельно в загальній взаємодії між собою протягом певного моменту часу.

Отже, виходячи з вищесказаного, амплітудно-частотні складові спектра мовного сигналу не рівноймовірні, а значить міра невизначеності буде залежати не тільки від загального числа рівнів квантування L амплітудних значень q_n , але і від розподілу ймовірності P між ними.

Кількість семантичної інформації I_ω в спектрі F мовного сигналу, амплітудно-частотні складові якого з'являються не рівноймовірно, проте статистичний зв'язок між ними відсутній, визначається наступним чином

$$I_\omega = -n \sum_{i=1}^L P_{Q_i} \log_2 P_{Q_i} \text{ [біт]}. \quad (2.3)$$

Виходячи з цього, середня кількість семантичної інформації H_ω , що припадає на одне значення квантованої спектральної амплітуди q_n знаходиться згідно виразу

$$H_{\omega} = - \sum_{i=1}^L P_{Q_i} \log_2 P_{Q_i} \text{ [біт/відлік]}, \quad (2.4)$$

де n – кількість семантичних частот в спектрі F , L – кількість рівнів квантування спектральних амплітуд, P_{Q_i} – апіорна ймовірність появи рівня квантування Q_i , причому $\sum_{i=1}^L P_{Q_i} = 1$ – сума ймовірностей появи всіх рівнів квантування $Q_{i,\dots,L}$ в спектрі F мовного сигналу. Розподіли ймовірностей $P_{Q_{i,\dots,L}}$ можна визначити через обчислення спектральної щільності потужності.

Визначена таким чином величина H_{ω} називається ентропією, яка визначає середню кількість семантичної інформації, що припадає на один символ амплітудно-частотної складової.

У разі оцінки семантичної інформації ми маємо справу виключно з енергетичною спектральною складовою F мовних сигналів, а саме, з амплітудами потужності спектра q_n . Даний підхід за своєю сутністю переводить математичний формалізм з часової в частотну область, адже саме там приховані глибинні закономірності про семантику мови, а значить і ентропія повинна показувати міру спектральної невизначеності, про яку було сказано на початку. Тоді можна дати визначення ентропії мовних сигналів з огляду на вищевикладені твердження.

Спектральна ентропія мовних сигналів – середня кількість інформації, що припадає на одне значення амплітудного спектра мови.

Семантична ентропія мовних сигналів – середня кількість інформації, що припадає на одне значення амплітудного спектра мови, що характеризує семантику.

Відповідно до амплітудно-частотної складової спектра мовного сигналу, можна судити щодо інформативності окремих частот: чим рідше (менша ймовірність) зустрічається частота з високою енергією, тим більше інформації вона несе, і навпаки, чим частіше (більша ймовірність) зустрічається частота з низькою енергією, тим менше інформації вона несе, тобто амплітуди спектра характеризують інформаційний внесок певної частоти в мовний сигнал, що

відповідає ентропії. Отже, низькі частоти (від 300 до 1500 Гц) несуть максимальний внесок в семантику мови, середні частоти (від 1500 до 2500 Гц) – мінімальний, а високі (від 2500 до 4000 Гц) зовсім не впливають на неї. Середні і високі частоти (від 1500 до 4000 Гц) в більшій мірі впливають на біометрію голосу.

У загальному випадку, амплітудно-частотні складові спектра мовного сигналу, можуть з'являтися не тільки з різною ймовірністю, а й бути статистично залежними. Статистична залежність може бути виражена умовною ймовірністю появи однієї частоти після іншої.

Кількість семантичної інформації I_ψ , що міститься в спектрі мовного сигналу, амплітудно-частотні складові якого з'являються з різною ймовірністю і мають статистичну залежність, визначається наступним чином

$$I_\psi = -n \sum_{i=1}^L P_{Q_i} \sum_{j=1}^L P(Q_j/Q_i) \log_2 P(Q_j/Q_i) \text{ [біт]}, \quad (2.5)$$

при цьому середня кількість семантичної інформації H_ψ , що припадає на одне значення квантованої спектральної амплітуди q_n , становитиме

$$H_\psi = - \sum_{i=1}^L P_{Q_i} \sum_{j=1}^L P(Q_j/Q_i) \log_2 P(Q_j/Q_i) \text{ [біт/відлік]}, \quad (2.6)$$

де $P(Q_j/Q_i)$ – умовна ймовірність появи частоти Q_j після Q_i .

При нерівній ймовірності появи амплітудно-частотних складових спектра F мовного сигналу, кількість інформації, що міститься в ньому, знижується. Іншим чинником, що знижує ентропію, а отже і кількість інформації в спектрі мовних сигналів, є наявність статистичної залежності між частотами – кореляції.

Через не рівноймовірну появу P_{Q_i} амплітудно-частотних складових спектра F мовного сигналу і кореляційних зав'язків між ними $P(Q_j/Q_i)$ кількість семантичної інформації в спектрі реальних мовних сигналах падає. Кількісно ці втрати інформації характеризуються коефіцієнтом надмірності

$$r_{\psi} = 1 - \frac{H_{\psi}}{\log_2 L}, \quad (2.7)$$

де H_{ψ} – середня кількість семантичної інформації, яку переносить одне значення квантованої спектральної амплітуди q_n в реальних мовних сигналах (формула 2.6); L – кількість рівнів квантування в алфавіті джерела мовних сигналів (обсяг алфавіту квантування).

Як видно з виразу (2.7), чим більше семантична ентропія спектра F , тим менше його надмірність і навпаки. Надмірність $r \neq 0$ говорить про те, що кількість рівнів квантування L мовного сигналу більша, ніж це потрібно було б при повному їх використанні, тобто за умови, що спектральні амплітудно-частотні складові з'являються рівномірно і взаємно незалежно. Для мовних сигналів семантична надмірність становить не менше 0,95.

В цьому випадку ми зменшуємо кількість амплітудно-частотних складових спектра мовного сигналу i , відповідно, зменшується його надмірність, тим самим підвищуючи ефективність семантичного кодування. Дане твердження засноване на тому факті, що семантика мови високого рівня (синтезовані мовні сигнали) знаходиться в смузі низьких частот від 300 до 1000 Гц (семантика первинного рівня представляється ще меншою частотою, близько 100 Гц), а тому, частоти поза цими діапазонами, нам не слід кодувати взагалі.

Особливий інтерес представляє оцінка швидкості передачі семантичної інформації, що міститься в спектрі мовних сигналів.

Тоді, швидкість передачі семантичної інформації R_{ψ} , що міститься в спектрі F мовних сигналів, з урахуванням ймовірності появи амплітудно-частотних складових P_{Q_i} і зав'язків між ними $P(Q_j/Q_i)$, визначається наступною формулою

$$R_{\psi} = W \times H_{\psi} \text{ [біт/с]}, \quad (2.8)$$

де $W = n/\tau$ [відліків/с] – швидкість передачі семантичної амплітудно-частотної складової (n – кількість семантичних частот), що міститься в спектрі

F мовних сигналів тривалістю $\tau = (1/F_{\text{дс}}) \times n$ секунд, при заданій семантичній частоті дискретизації $F_{\text{дс}}$ вихідного сигналу; H_{ψ} [біт/відлік] – середня кількість семантичної інформації що припадає на одне значення квантованої спектральної амплітуди q_n , враховуючи ймовірність появи амплітудно-частотних складових P_{Q_i} і зв'язки між ними $P(Q_j/Q_i)$.

Тоді згідно наведених вище тверджень, практично досяжна найменша швидкість передачі семантичної інформації R_{ψ} , що еквівалентно максимальному підвищенню ефективності семантичного кодування розраховується за формулою (2.8), а теоретично можливі межі підвищення ефективності семантичного кодування, при існуванні близьких до ідеальних семантичних перетворювачів, можна задати умовою

$$R_{min} = \begin{cases} W \rightarrow 1 \\ H_{\psi} \rightarrow 1 \end{cases} \text{ [біт/с]}, \quad (2.9)$$

де швидкість передачі W амплітудно-частотних складових і їх ентропія H_{ψ} прагнуть до одиниці.

Автор вбачає головний сенс даного дослідження в спробі об'єднати методи спектрального аналізу з теорією інформації для системної оцінки семантичної складової мовних сигналів, де спектральний аналіз відповідає за оцінку якості інформації, а теорія інформації за оцінку кількісної її сторони. Саме на основі їх глибинного взаємозв'язку вдасться надати мовній інформації смислове забарвлення, що є досить важливим в епоху інтелектуального аналізу даних.

2.2. Дискретне вейвлет-перетворення

Властивості частотно-часової локалізації і добре розроблені алгоритми швидкого вейвлет-перетворення (ШВП) обумовлюють широке застосування вейвлет-перетворення в галузі аналізу нестационарних сигналів. ВП володіє істотними перевагами в порівнянні з перетворенням Фур'є, тому що з його допомогою можна аналізувати короткочасні локальні особливості сигналів, наприклад, короткі сплески чи провали, розриви і т.д. Унікальні властивості

ВП дозволяють сконструювати базис, в якому представлення даних може виражатися невеликою кількістю ненульових коефіцієнтів. Ця властивість робить ВП привабливим для використання його, як метод первинної обробки мовного сигналу для підвищення ефективності його стиснення. Безпосередньо стиск виконується після ВП, для коефіцієнтів вейвлет-розкладання сигналу, а його відновлення за цими коефіцієнтам проводиться на етапі вейвлет-реконструкції.

Дискретне вейвлет-перетворення (ДВП) може бути реалізовано у вигляді каскадного з'єднання низькочастотних і високочастотних фільтрів або пірамідального алгоритму Малла. У цьому випадку смуговий фільтр для кожного масштабу складається з пари доповнюючи один одного фільтрів низьких і високих частот, що відносяться до класу квадратурних дзеркальних фільтрів. Особливістю цього класу фільтрів є те, що фільтр високих частот виходить з відповідного фільтра низьких частот простою перестановкою його коефіцієнтів у зворотному порядку і зміною знака половини з них (тільки парних або тільки непарних). Процес субполосної фільтрації вихідного сигналу $s(n)$, що складається з N відліків ($N = 2^J$, де J – число масштабів або число каскадування фільтрів), може бути представлений у матричній формі дискретного вейвлет-перетворення: $\vec{c}_j = H_j \vec{c}_{j-1}$, $\vec{d}_j = G_j \vec{c}_{j-1}$ де, $\vec{c}_j = (c_j(0), c_j(1), \dots, c_j(N/2^j - 1))$ і $\vec{d}_j = (d_j(0), d_j(1), \dots, d_j(N/2^j - 1))$ – вектори-стовпці виходів скейлинг-фільтра і вейвлет-фільтра для деякого j , що складаються з коефіцієнтів, що характеризують спектр сигналу $s(t)$ і проріджених в два рази [4].

$$H_j = \begin{bmatrix} \vec{h}^T(1) \\ \vec{h}^T(2) \\ \vdots \\ \vec{h}^T(N/2^j) \end{bmatrix}_{(N/2^j)(N/2^{j-1})} \quad G_j = \begin{bmatrix} \vec{g}^T(1) \\ \vec{g}^T(2) \\ \vdots \\ \vec{g}^T(N/2^j) \end{bmatrix}_{(N/2^j)(N/2^{j-1})}$$

– матриці перетворення розміром $(N/2^j)(N/2^{j-1})$ на j -й ітерації, що представляють собою набори векторів-рядків $\vec{h}^T(n)$ і $\vec{g}^T(n)$ і описують скейлінг і вейвлет-фільтри відповідно; $\vec{h}^T(n)$ і $\vec{g}^T(n)$ – вектори-рядки, що складаються з коефіцієнтів, доповнених нульовими коефіцієнтами, для вейвлет-фільтрів і скейлінгових фільтрів відповідно. Коефіцієнти $c_j(n)$ на початковому значенні масштабу $j=0$ приймаються часові відліки вихідного сигналу, тобто $\{c_0(n)\}_{n=0}^{N-1} = \{s(kn)\}_{n=0}^{N-1}$.

Ітераційна процедура ДВП, закінчується при деякому значенні $j=J$, яке вибирається виходячи з апіорної інформації про сигнал, тобто з його тривалості [4]. На першому кроці багатокрокової ітераційної процедури проводиться обробка часових відліків сигналу $\{s(n)\}_{n=0}^{N-1}$, а на кожному наступному – відповідних коефіцієнтів c_j . Вейвлет-коефіцієнти $\{d_1(n)\}_{n=1}^{N/2}$ зберігаються як кінцевий результат, а скейлінгові коефіцієнти $\{c_1(n)\}_{n=1}^{N/2}$ використовуються в якості вихідних даних і рекурсивно обробляються аж до кінцевого масштабу J . В результаті рекурсивного виконання процедури будемо мати один вектор коефіцієнтів $\{c_J(n)\}$, обчислений на останньому масштабі, і набір векторів коефіцієнтів $\{d_j(n)\}_1^J$, обчислених на попередніх масштабах.

В якості ілюстрації наведемо наступний приклад. Візьмемо фільтр довжиною $L=4$, вихідний сигнал довжиною $N=8$, а в якості початкового значення – $j=1$. Послідовність значень $g(n)$ отримуємо з $h(n)$ за формулою $g_n = (-1)^n h_{L-1-n}$, де L – довжина коефіцієнтів вейвлет-фільтра. Тоді операція матрично-векторного множення може бути представлена у вигляді:

$$\begin{bmatrix} c_{1,0} \\ c_{1,1} \\ c_{1,2} \\ c_{1,3} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{H}_1 \bar{c}_0 = \begin{bmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_{0,0} \\ c_{0,1} \\ c_{0,2} \\ c_{0,3} \\ c_{0,4} \\ c_{0,5} \\ c_{0,6} \\ c_{0,7} \end{bmatrix}$$

$$\begin{bmatrix} d_{1,0} \\ d_{1,1} \\ d_{1,2} \\ d_{1,3} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{G}_1 \bar{c}_0 = \begin{bmatrix} h_3 & -h_2 & h_1 & -h_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_3 & -h_2 & h_1 & -h_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_3 & -h_2 & h_1 & -h_0 \\ h_1 & -h_0 & 0 & 0 & 0 & 0 & h_3 & -h_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_{0,0} \\ c_{0,1} \\ c_{0,2} \\ c_{0,3} \\ c_{0,4} \\ c_{0,5} \\ c_{0,6} \\ c_{0,7} \end{bmatrix}$$

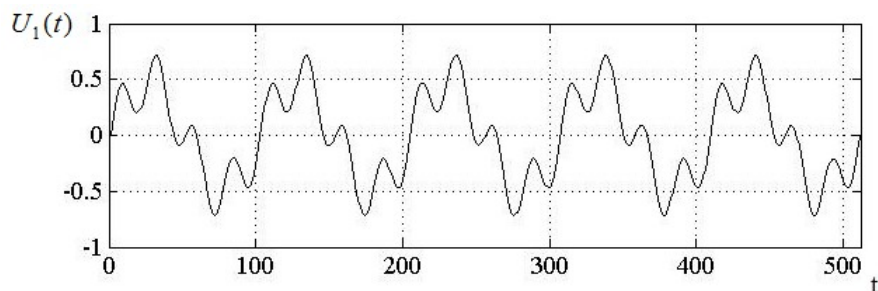
\mathbf{H}_1 і \mathbf{G}_1 – матриці субполосної низькочастотної і високочастотної фільтрації розміром 4×8 ; $\{h(n)\}_{n=1}^4$ і $\{g(n)\}_{n=1}^4$ – коефіцієнти скейлінг і вейвлет-фільтрів довжиною 4; $\{c_0(n)\}_{n=0}^{N-1} = \{s(n)\}_{n=0}^{N-1}$ – значення відліків вихідного сигналу.

Блок-схема одного кроку ітераційної процедури субполосного вейвлет-аналізу та синтезу сигналу по його коефіцієнтам розкладання. Згідно їй сигнал рівня $j = 0$ розкладається на два сигнали рівня $j = 1$. Розрішення сигналу, що є мірою кількості детальної інформації в сигналі, змінюється за рахунок фільтрації сигналу, а масштаб змінюється за рахунок децимації та інтерполяції [5].

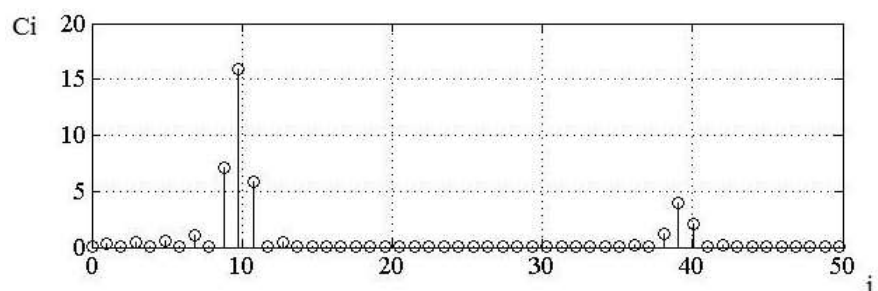
Операції децимації сигналу та інтерполяції в 2 рази, позначені символами \downarrow і \uparrow відповідно, виконуються в матрицях \mathbf{H} , \mathbf{G} , \mathbf{E} і \mathbf{F} . У нижній частині блок-схеми (рис. 1) виконується низькочастотна фільтрація. У результаті виходить апроксимація сигналу, позбавлена деталей – низькочастотна субполоса. У верхній частині схеми виділяється високочастотна субполоса. Дискретне вейвлет-перетворення виходить

шляхом рекурсивного застосування даної процедури до низькочастотної частини. Послідовний розподіл субполос тільки низькочастотних фільтрів обумовлено тим фактом, що саме низькочастотна область несе більше інформації, ніж високочастотна область. Дискретне вейвлет-перетворення вхідного сигналу виходить шляхом об'єднання вейвлеткоєфіцієнтів усіх рівнів перетворення і приєднання до них скейлінгових коефіцієнтів останнього рівня. У цьому випадку число вейвлет-коефіцієнтів дорівнює числу відліків у вхідному сигналі. Число операцій множення, необхідне для обчислення всіх коефіцієнтів дискретного вейвлет-перетворення для масиву даних N і довжини векторів h і g , рівний L , буде $2LN$. Стільки ж операцій потрібно виконати, щоб відновити або обчислити всі спектральні компоненти. Отже, для аналізу – синтезу сигналу $s(t)$ у вейвлетному базисі необхідно виконати $4LN$ операцій. Число операцій комплексного множення для ДПФ – $N \log_2 N$, що порівняно або навіть більше, ніж у випадку дискретного вейвлет-перетворення.

Недоліки ПФ демонструється на рис. 2.1а, б, і рис. 2.2а, б.



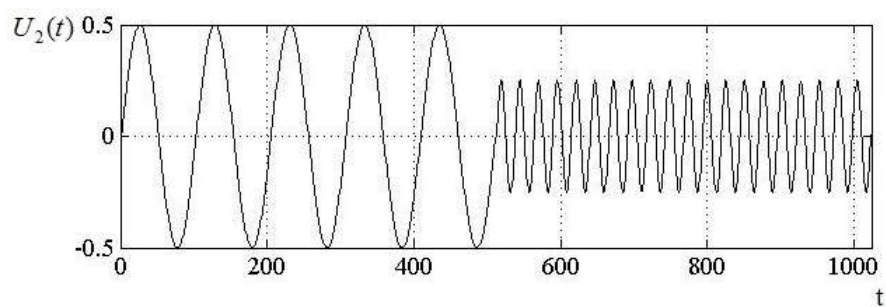
а



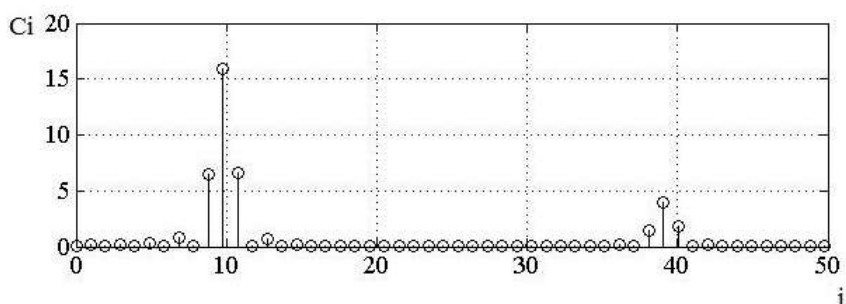
б

Рис. 2.1. ПФ сигнала $U_1(t)$

а – сигнал $U_1(t)$; б – спектр ПФ сигнала $U_1(t)$.



а



б

Рис. 2.2. ПФ сигнала $U_2(t)$

а – сигнал $U_2(t)$; б – спектр ПФ сигнала $U_2(t)$.

На рис. 2.1а і рис. 2.2а наведено дві гармонійні компоненти $S_1(t) = A_1 \cdot \sin(\omega_1 t)$ і $S_2(t) = A_2 \cdot \sin(\omega_2 t)$, з кутовими частотами $\omega_1 = 63$ рад/с і $\omega_2 = 252$ рад/с.

Кутова частота ω в рад/с виражається через частоту f в Гц, як $\omega = 2\pi f$, а $f = \frac{\omega}{2\pi}$. Виходячи з цього $f_1 = 10$ Гц, а $f_2 = 40$ Гц.

Процес, який представлений на рис. 2.1а, є адаптивною сукупністю двох синусоїд $S_1(t)$ і $S_2(t)$

$$U_1(t) = S_1(t) + S_2(t), \quad t \in (0, T],$$

де $A_1 = 0.5$, $A_2 = 0.25$ відповідно і $T = 512$, а процес, наведений на рис. 2.2а, описується таким чином

$$U_2(t) = \begin{cases} S_1(t), & t \in (0, t_0], \\ S_2(t), & t \in (t_0, T], \end{cases}$$

де $A_1 = 0.5$, $A_2 = 0.25$ відповідно і $t_0 = 512$, де $T = 1024$.

Поза інтервалу $(0, T]$ функції $U_1(t)$ і $U_2(t)$ рівні 0.

В результаті ПФ сигналів $U_1(t)$ і $U_2(t)$ отримані слабо различаючіе спектральні образи, які представлені на рис. 2.1б і рис. 2.2б.

Наступний приклад також показує малу інформативність ПФ. Представлений на рис. 2.3а сигнал $U_3(t)$ в околиці $t = 253:260$ містить короткий імпульс $I(t)$ (аномалію), де $t \in (-3, 3]$.

$$U_3(t) = \begin{cases} U_1(t), & t \in (0, t_1], \\ I(t), & t \in (t_1, t_2], \\ U_1(t), & t \in (t_2, T], \end{cases}$$

де $t_1 = 253$, $t_2 = 260$.

ПФ дозволило явно виділити дві гармонійні складові сигналу, а спектральні складові аномалії, як і слід було чекати, виявилися розподіленими по всій осі частот.

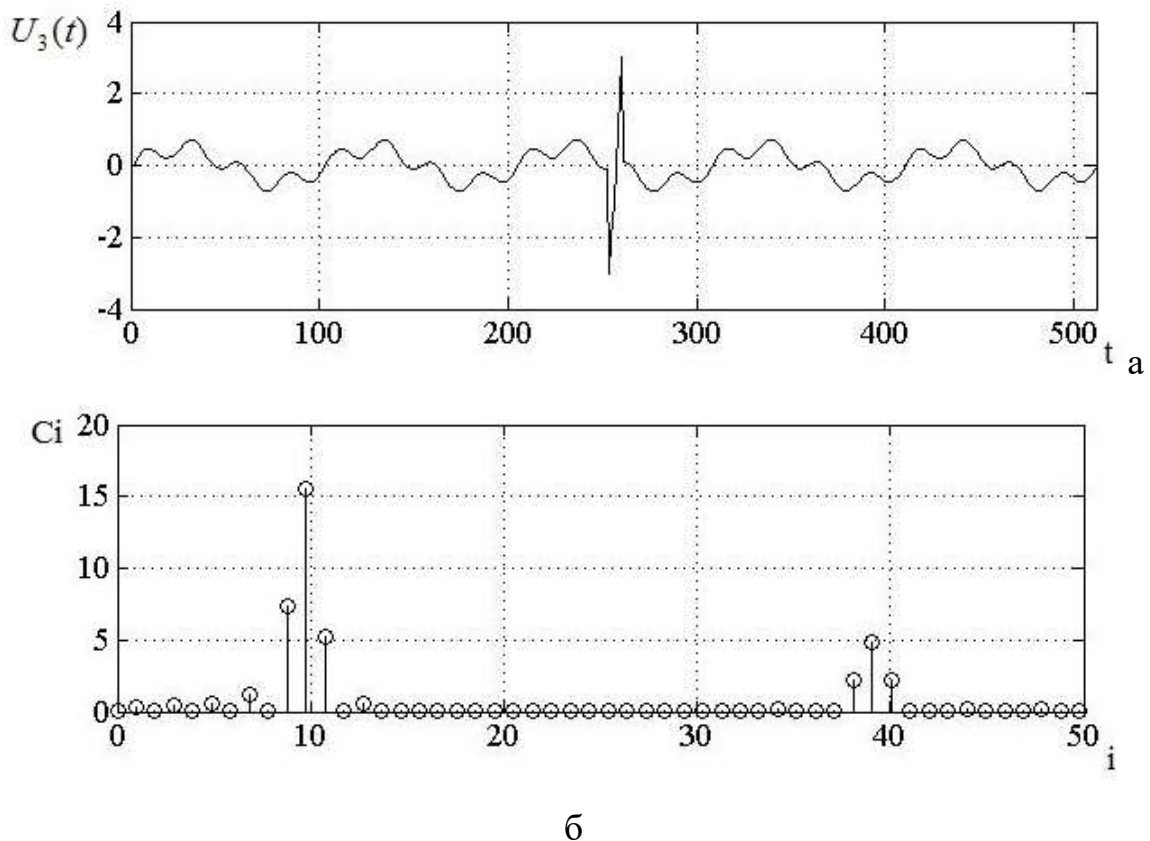


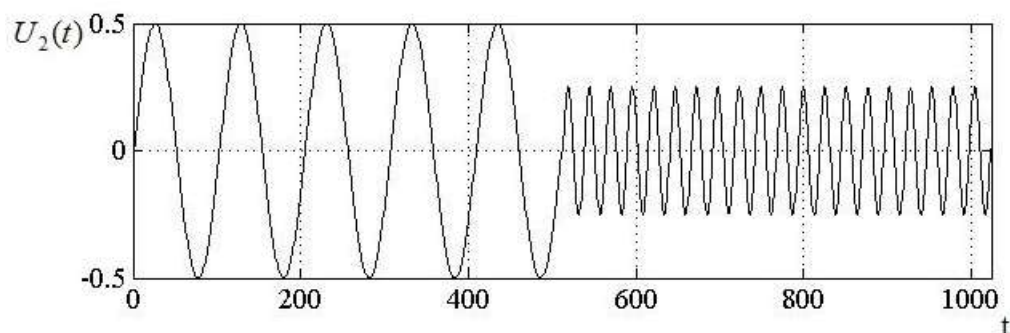
Рис. 2.3. ПФ сигнала $U_3(t)$

а – сигнал $U_3(t)$; б – спектр ПФ сигнала $U_3(t)$.

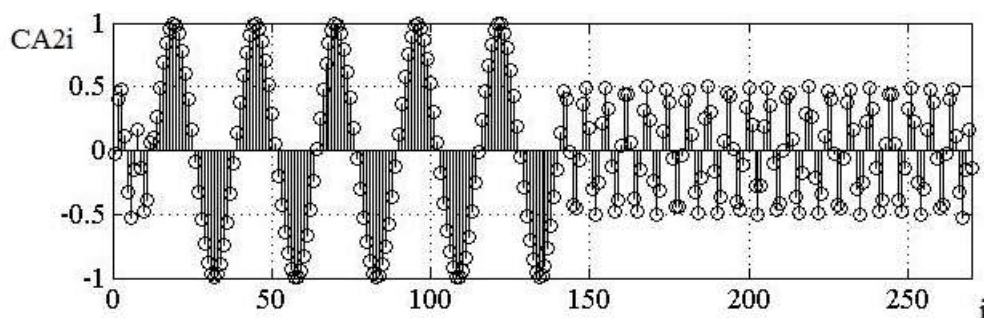
На рис. 2.1, рис. 2.2, рис. 2.3 були показані конкретні приклади недоліків ПФ, які можуть бути подолані при використанні ВП.

Слід зазначити, що в наведених спектрах ПФ міститься вся інформація про вихідні сигнали. Ця інформація розподілена в значеннях фаз і амплітуд всіх спектральних складових. Вихідні сигнали можуть бути повністю відновлені після зворотного ПФ.

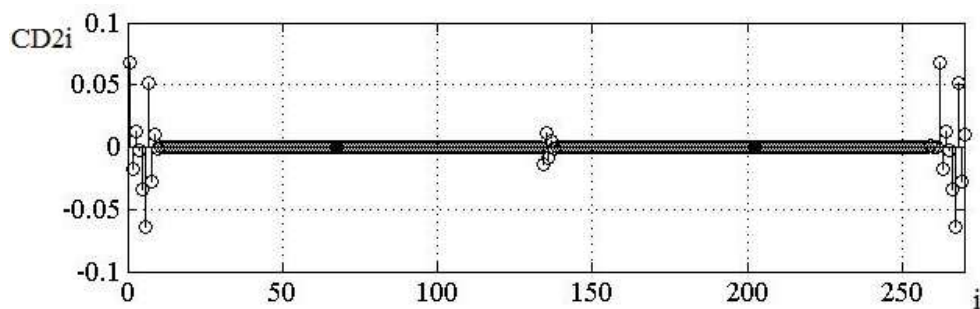
Переваги ВП демонструються на рис. 2.4а, б, в, і рис. 2.5а, б, в.



а



б

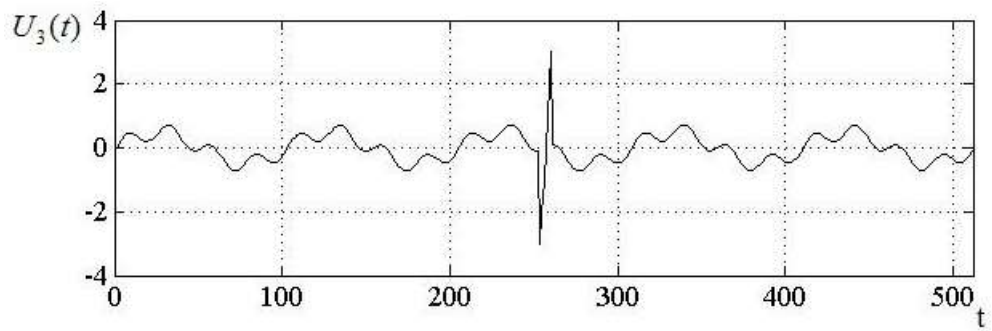


в

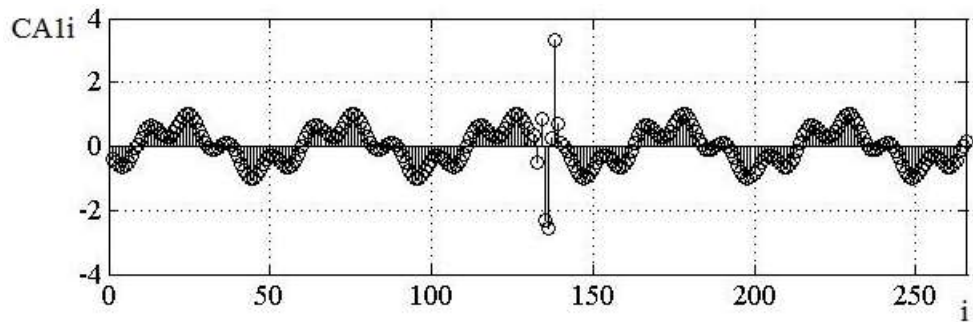
Рис. 2.4. ВП сигнала $U_2(t)$

а – сигнал $U_2(t)$; б – коефіцієнти апроксимації ВП сигнала $U_2(t)$;

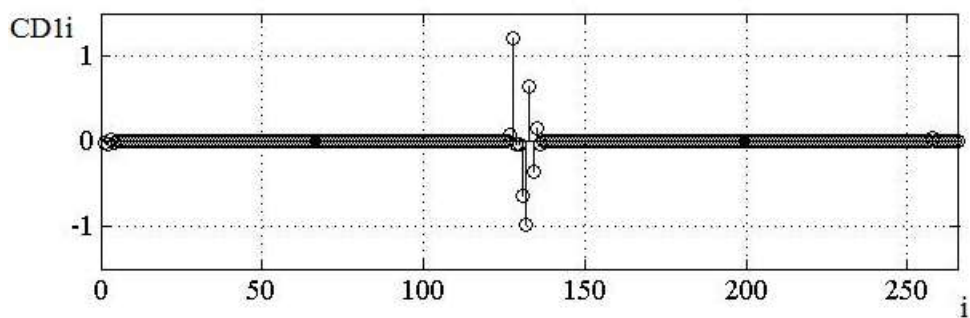
в – коефіцієнти деталізації ВП сигнала $U_2(t)$.



а



б



в

Рис. 2.5. ВП сигнала $U_3(t)$

а – сигнал $U_3(t)$; б – коефіцієнти апроксимації ВП сигнала $U_3(t)$;

в – коефіцієнти деталізації ВП сигнала $U_3(t)$.

2.2. Адаптивні вейвлет-фільтри Мейєра

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 напряду 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 Хуангом та іншими співавторами в публікації, 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 2π 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 ω_n як границі між кожним сегментом (де $\omega_0 = 0$ і $\omega_N = \pi$), див. Рис. 2.6.

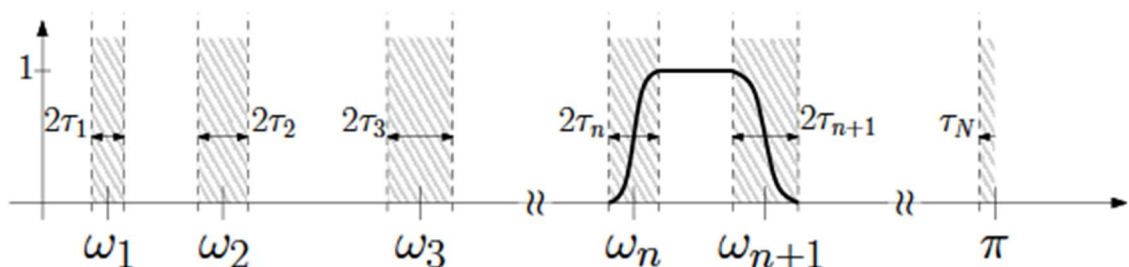


Рис. 2.6. Розділення спектру Фур'є

Кожен сегмент спектру позначається як $\Lambda_n = [\omega_{n-1}, \omega_n]$, тоді очевидно, що $\cup_{n=1}^N \Lambda_n = [0, \pi]$. У центрі кожного ω_n , ми визначаємо перехідну фазу (сірі заштриховані області на Рис. 2.6) T_n шириною $2\tau_n$. Емпіричні вейвлет-функції визначаються як смугові фільтри для кожного сегмента спектру Λ_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 (2.10) і (2.11) відповідно.

$$\hat{\phi}_n(\omega) = \begin{cases} 1, & |\omega| \leq \omega_n - \tau_n \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_n} (|\omega| - \omega_n + \tau_n) \right) \right], & \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0, & otherwise \end{cases} \quad (2.10)$$

і

$$\hat{\psi}_n(\omega) = \begin{cases} 1, & \omega_n + \tau_n \leq |\omega| \leq \omega_{n+1} - \tau_{n+1} \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_{n+1}} (|\omega| - \omega_{n+1} + \tau_{n+1}) \right) \right], & \omega_{n+1} - \tau_{n+1} \leq |\omega| \leq \omega_{n+1} + \tau_{n+1} \\ \sin \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_n} (|\omega| - \omega_n + \tau_n) \right) \right], & \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n \\ 0, & otherwise \end{cases} \quad (2.11)$$

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

$$\beta(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x \geq 1 \end{cases} \text{ і } \beta(x) + \beta(1-x) = 1 \quad \forall x \in [0,1]. \quad (2.12)$$

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

$$\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3). \quad (2.13)$$

Що стосується вибору τ_n , то можливо кілька варіантів. Найпростіший – вибрати τ_n пропорційно до ω_n : $\tau_n = \gamma\omega_n$ де $0 < \gamma < 1$. Отже, $\forall n > 0$, рівняння (2.10) і (2.11) спрощуються до рівнянь (2.14) і (2.15) відповідно

$$\hat{\phi}_n(\omega) = \begin{cases} 1, & |\omega| \leq (1-\gamma)\omega_n \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\gamma\omega_n} (|\omega| - (1-\gamma)\omega_n) \right) \right], & (1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n \\ 0, & otherwise \end{cases} \quad (2.14)$$

$$\hat{\psi}_n(\omega) = \begin{cases} 1, & (1 + \gamma)\omega_n \leq |\omega| \leq (1 - \gamma)\omega_{n+1} \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\gamma\omega_{n+1}} (|\omega| - (1 - \gamma)\omega_{n+1}) \right) \right], & (1 - \gamma)\omega_{n+1} \leq |\omega| \leq (1 + \gamma)\omega_{n+1} \\ \sin \left[\frac{\pi}{2} \beta \left(\frac{1}{2\gamma\omega_n} (|\omega| - (1 - \gamma)\omega_n) \right) \right], & (1 - \gamma)\omega_n \leq |\omega| \leq (1 + \gamma)\omega_n \\ 0, & \text{otherwise} \end{cases} \quad (2.15)$$

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. 2.7.

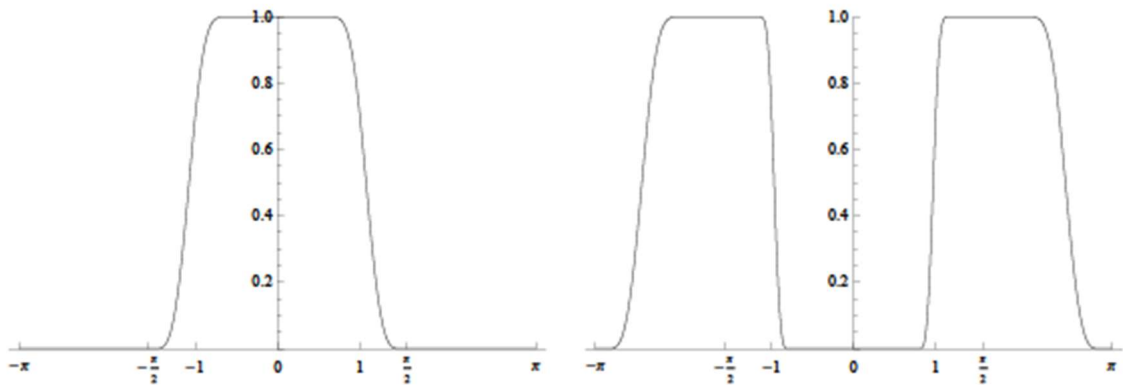


Рис. 2.7. Зліва: перетворення Фур'є функції масштабування для $\nu_n = 1$, $\gamma = 0.5$. Справа: перетворення Фур'є вейвлет-функції для $\nu_n = 1$, $\nu_{n+1} = 2.5$, $\gamma = 0.2$.

2.3. Сегментація спектра Фур'є

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 π , 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 π 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 π , ми визначаємо границі ω_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.

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 $\{M_i\}_{k=1}^M$. Припустимо, що ця множина відсортована в порядку спадання значень ($M_1 \geq M_2 \geq \dots M_M$) 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 α відповідає відносному співвідношенню амплітуд. Завдання полягає в тому, щоб підібрати таке α , 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.

2.4. Структура кадру вейвлет-фільтра Мейера

Наступне твердження показує, що при відповідному виборі параметра γ , 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 $\{\phi_1(t), \{\psi_n(t)\}_{n=1}^N\}$ є щільною структурою кадру $L^2(\mathbb{R})$.

Доведення. Ми дотримуємося ідеї побудови вейвлет-функції Мейера.

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

$$\sum_{k=-\infty}^{+\infty} \left(|\hat{\phi}_1(\omega + 2k\pi)|^2 + \sum_{n=1}^N |\hat{\psi}_n(\omega + 2k\pi)|^2 \right) = 1. \quad (2.16)$$

Згідно з періодичністю 2π (див. Рис. 2.3), достатньо зосередитися на інтервалі $[0, 2\pi]$.

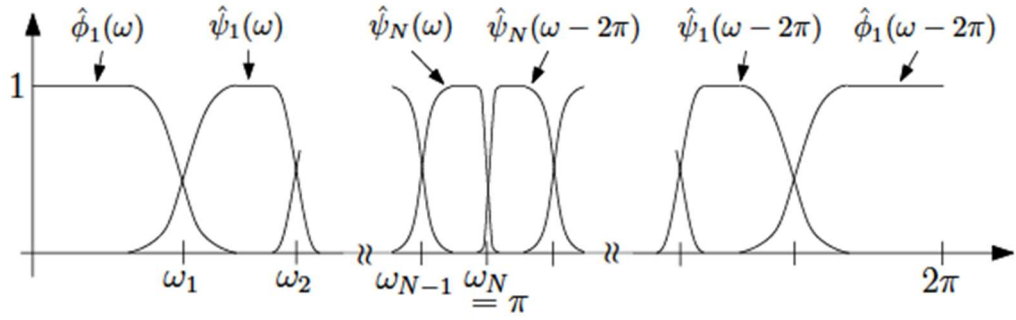


Рис. 2.8. Періодичність банку фільтрів

Дотримуючись попередніх визначень, ми можемо записати наступний вираз

$$[0, 2\pi] = \bigcup_{n=1}^N \Lambda_n \cup \bigcup_{n=1}^N \Lambda_{\sigma(n)}, \quad (2.17)$$

де $\Lambda_{\sigma(n)}$ є копією Λ_n але центрованою на $2\pi - \nu_n$ замість ν_n . По-перше, це легко побачити згідно виразів (2.18) і (2.19), тобто що для

$$\omega \in \left(\frac{\bigcup_{n=1}^N \Lambda_n}{\bigcup_{n=1}^N T_n} \right) \cup \left(\frac{\bigcup_{n=1}^N \Lambda_{\sigma(n)}}{\bigcup_{n=1}^N T_{\sigma(n)}} \right), \quad (2.18)$$

ми маємо

$$|\hat{\phi}_1(\omega)|^2 + |\hat{\phi}_1(\omega - 2\pi)|^2 + \sum_{n=1}^N (|\hat{\psi}_n(\omega)|^2 + |\hat{\psi}_n(\omega - 2\pi)|^2) = 1. \quad (2.19)$$

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

$$\tau_n + \tau_{n+1} < \omega_{n+1} - \omega_n, \quad (2.20)$$

$$\Leftrightarrow \gamma\omega_n + \gamma\omega_{n+1} < \omega_{n+1} - \omega_n, \quad (2.21)$$

$$\Leftrightarrow \gamma < \frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n}. \quad (2.22)$$

Умова (2.22) має виконуватися для всіх n , а також, для найменшого T_n , що є рівнозначним, внаслідок, ми отримаємо необхідний результат, якщо

$$\gamma < \min_n \frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n}. \quad (2.23)$$

На Рис. 2.9 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 $\gamma < 0.057$).

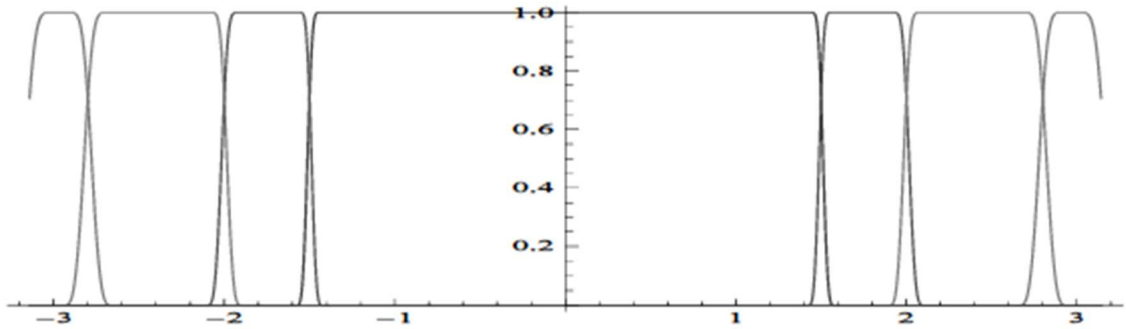


Рис. 2.4. Приклад розділення спектра Фур'є банком емпіричних вейвлет-фільтрів

2.5. Емпіричне вейвлет-перетворення

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:

$$W_f^\varepsilon(n, t) = \langle f, \psi_n \rangle = \int f(\tau) \overline{\psi_n(\tau - t)} d\tau = \left(\hat{f}(\omega) \overline{\hat{\psi}_n(\omega)} \right)^\vee, \quad (2.24)$$

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

$$W_f^\varepsilon(0, t) = \langle f, \phi_1 \rangle = \int f(\tau) \overline{\phi_1(\tau - t)} d\tau = \left(\hat{f}(\omega) \overline{\hat{\phi}_1(\omega)} \right)^\vee, \quad (2.25)$$

де $\hat{\psi}_n(\omega)$ і $\hat{\phi}_1(\omega)$ визначаються рівняннями 2.15 і 2.14 відповідно.

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

$$f(t) = W_f^\varepsilon(0, t) \phi_1(t) + \sum_{n=1}^N W_f^\varepsilon(n, t) \psi_n(t) = \quad (2.26)$$

$$= \left(\widehat{W}_f^\varepsilon(0, \omega) \widehat{\phi}_1(\omega) + \sum_{n=1}^N \widehat{W}_f^\varepsilon(n, \omega) \widehat{\psi}_n(\omega) \right)^\vee.$$

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

$$f_k(t) = F_k(t) \cos(\varphi_k(t)), \quad (2.27)$$

де $F_k(t), \dot{\varphi}_k(t) > 0 \forall t$, та є такими, що

$$f(t) = \sum_{k=0}^N f_k(t), \quad (2.28)$$

задається формулами

$$f_0(t) = W_f^\varepsilon(0, t) \phi_1(t), \quad f_k(t) = W_f^\varepsilon(k, t) \psi_k(t). \quad (2.29)$$

Висновки до другого розділу

1. Запропоновано підхід, для кількісного виміру семантичної інформації, яка знаходиться в мовних сигналах, на сьогоднішній день немає остаточного теоретичного обґрунтування проблеми пов'язаної з доведенням можливості кількісного виміру семантичної інформації.

2. 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. Доведено адаптивність емпіричного вейвлет-перетворення за допомогою сегментації спектра Фур'є, де компактні носії вейвлет-фільтрів Мейера напряму залежать від того, де знаходиться потрібна нам семантична інформація в спектрі мовного сигналу.

Список використаних джерел у другому розділі

- [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.
- [2] Г. Ф. Конахович, 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.
- [3] J. Gilles, "Empirical Wavelet Transform," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 3999-4010, 2013.
- [4] 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.
- [5] R. R. Coifman, M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 713-718, 1992.
- [6] 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.
- [7] I. Daubechies, J. Lu, H-T. Wu, "Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool," *Journal of Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243-261, 2011.
- [8] 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.

- [9] А. Ю. Лавриненко, Ю. А. Кочергин, Г. Ф. Конахович, “Система защищённого голосового управления беспилотным летательным аппаратом,” *Авиационно-космическая техника и технология*, Т. 148, № 4, С. 4-11, 2018.
- [10] P. Flandrin, P. Goncalves, “Empirical mode decompositions as data-driven wavelet-like expansions,” *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 2, no. 4, pp. 477-496, 2004.
- [11] Г. Ф. Конахович, Д. І. Бахтіяров, О. Ю. Лавриненко, “Комп’ютерне моделювання захищеного каналу керування безпілотним літальним апаратом,” *Наукоємні технології*, Т. 28, № 4, С. 283-290, 2015.
- [12] Н. М. Астафьева, “Вейвлет-анализ: основы теории и примеры применения,” *Успехи физических наук*, Т. 166, № 11, С. 1145-1170, 1996.
- [13] 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.
- [14] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674-693, 1989.
- [15] 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.
- [16] А. Ю. Лавриненко, Г. Ф. Конахович, Р. С. Одарченко, Д. И. Бахтияров, “Алгоритм сжатия сигналов речевых команд управления функциями беспилотного летательного аппарата,” *Авиационно-космическая техника и технология*, Т. 130, № 3, С. 51-61, 2016.
- [17] P. Flandrin, G. Rilling, P. Goncalves, “Empirical mode decomposition as a filter bank,” *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112-114, 2004.
- [18] O. Yu. Lavrynenko, 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.

[19] M. J. Shensa, "The discrete wavelet transform: wedding the a trous and Mallat algorithms," *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2464-2482, 1992.

[20] O. Lavrynenko, G. Konakhovych, D. Bakhtiarov, "Comparative analysis of speech recognition algorithms in UAV voice control system," *Science-Based Technologies*, vol. 38, no. 2, pp. 137-142, 2018.

[21] О. Ю. Лавриненко, Г. Ф. Конахович, 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, № и 2017 03401, заявл. 10.04.2017, опубл. 25.09.2017, бюл. № 18, 6 с.

РОЗДІЛ 3

МЕТОД СЕМАНТИЧНОГО КОДУВАННЯ МОВНИХ СИГНАЛІВ НА ОСНОВІ ЕМПІРИЧНОГО ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ

3.1. Структурна схема запропонованого методу і первинна обробка мовного сигналу

Метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення включає в себе наступні кроки (див. Рис. 3.1).

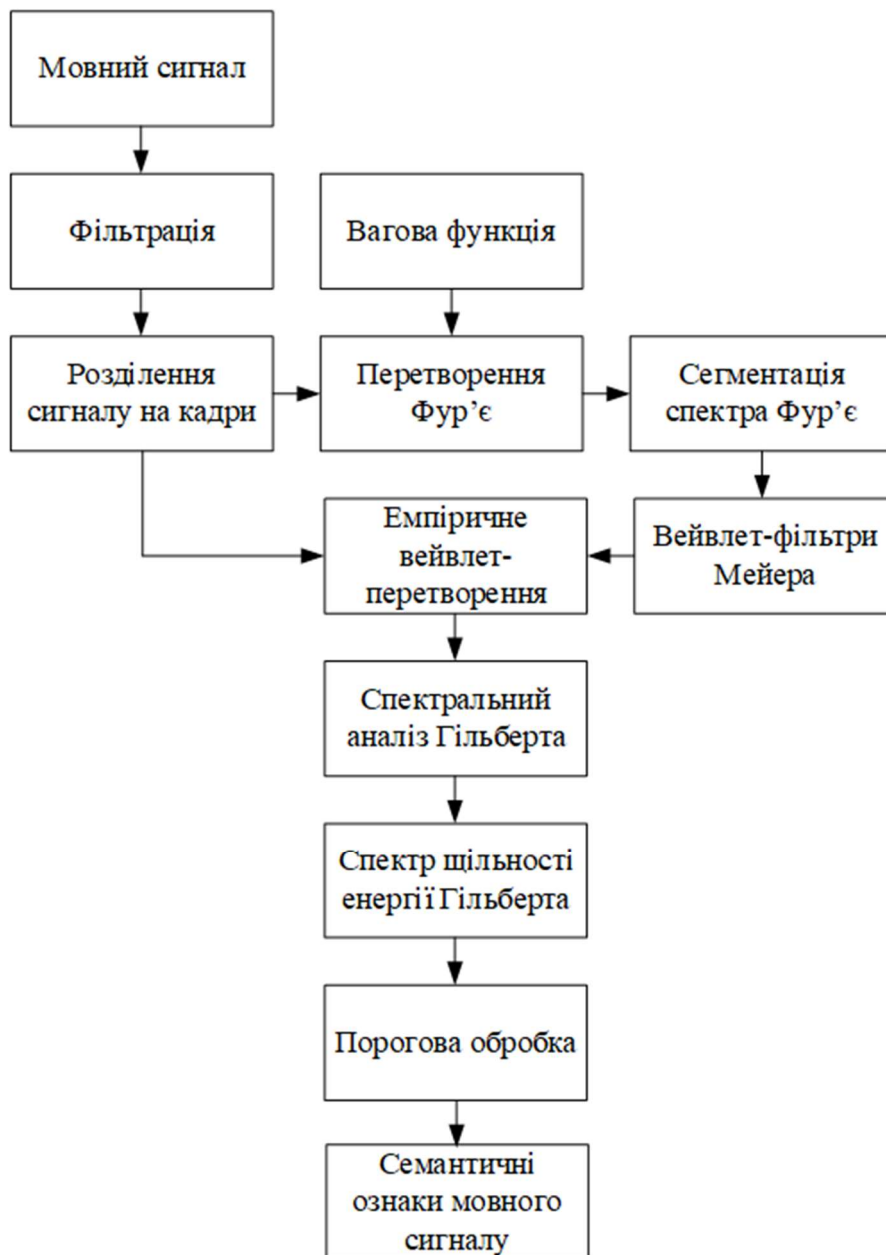


Рис. 3.1. Метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення

Попередня обробка мовного сигналу є важливим моментом в системах розпізнавання мови орієнтованих на вирішення завдання управління голосом. За допомогою такої обробки, по-перше, досягається істотне стиснення інформації за рахунок частотного аналізу мовного коливання, що закінчується отриманням короткочасного амплітудного спектра сигналу з інтервалом 5-20 мс. Для цих цілей зазвичай використовується гребінка з порівняно невеликої кількості смугових фільтрів. По-друге, створюється стійке до рівня сигналу, частотним спотворень і фоновим шумів первинне опис мовного сигналу, необхідних для виявлення його семантичних ознак, що використовуються для розпізнавання фонем. В результаті розпізнавання фонем безперервний мовний сигнал представляється у вигляді послідовності дискретних елементів - фонем, необхідних для подальшого розпізнавання більших фрагментів мови, прикладами яких є склади, слова і фрази.

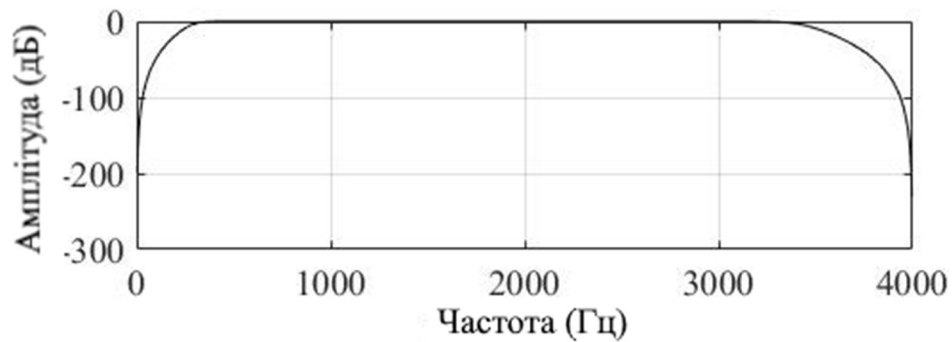
На вхід алгоритму подається мовний сигнал частотний діапазон, якого вельми обмежений і розташовується в інтервалі від 300 до 3400 Гц. З даного факту випливає, що шляхом застосування смугового фільтра можна відкинути частотні складові, які знаходяться за межами цього діапазону і відповідно, не несуть смислового навантаження [28].

Для зниження рівня шуму і нормалізації смуги частот оцифрований мовний сигнал надходить на смуговий фільтр Баттерворта $2n$ -го порядку, де $n = 5$ зі смугою пропускання 300 ... 3400 Гц, який представляється у вигляді векторів-строк b і a , які мають довжину $2n + 1$ та містять коефіцієнти поліномів чисельника і знаменника функції передачі в порядку спадання степенів змінної z :

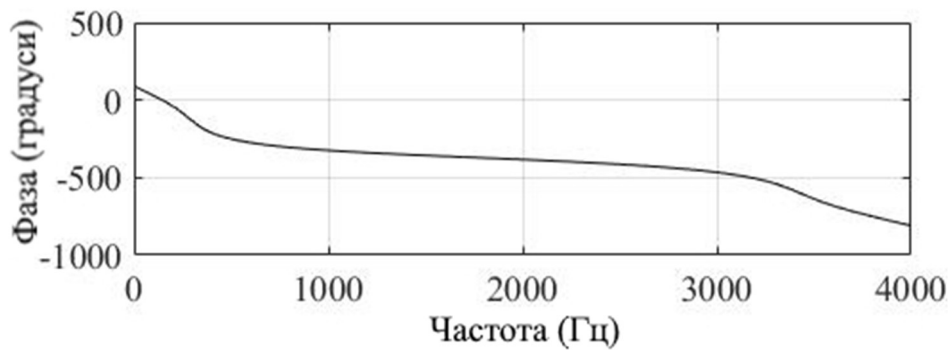
$$H(z) = \frac{B(z)}{A(z)} = \frac{b(1) + b(2)z^{-1} + \dots + b(n+1)z^{-n}}{1 + a(2)z^{-1} + \dots + a(n+1)z^{-n}}. \quad (3.1)$$

Частотою зрізу фільтра Баттерворта називається частота, на якій модуль коефіцієнта передачі дорівнює $\sqrt{1/2}$. На Рис. 3.2 зображені графіки амплітудно-частотної характеристики (АЧХ), фазово-частотної

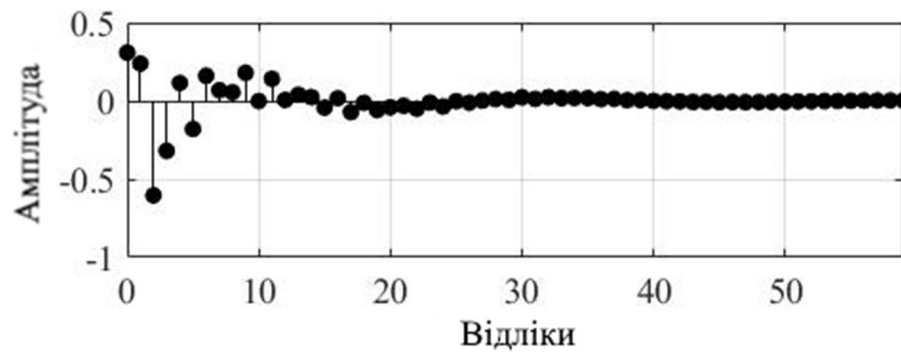
характеристики (ФЧХ), а також імпульсної характеристики (ІХ) синтезованого фільтра Баттерворта 10-го порядку зі смугою пропускання 300 ... 3400 Гц [29].



а



б



в

Рис. 3.2. а – АЧХ, б – ФЧХ, в – ІХ фільтра Баттерворта 10-го порядку зі смугою пропускання 300 ... 3400 Гц

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 інформативних ознак розпізнавання однакової довжини, потрібно розбити мовний сигнал на рівні кадри, а потім виконувати перетворення, вважаючи, що сигнал на такому відрізку приблизно стаціонарний (див. Рис. 3.3) [30-32].

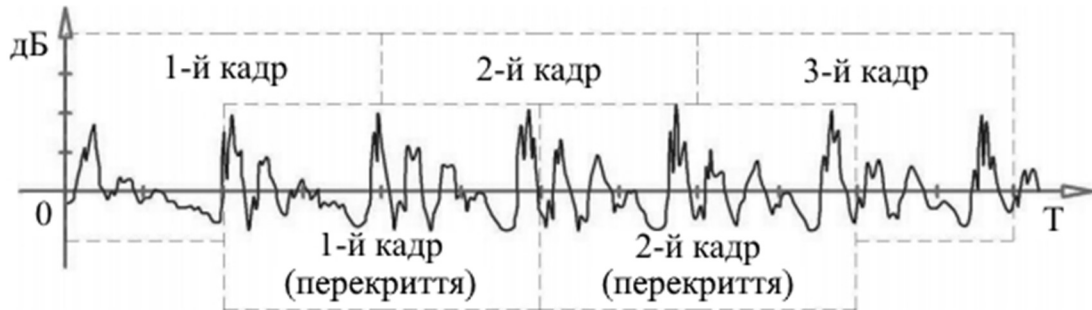


Рис. 3.3. Розділення мовного сигналу на кадри

Для мовного сигналу розмір кадру зазвичай вибирається в межах 10-20 мс. Для більш точного подання сигналу між кадрами роблять перекриття, що дорівнює половині довжини кадру. Перекриття кадрів використовується для запобігання втрати інформації про сигнал на границі. Чим менше перекриття, тим меншою розмірністю в результаті буде мати набір ознак розпізнавання, характерний для даної ділянки мовного сигналу. Потім до кожного кадру застосовується алгоритм виділення семантичних ознак мовного сигналу (МЧКК) [33].

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

На вхід блоку дискретного перетворення Фур'є (ДПФ) подається послідовність відліків ділянки 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 Фур'є аналізі, викликаних скінченністю вибірки [34].

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

$$w[n] = 0.53836 - 0.46164 \times \cos\left(2\pi \frac{n}{N-1}\right), \quad n = 0, \dots, N-1, \quad (3.2)$$

де, N – довжина вікна, виражена у відліках.

Тоді ДПФ мовного сигналу з попередньою обробкою ваговою функцією можна записати у вигляді формули:

$$X[k] = \sum_{n=0}^{N-1} x[n] w[n] e^{\frac{-2\pi j}{N}kn}, \quad k = 0, \dots, N-1. \quad (3.3)$$

Значення індексів k відповідають частотам:

$$f[k] = \frac{F_s}{N}k, \quad k = 0, \dots, N/2, \quad (3.4)$$

де, F_s – частота дискретизації мовного сигналу [35-37].

3.2. Емпіричне вейвлет-перетворення сегментованого спектра Фур'є

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 $[0; 1]$. 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 α відповідає відносному співвідношенню амплітуд. Завдання полягає в тому, щоб підібрати таке α , 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.

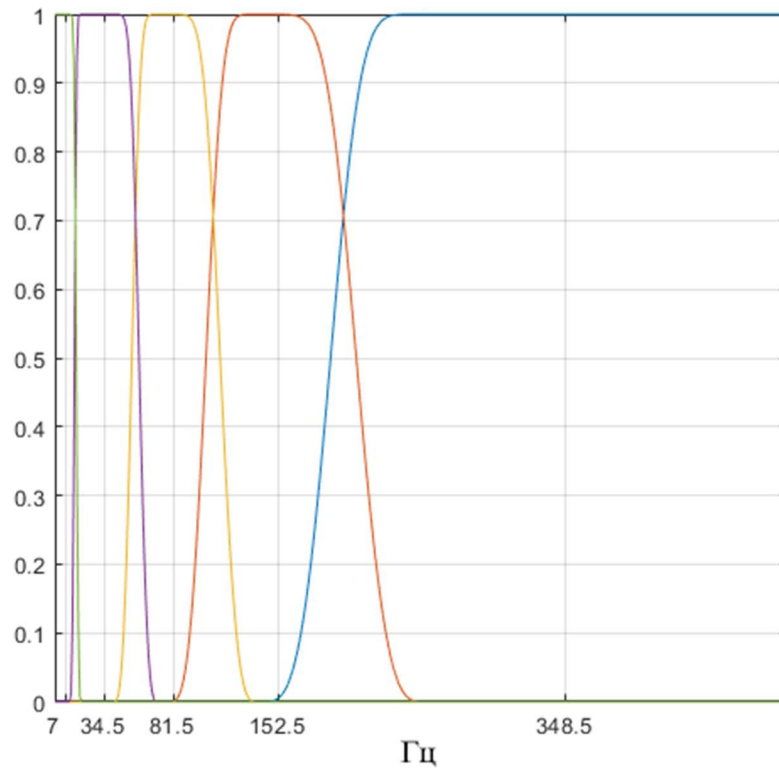


Рис. 3.4. Емпіричний вейвлет-фільтр Мейера

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:

$$W_f^\varepsilon(n, t) = \langle f, \psi_n \rangle = \int f(\tau) \overline{\psi_n(\tau - t)} d\tau = \left(\hat{f}(\omega) \overline{\hat{\psi}_n(\omega)} \right)^\vee, \quad (3.5)$$

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

$$W_f^\varepsilon(0, t) = \langle f, \phi_1 \rangle = \int f(\tau) \overline{\phi_1(\tau - t)} d\tau = \left(\hat{f}(\omega) \overline{\hat{\phi}_1(\omega)} \right)^\vee, \quad (3.6)$$

де $\hat{\psi}_n(\omega)$ і $\hat{\phi}_1(\omega)$ визначаються рівняннями 2.15 і 2.14 (див. Рис. 3.4).

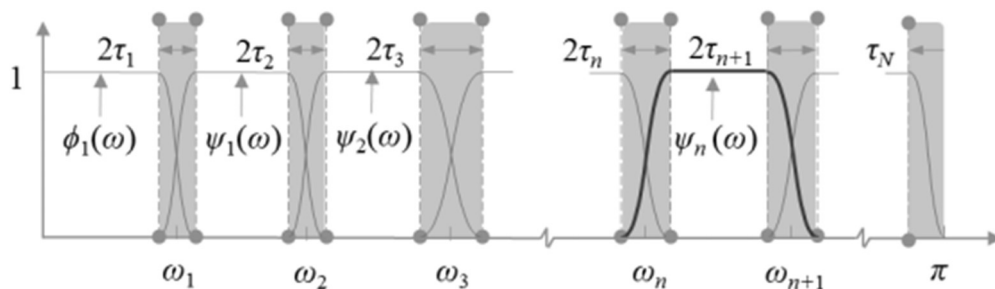


Рис. 3.5. Основна конструкція емпіричного вейвлет-перетворення

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

$$f_k(t) = F_k(t) \cos(\phi_k(t)), \quad (3.7)$$

де $F_k(t), \phi_k(t) > 0 \forall t$, та є такими, що

$$f(t) = \sum_{k=0}^N f_k(t), \quad (3.8)$$

задається формулами

$$f_0(t) = W_f^\varepsilon(0, t)\phi_1(t), \quad (3.9)$$

$$f_k(t) = W_f^\varepsilon(k, t)\psi_k(t). \quad (3.10)$$

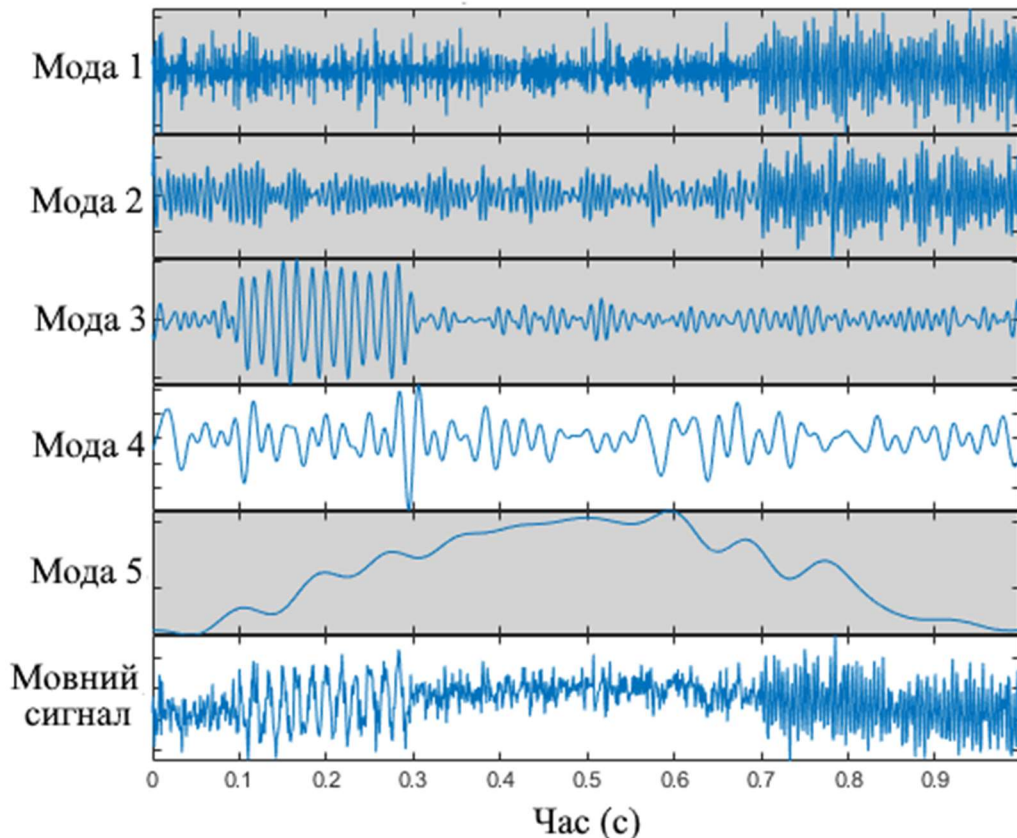


Рис. 3.6. Розкладання мовного сигналу на внутрішні емпіричні моди

3.3. Спектральний аналіз Гільберта

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

$$y(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau, \quad (3.11)$$

де P – головне значення Коші сингулярного інтеграла.

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

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)}, \quad (3.12)$$

де $i = (-1)^{1/2}$.

Тоді $a(t)$ можна виразити як

$$a(t) = \sqrt{(x^2 + y^2)}, \quad (3.13)$$

де $a(t)$ – миттєва амплітуда.

Миттєва фазова функція може бути виражена наступним чином

$$\theta(t) = \arctan \frac{y}{x}. \quad (3.14)$$

Миттєва частота визначається виразом

$$\omega = \frac{d\theta}{dt}. \quad (3.15)$$

Застосовуючи ПГ до окремих компонентів внутрішніх емпіричних мод, вихідні дані можуть бути виражені як дійсна частина наступного рівняння [3]

$$x(t) = \operatorname{Re} \left\{ \sum_{j=1}^n a_j(t) \exp \left[i \int \omega_j(t) dt \right] \right\}. \quad (3.16)$$

Рівняння (3.16) визначає дійсну частину амплітуди (3.13) і частоти (3.15) 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 визначається як

$$S_{i,j} = H(t_i, \omega_j) = \frac{1}{\Delta t \times \Delta \omega} H \left[\sum_{k=1}^n a_k^2(t) \right]. \quad (3.17)$$

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

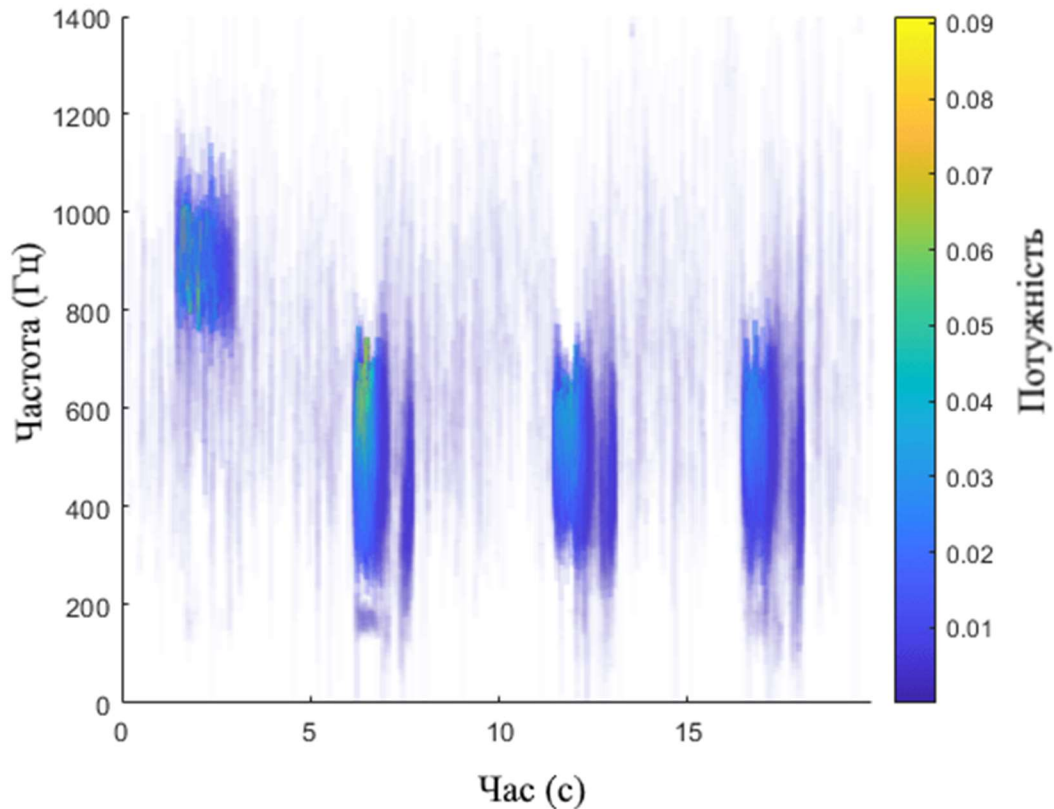


Рис. 3.7. Спектр щільності енергії Гільберта мовного сигналу

Таким чином, застосувавши перетворення Гільберта до модової функції емпіричного вейвлет-перетворення, отриманої в результаті декомпозиції мовного сигналу, одержали частотний спектр Гільберта, що залежить від частоти і часу. Інтеграл цієї величини за часом дає інтегральний спектр Гільберта - аналог спектра Фур'є.

Перетворення Гільберта та емпіричне вейвлет-перетворення відкривають для аналізу мовних сигналів нові можливості в детальному аналізі частотної і часової структури спектра мови.

3.4. Порогова обробка спектра Гільберта

На практиці широко використовуються дві порогові функції:

1) порогова функція $T_H(\tilde{d}, \lambda)$ виду:

$$T_H(\tilde{d}, \lambda) = \begin{cases} 0, & |\tilde{d}| \leq \lambda \\ \tilde{d}, & |\tilde{d}| > \lambda \end{cases} \quad (3.18)$$

2) порогова функція $T_S(\tilde{d}, \lambda)$ виду:

$$T_S(\tilde{d}, \lambda) = \begin{cases} 0, & |\tilde{d}| \leq \lambda \\ \text{sign}(\tilde{d}) \times [|\tilde{d}| - \lambda], & |\tilde{d}| > \lambda \end{cases} \quad (3.19)$$

де λ - величина порога, \tilde{d} - оброблюваний коефіцієнт розкладання. Графіки функцій (3.18), (3.19) наведені на Рис. 3.8 для $\lambda = 1$ (1 - графік функції (3.18), 2 - функція (3.19)).

Відзначимо характерні особливості цих функцій:

1) через зменшення амплітуди коефіцієнта розкладання на величину λ в функції $T_S(\tilde{d}, \lambda)$ можливо згладжування контрастних елементів оброблюваного сигналу, особливо при великих значеннях λ ;

2) наявність у функції $T_H(\tilde{d}, \lambda)$ розриву в оточенні λ може викликати появу осциляцій (ефект Гіббса) в оброблюваному сигналі.

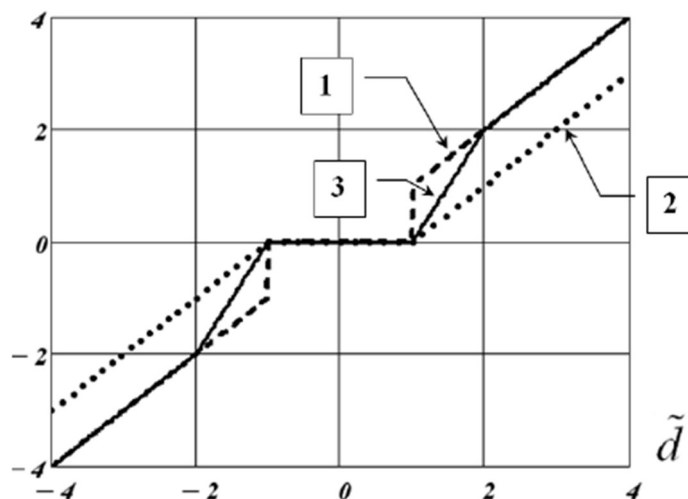


Рис. 3.8. Графіки порогових функцій (3.18), (3.19), (3.20)

Для подолання цих недоліків були запропоновані двопараметричні функції, які зараз будуть розглянуті.

1. Порогова функція $T_{SS}(\tilde{d}, \lambda_1, \lambda_2)$ виду:

$$T_{SS}(\tilde{d}, \lambda_1, \lambda_2) = \begin{cases} 0, & |\tilde{d}| \leq \lambda_1 \\ \text{sign}(\tilde{d}) \times \left[\frac{\lambda_2(|\tilde{d}| - \lambda_1)}{\lambda_2 - \lambda_1} \right], & \lambda_1 < |\tilde{d}| \leq \lambda_2 \\ \tilde{d}, & |\tilde{d}| > \lambda_2 \end{cases} \quad (3.20)$$

яка включає вже дві порогові величини λ_1, λ_2 . Графік цієї функції (при $\lambda_1 = 1, \lambda_2 = 2$) наведено на Рис. 3.8 (крива 3).

В роботі [7] були досліджені статистичні характеристики помилок фільтрації для трьох порогових функцій (3.18), (3.19), (3.20). Було показано (аналітично і в обчислювальному експерименті), що функція (3.18) має найменше зміщення, функція (3.19) - найменшу дисперсію, функція (3.20) - найменше середньоквадратичне відхилення фільтрації.

2. Порогова функція $T_Z(\tilde{d}, \lambda_1, \lambda_2)$, визначається виразом:

$$T_Z(\tilde{d}, \lambda_1, \lambda_2) = \begin{cases} 0, & |\tilde{d}| \leq \lambda_1 \\ \frac{\tilde{d}}{e - 1} \times \left[e^{\frac{|\tilde{d}| - \lambda_1}{\lambda_2 - \lambda_1}} - 1 \right], & \lambda_1 < |\tilde{d}| \leq \lambda_2 \\ \tilde{d}, & |\tilde{d}| > \lambda_2 \end{cases} \quad (3.21)$$

На Рис. 2 представлені графіки функції (3.21) (крива 1) і функції (3.20) (крива 2), побудовані при $\lambda_1 = 1, \lambda_2 = 2$. Видно, що на інтервалі $[\lambda_1, \lambda_2]$ функція (3.21) відрізняється від прямої (яка присутня в функції (3.20)).

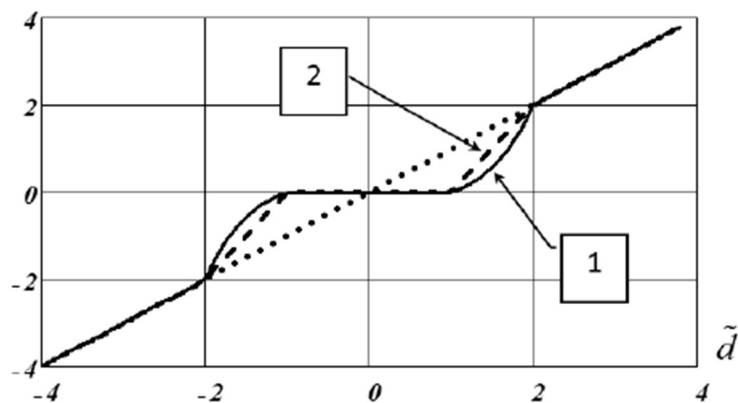


Рис. 3.9. Графіки порогових функцій (3.20), (3.21)

3. Комбінована порогова функція $T_{COM}(\tilde{d}, \lambda, \alpha)$, визначається виразом:

$$T_{COM}(\tilde{d}, \lambda, \alpha) = \begin{cases} 0, & |\tilde{d}| \leq \lambda \\ (1 - \mu) \times T_H(\tilde{d}, \lambda) + \mu T_S(\tilde{d}, \lambda), & |\tilde{d}| > \lambda \end{cases} \quad (3.22)$$

де $\mu = \alpha(|\tilde{d}| - \lambda)^2$. Ця порогова функція є лінійна комбінація порогових функцій (3.18) и (3.19), де вона також включає два параметри λ і α , при цьому $0 \leq \alpha \leq 1$. Очевидно, що при $\alpha = 0$ значення $\mu = 0$ і $T_{COM}(\tilde{d}, \lambda, 0) = T_H(\tilde{d}, \lambda)$. Якщо $\alpha = 1$, то $\mu = 1$ і $T_{COM}(\tilde{d}, \lambda, 1) = T_S(\tilde{d}, \lambda)$. На Рис. 3.10 наведені графіки функції $T_{COM}(\tilde{d}, 3, \alpha)$ при різних значеннях параметра α : крива 1 - $\alpha = 0.001$; крива 2 - $\alpha = 0.5$; крива 3 - $\alpha = 0.99$. Ці криві добре ілюструють зміну порогової функції при варіації параметра α .

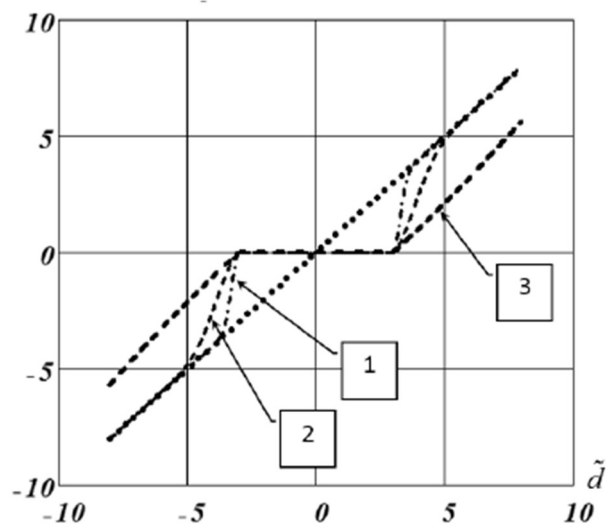


Рис. 3.10. Графіки порогової функції (3.22) при різних значеннях параметра α

3.5. Вибір оптимальної порогової функції

Вибір оптимальної порогової функції проводиться за наступним алгоритмом.

Вихідні дані формувалися як

$$\tilde{f}_i = f_i + \eta_i, \quad i = 1, 2, \dots, N, \quad (3.23)$$

де \tilde{f}_i - значення функції, η_i - псевдовипадкові числа (шум вимірювання), підпорядковані нормальному розподілу з нульовим середнім і дисперсією σ^2 . Значення дисперсії задавалося через відносний рівень $\delta_\eta = \|\eta\|/\|f\|$ шуму

вимірювання, де $\|\eta\|$ і $\|f\|$ - евклідові норми відповідних векторів. Точність вейвлет-фільтрації визначалася величиною відносної помилки:

$$\delta_f(T) = \frac{\|\hat{f}(T) - f\|}{\|f\|}, \quad (3.24)$$

де вектор $\hat{f}(T)$ - результат вейвлет-фільтрації з пороговою функцією T . Очевидно, що $\delta_f(T)$ є випадковою величиною і тому за вибіркою обчислювалася вибіркова оцінка математичного очікування цієї випадкової величини:

$$\bar{\delta}_f(T) = \frac{1}{N_s} \times \sum_{l=1}^{N_s} \delta_f^{(l)}(T), \quad (3.25)$$

де N_s - об'єм вибірки, $\delta_f^{(l)}(T)$ - відносна помилка фільтрації l -ої реалізації даних $\tilde{f}^{(l)} = f + \eta^{(l)}$. Далі знаходимо мінімальну помилку фільтрації для кожної порогової функції шляхом мінімізації $\bar{\delta}_f(T)$ за двома параметрами порогової функції та відношення цих мінімальних помилок до мінімальної помилку порогової функції.

На практиці в якості оптимальної двопараметричної порогової функції можна прийняти функції (3.20) и (3.21), які еквівалентні по точності фільтрації і своєю структурою, залежать від двох параметрів λ_1, λ_2 і ці параметри істотно впливають на помилку фільтрації.

Проведемо вибір оптимальної функції порогової обробки за допомогою оцінки параметрів λ_1, λ_2 , порогової функції, що дозволяє знайти оптимальні пороги $\lambda_{1opt}, \lambda_{2opt}$, з мінімальним середньоквадратическим відхиленням, яке визначається виразом:

$$\Delta(\lambda_1, \lambda_2) = M \left[\|\hat{f}_{\lambda_1, \lambda_2} - f\|^2 \right], \quad (3.26)$$

де $M[]$ - оператор математичного очікування по щільності розподілу шуму, $f, \hat{f}_{\lambda_1, \lambda_2}$ - вектори, проєкції яких дорівнюють значенням сигналів (при заданих порогових величинах λ_1, λ_2).

Показано, що побудований алгоритм, дозволяє досить точно оцінити оптимальне значення порога λ_{opt} , при якому середньоквадратичне відхилення фільтрації мінімальне.

Оцінювання оптимальних порогових величин $\lambda_{1opt}, \lambda_{2op}$.

Припустимо, що:

1) замість точних значень f_i дискретного сигналу дані (відомі) зашумлені значення

$$\tilde{f}_i = f_i + \eta_i, \quad i = 1, 2, \dots, N, \quad (3.27)$$

де шум вимірювання η_i має нульове середнє $M[\eta_i] = 0$, дисперсію σ^2 і значення η_i, η_j не корельовані при $i \neq j$;

2) базисні функції $\{\varphi_{j,k}(t)\}, \{\psi_{j,k}(t)\}$ є ортонормованим, що відповідає ортогональним вейвлетам (вейвлети Мейера, Добеши, Сімлета, Кайфлета);

Тоді визначимо вектор нев'язки $e_{\lambda_1, \lambda_2} = \tilde{f} - \hat{f}_{\lambda_1, \lambda_2}$ і введемо статистику:

$$\rho_W(\lambda_1, \lambda_2) = \frac{1}{\sigma^2} \langle e_{\lambda_1, \lambda_2}, \tilde{f} \rangle = \frac{1}{\sigma^2} \sum_{i=1}^N (e_{\lambda_1, \lambda_2}) \times \tilde{f}_i. \quad (3.28)$$

Також як і в лінійних алгоритмах фільтрації в якості оцінок для $\lambda_{1opt}, \lambda_{2opt}$, приймемо значення $\lambda_{1W}, \lambda_{2W}$, для яких статистика $\rho_W(\lambda_{1W}, \lambda_{2W})$ задовольняє нерівності:

$$\vartheta_{m, \gamma/2} \leq \rho_W(\lambda_{1W}, \lambda_{2W}) \leq \vartheta_{m, 1-\gamma/2}, \quad (3.29)$$

де $\vartheta_{m, \gamma/2}, \vartheta_{m, 1-\gamma/2}$ квантилі χ_m^2 - розподіл з m ступенями свободи рівнів $\gamma/2, 1 - \gamma/2$ відповідно, γ - ймовірність помилки першого роду при перевірці статистичної гіпотези про оптимальність параметра згладжування (зазвичай $\gamma = 0.05$), $m = N$ - число значень фільтруемого сигналу (проекції вектора \tilde{f}). Якщо число ступенів свободи $m > 30$ (в задачах фільтрації це завжди виконується), то χ_m^2 - розподіл добре апроксимується нормальним розподілом з математичним очікуванням m і дисперсією $2m$. Тоді, прийнявши ймовірність помилки першого роду $\gamma = 0.05$, отримуємо формули для обчислення квантилів, що входять в нерівність (3.29)

$$\vartheta_{m,0.025} = m - 1.96\sqrt{2m}, \quad \vartheta_{m,0.975} = m + 1.96\sqrt{2m}. \quad (3.30)$$

Для обчислення оцінок λ_{1W} , λ_{2W} , порогові величини λ_1 , λ_2 , функцій (3.20), (3.21) визначимо у вигляді:

$$\lambda_1(\beta) = \beta \sqrt{2 \ln(N_j)}, \quad \lambda_2(\beta, C) = \beta \times C \sqrt{2 \ln(N_j)}, \quad (3.31)$$

де N_j - кількість оброблюваних коефіцієнтів j -го рівня, множник $C > 1$ впливає з нерівності $\lambda_2 > \lambda_1$ (див. (3.20), (3.21)). Зауважимо, що співмножник $\sqrt{2 \ln(N_j)}$ робить пороги рівнозалежними і забезпечує асимптотичну оптимальність порогових величин по порядку при $N_j \rightarrow \infty$. Отже, необхідно оцінити β_{opt} , C_{opt} , використовуючи статистику (3.28), тобто обчислити значення β_W , C_W , які задовольняють нерівності:

$$\vartheta_{m,\gamma/2} \leq \rho_W(\beta_W, C_W) \leq \vartheta_{m,1-\gamma/2}. \quad (3.32)$$

Тоді оцінки λ_{1W} , λ_{2W} визначаються виразами:

$$\lambda_{1W} = \beta_W \sqrt{2 \ln(N_j)}, \quad \lambda_{2W} = \beta_W \times C_W \sqrt{2 \ln(N_j)}. \quad (3.33)$$

для обчислення β_W , C_W замість вирішення нелінійного рівняння

$$\rho_W(\beta, C) = m, \quad (3.34)$$

що включає дві невідомі величини β , C , розглянемо задачу мінімізації функціонала

$$F(\beta, C) = |\rho_W(\beta, C) - m|^2, \quad (3.35)$$

Зауважимо, що рішення цієї задачі завжди існує і для його знаходження можна використовувати відомі процедури мінімізації. В якості β_W , C_W приймається такий елемент $\{\beta^{(n)}, C^{(n)}\}$ що мінімізує послідовності, для якого виконується нерівність (3.32).

Можна показати, що при використанні ортогональних вейвлетів критерій $\rho_W(\beta, C)$ обчислюється через коефіцієнти вейвлет-розкладання:

$$\rho_W(\beta, C) = \frac{1}{\sigma^2} \sum_{j=j_0+1}^{j_0+J} \sum_k \tilde{d}_{j,k} \times \left(\tilde{d}_{j,k} - T \left(\tilde{d}_{j,k}, \lambda_1(\beta), \lambda_2(\beta, C) \right) \right). \quad (3.36)$$

Це дозволяє знаходити значення критерію (при реалізації процедури мінімізації) в просторі коефіцієнтів вейвлет-розкладання, а потім (при знайдених β_W , C_W і обчислених $\hat{d}_{j,k}$) тільки один раз виконати зворотне вейвлет-перетворення і отримати згладжені значення функції.

Відзначимо деякі властивості $\rho_W(\beta, C)$, які виходять з (3.36):

1) всі складові, що входять в формулу (3.36), не від'ємні (можуть змінюватися від 0 до $\tilde{d}_{j,k}^2$) і тому $\rho_W(\beta, C) \geq 0$;

2) при $\beta \rightarrow 0$ і $C < \infty$ справедлива границя $\rho_W(\beta, C) \rightarrow 0$;

3) при $\beta \rightarrow \infty$ і $C < \infty$ справедлива границя

$$\rho_W(\beta, C) \rightarrow \frac{1}{\sigma^2} \sum_{j=j_0+1}^{j_0+J} \sum_k \tilde{d}_{j,k}^2 = \frac{1}{\sigma^2} \|\tilde{f}\|^2. \quad (3.37)$$

Остання рівність має місце для ортогональних вейвлетів при відповідного нормуванні базисних функцій. Дві останні властивості обумовлюють наступне твердження.

Твердження. Якщо виконується нерівність

$$\rho_W(\infty, C) = \frac{1}{\sigma^2} \sum_{i=1}^N \tilde{f}_i^2 > \vartheta_{m,1-\gamma/2}, \quad (3.38)$$

то існують кінцеві значення β_W , C_W , для яких виконується нерівність (3.32). Невиконання умови (3.32) означає, що значення $\tilde{f}_i = \eta_i$, тобто $f_i \equiv 0$. В цьому випадку $\beta_W = \infty$ і згладжені значення рівні 0.

Суттєвою рисою наведеного алгоритму обчислення θ_W є використання дисперсії шуму σ^2 . На практиці, як правило, ця величина невідома, і в цьому випадку можна використовувати оцінку для середньоквадратичного відхилення σ :

$$\hat{\sigma} = \frac{\text{median}(|\tilde{d}_{1,k}|)}{0.6745}, \quad (3.39)$$

де оператор $\text{median}(|\tilde{d}_{1,k}|)$ обчислює медіану абсолютних величин деталізуючих коефіцієнтів рівня розкладання $j_0 + 1$ (об'єм вибірки дорівнює $N/2$). Ця оцінка широко використовується в робастних алгоритмах регресійного аналізу. Стосовно до алгоритмів вейвлет-фільтрації ця оцінка досліджувалася, де була показана прийнятна точність, а саме, для заданої дисперсії $\sigma^2 = 0.91 \times 10^{-1}$ значення оцінки (3.39), обчислені за 30 реалізаціями довжиною $N/2 = 1024$ знаходилися в інтервалі $[0.88 \times 10^{-1}, 0.97 \times 10^{-1}]$.

Висновки до третього розділу

1. 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. Проведено адаптивну порогову обробку мовного сигналу з розрахунком оптимальних значень порогів функції вейвлет-фільтрації, для відсіювання коефіцієнтів які характеризують амплітуди і частоти спектра Гільберта малої потужності.

3. Проведено вибір оптимальної функції порогової обробки за допомогою оцінки параметрів λ_1, λ_2 , порогової функції вейвлет-фільтрації, що дозволяє знайти оптимальні пороги $\lambda_{1opt}, \lambda_{2opt}$, з мінімальним середньоквадратическим відхиленням $\Delta(\lambda_1, \lambda_2)$, що дозволить підвищити ефективність визначення семантичних ознак мовного сигналу.

Список використаних джерел в третьому розділі

- [1] Г. Ф. Конахович, 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.
- [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.
- [3] А. Ю. Лавриненко, Ю. А. Кочергин, Г. Ф. Конахович, “Система захищеного голосового управління беспилотным летательным аппаратом,” *Авиационно-космическая техника и технология*, Т. 148, № 4, С. 4-11, 2018.
- [4] M. E. Torres, M. A. Colominas, G. Schlotthauer, P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise,” *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, May 22-27, 2011, pp. 4144-4147.
- [5] 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.
- [6] F. G. Meyer, R. R. Coifman, “Brushlets: A tool for directional image analysis and image compression,” *Journal of Applied and Computational Harmonic Analysis*, vol. 4, no. 2, pp. 147-187, 1997.
- [7] Z. K. Peng, P. W. Tse, F. L. Chu, “A comparison of improved Hilbert-Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing,” *Mechanical Systems and Signal Processing*, vol. 19, no. 5, pp. 974-988, 2005.

- [8] 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.
- [9] 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.
- [10] А. Ю. Лавриненко, Г. Ф. Коначович, Р. С. Одарченко, Д. И. Бахтияров, "Алгоритм сжатия сигналов речевых команд управления функциями беспилотного летательного аппарата," *Авиационно-космическая техника и технология*, Т. 130, № 3, С. 51-61, 2016.
- [11] C. R. Johnson *et al.*, "Image processing for artist identification," *IEEE Signal Processing Magazine*, vol. 25, no. 4, pp. 37-48, 2008.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] Liwei Liu, Geheng Chen, Feng Qian, "The application of Hilbert-Huang Transform in speech enhancement," *Proc. IEEE Int. Conf. on Computer, Mechatronics, Control and Electronic Engineering*, Changchun, Aug. 24-26, 2010, pp. 207-210.

[16] O. Yu. Lavrynenko, 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.

[17] S. Jaggi, A. S. Willsky, W. C. Karl, S. Mallat, “Multiscale geometrical feature extraction and object recognition with wavelets and morphology,” *Proc. IEEE Int. Conf. on Image Processing*, Washington, Oct. 23-26, 1995, pp. 372-375.

[18] Г. Ф. Конахович, 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, № у 2016 12187, заявл. 01.12.2016, опубл. 25.05.2017, бюл. № 10, 5 с.

РОЗДІЛ 4

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ СЕМАНТИЧНОГО КОДУВАННЯ МОВНИХ СИГНАЛІВ

4.1. Показники ефективності семантичного кодування мови

Нехай дано два вектора семантичних ознак мовного сигналу $x = (x_1 \dots x_L)$, $y = (y_1 \dots y_L)$ довжиною L відліків, тоді коефіцієнт кореляції Пірсона (КК) розраховується згідно наступного виразу

$$КК = \frac{1}{L} \sum_{i=1}^L \left(\frac{x_i - M_x}{S_x} \right) \left(\frac{y_i - M_y}{S_y} \right), \quad (4.1)$$

де $M_x = \frac{1}{L} \sum_{i=1}^L x_i$, $M_y = \frac{1}{L} \sum_{i=1}^L y_i$ – математичні очікування векторів x і y , $S_x =$

$\sqrt{\frac{1}{L} \sum_{i=1}^L (x_i - M_x)^2}$, $S_y = \sqrt{\frac{1}{L} \sum_{i=1}^L (y_i - M_y)^2}$ – середньоквадратичні відхилення

векторів x і y . За допомогою коефіцієнта кореляції Пірсона можна визначити силу лінійної залежності між двома векторами значень x і y , тобто якщо $|КК| = 1$ – існує функціональна лінійна залежність, а якщо $КК = 0$ – лінійна залежність відсутня. У випадках, коли розраховане значення коефіцієнта кореляції лежить згідно умови $0 \leq |КК| \leq 1$, то з допустимою похибкою, коефіцієнту кореляції можна дати якісну оцінку відповідно до таблиці 4.1.

Таблиця 4.1

Кількісна міра тісноти зв'язку, КК	Якісна характеристика сили зв'язку
0 – 0,1	Відсутній
0,1 – 0,3	Слабкий
0,3 – 0,5	Помірний
0,5 – 0,7	Помітний
0,7 – 0,9	Високий
0,9 – 0,99	Дуже високий
0,99 – 1	Функціональний

Середньоквадратична помилка (СКП) обчислюється за формулою

$$\text{СКП} = \sqrt{\frac{1}{L} \sum_{i=1}^L (x_i - y_i)^2}. \quad (4.2)$$

В даному випадку нас цікавить найменша помилка з найбільшою геометричною близькістю між порівнюваними семантичними ознаками мовного сигналу, тобто $\text{СКП} \rightarrow 0$.

Відношення сигнал/шум (ВСШ) і пікове відношення сигнал/шум (ПВСШ) розраховуються згідно наступних формул

$$\text{ВСШ} = 10 \log_{10} \left(\frac{\frac{1}{L} \sum_{i=1}^L x_i^2}{\frac{1}{L} \sum_{i=1}^L (x_i - y_i)^2} \right) [\text{дБ}]. \quad (4.3)$$

$$\text{ПВСШ} = 10 \log_{10} \left(\frac{\max(x_i^2)}{\frac{1}{L} \sum_{i=1}^L (x_i - y_i)^2} \right) [\text{дБ}]. \quad (4.4)$$

Очевидно, що чим більше геометрична близькість між порівнюваними семантичними ознаками мовного сигналу, тим більше ВСШ і ПВСШ відповідно, тобто $\text{ВСШ}(\text{ПВСШ}) \rightarrow \infty$, в іншому випадку $\text{ВСШ}(\text{ПВСШ}) \rightarrow 0$.

Коефіцієнт стиснення (КС) мовних даних характеризує наскільки ефективно функціонує алгоритм семантичного кодування мови і розраховується відповідно до виразу

$$\text{КС} = \frac{S_o}{S_c}, \quad (4.5)$$

де S_o – об'єм вихідних мовних даних, S_c – об'єм семантично закодованих мовних даних. Таким чином, чим вище коефіцієнт стиснення, тим алгоритм ефективніше. Слід зазначити, що якщо $\text{КС} = 1$, то алгоритм не здійснює стиснення, тобто вихідне повідомлення виявляється за об'ємом рівним вхідному.

Швидкість передачі бітів (ШПБ)

$$\text{ШПБ} = v \times \log_2 L_v \left[\frac{\text{біт}}{\text{с}} \right], \quad (4.6)$$

де v [відліків/с] – швидкість передачі v відліків мовного сигналу за 1 секунду;
 L_v – загальне число рівнів квантування відліків мовного сигналу.

4.2. Порівняльний аналіз ПФ, КП та ВП як спектрального аналізу цифрових мовних сигналів

У роботі проводилося порівняння перетворення Фур'є (ПФ), косинусного перетворення (КП) та вейвлет-перетворення (ВП) в якості провідних технологій спектрального аналізу, які використовуються в алгоритмах стиснення мовних сигналів. В якості вхідного цифрового мовного сигналу для стиснення використовується запис чоловічого голосу с частотою дискретизації 8 кГц і розрядністю квантування 8 біт, що відповідає основному цифровому каналу телефонної мережі – 64 кбіт/с.

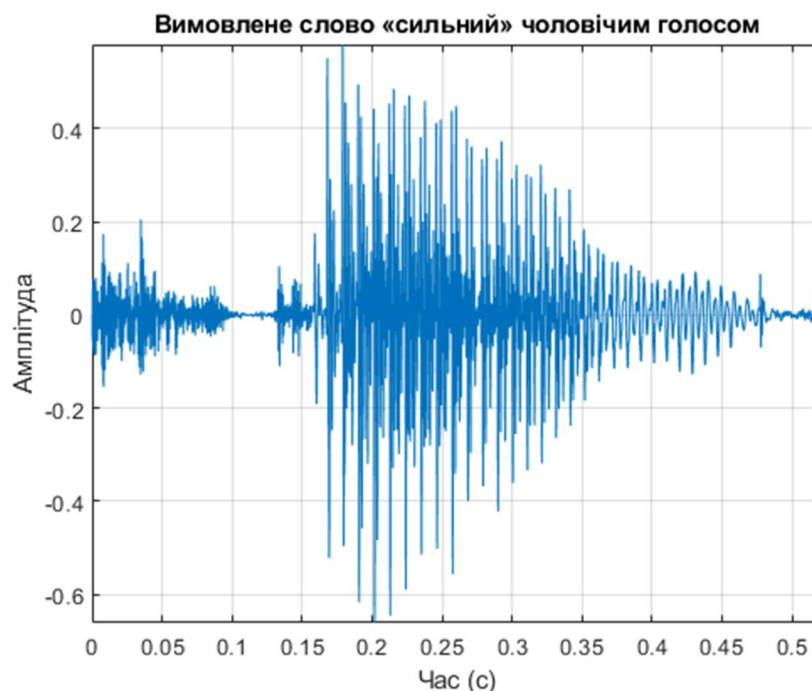


Рис. 4.1. Цифровий мовний сигнал

Досягнення в області цифрової мікросхемотехніки є фундаментом практичного впровадження методів цифрової передачі мовних сигналів. На основі таких методів створені системи високоякісної передачі, які здатні

функціонувати в реальних умовах і забезпечують розпізнавання мови на тлі шумів і завад. З точки зору теорії інформації подібні системи здійснюють кодування мови. Мета такого перетворення полягає в зменшенні числа бітів, що відображають задану кількість інформації за певний час, при цьому надмірність цифрового мовного сигналу зменшується. Основне завдання стиснення мовних сигналів є зменшення потоку переданих даних по цифровому каналу зв'язку при незначному погіршенні якості відновленої мови на приймальній стороні. Оскільки мовний сигнал являє собою нестационарний випадковий процес, то для його обробки було запропоновано використовувати дискретне вейвлет-перетворення. Основна властивість ВП полягає в тому, що перетворений сигнал представляється великою кількістю надлишкових вейвлет-коефіцієнтів які після порогової обробки будуть обнулені. Обнулення коефіцієнтів відбувається за допомогою заданої порогової функції, коефіцієнти нижче або дорівнюють якій будуть прирівняні до нуля, а всі інші залишаться без змін, тому вейвлет-коефіцієнти з абсолютним значенням близьким до нуля містять лише невелику частину енергії сигналу. Обнулення цих коефіцієнтів призводить до незначних втрат енергії. Ця властивість робить ВП привабливим для стиснення мовних сигналів. Слід зауважити, що стиснення сигналів з використанням ВП буде, відноситься до стиснення даних з втратами, так як обнулення надлишкові вейвлет-коефіцієнти будуть впливати на якість мовного сигналу в пропорційному співвідношенні [5].

Порогова функція $T_{SS}(\tilde{d}, \lambda_1, \lambda_2)$ виду:

$$T_{SS}(\tilde{d}, \lambda_1, \lambda_2) = \begin{cases} 0, & |\tilde{d}| \leq \lambda_1 \\ \text{sign}(\tilde{d}) \times \left[\frac{\lambda_2(|\tilde{d}| - \lambda_1)}{\lambda_2 - \lambda_1} \right], & \lambda_1 < |\tilde{d}| \leq \lambda_2 \\ \tilde{d}, & |\tilde{d}| > \lambda_2 \end{cases} \quad (4.7)$$

яка включає вже дві порогові величини λ_1, λ_2 . Графік цієї функції (при $\lambda_1 = 1, \lambda_2 = 2$) наведено на Рис. 4.2 (крива 3).

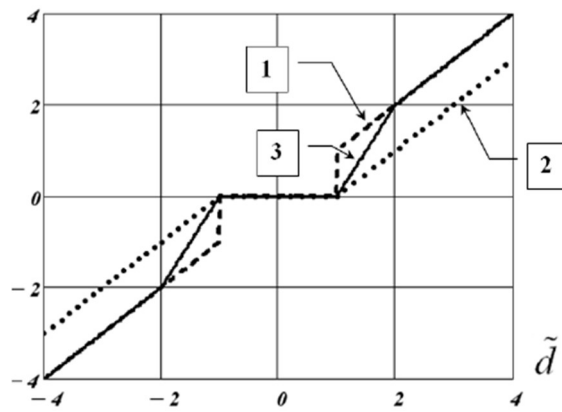


Рис. 4.2. Порогова функція (4.7)

На Рис. 4.2 зображена порогова функція для обробки мовного сигналу. Підвищення порогу буде підвищувати ступінь зменшення надмірності, але разом з тим, зростатимуть втрати якості розбірливості мови. Зниження порогу дозволяє зменшити втрати інформаційних вейвлет-коефіцієнтів, але також знижується ефективність стиснення сигналу.

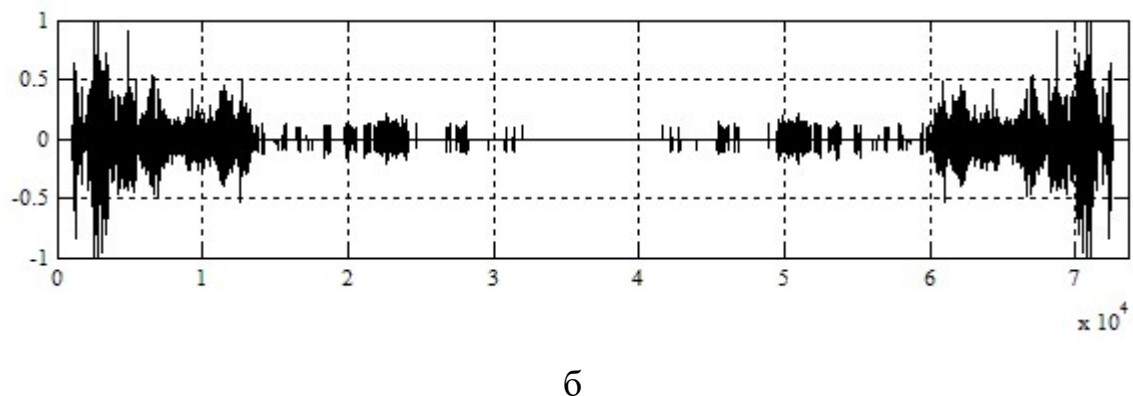
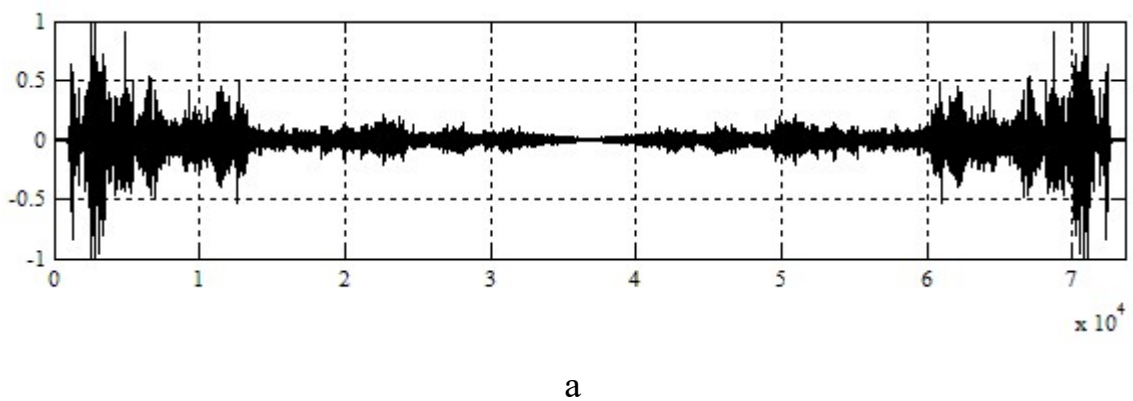
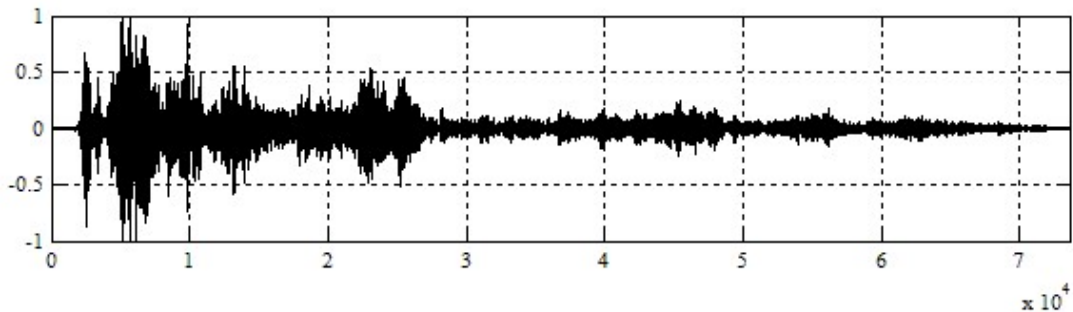


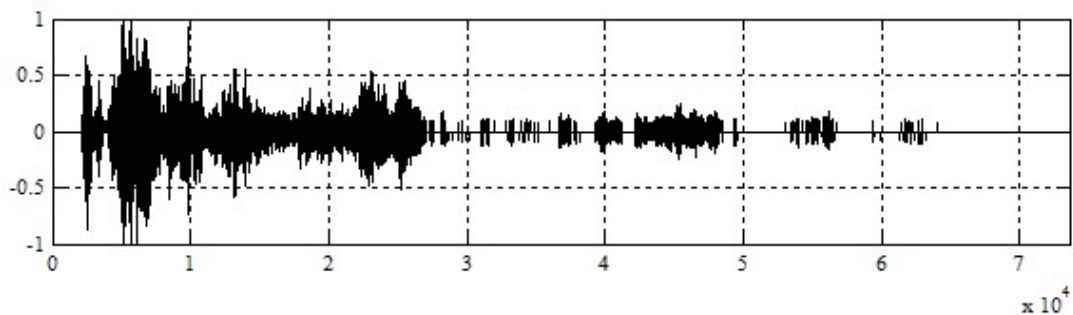
Рис. 4.3. Спектральна складова ПФ

а – коефіцієнти ПФ до порогової обробки, б – коефіцієнти ПФ після порогової обробки

На Рис. 4.3 зображені графіки коефіцієнтів ПФ досліджуваного мовного сигналу до порогової обробки та після порогової обробки відповідно. Критерієм для визначення порогу була залежність КС від КК (Табл. 4.1). На Рис. 4.3б видно як надмірні коефіцієнти ПФ обнуляються.



а



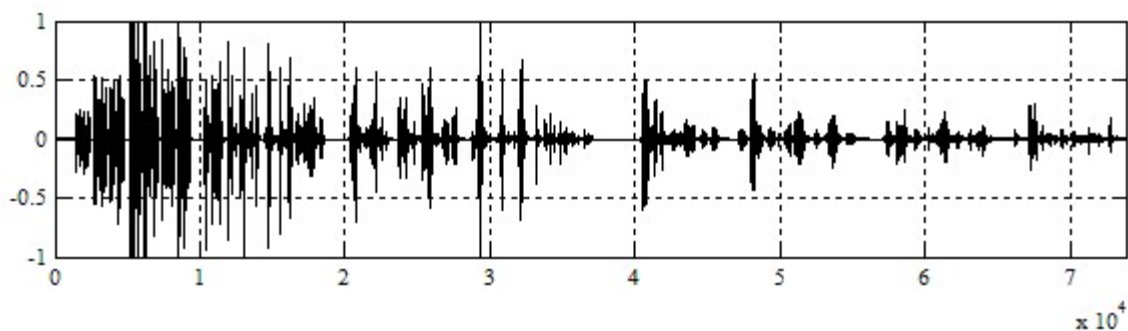
б

Рис. 4.4. Спектральна складова КП

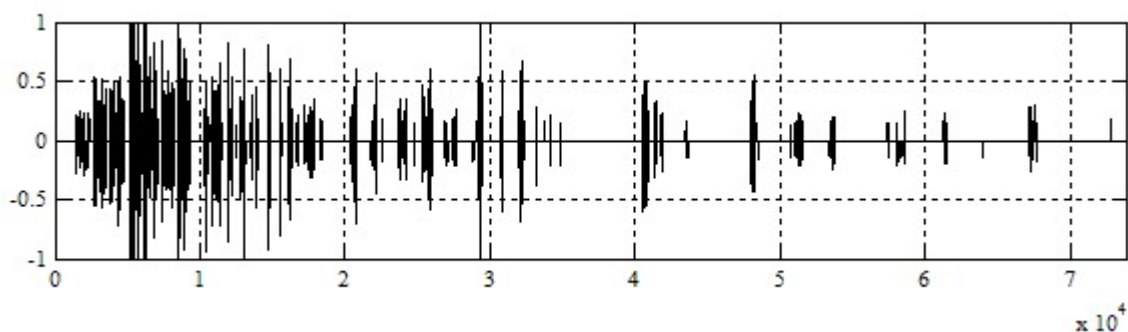
а – коефіцієнти КП до порогової обробки, б – коефіцієнти КП після порогової обробки

На Рис. 4.4 зображено графіки коефіцієнтів КП досліджуваного мовного сигналу до порогової обробки та після порогової обробки відповідно. Критерієм для визначення порогу була залежність КС від КК. (Табл. 4.1). На Рис. 4.4б видно як надмірні коефіцієнти КП обнуляються.

На Рис. 4.5 зображено графіки коефіцієнтів ВП мовного сигналу на базі вейвлет-функції Добеші 12-го порядку до порогової обробки та після порогової обробки відповідно. Головним критерієм для визначення порогу була залежність КС від КК (Табл. 4.1). На Рис. 4.5б видно як надмірні коефіцієнти ВП обнуляються в залежності від значення порогової функції.



а



б

Рис. 4.5. Спектральна складова ВП

а – коефіцієнти ВП до порогової обробки, б – коефіцієнти ВП після порогової обробки

В якості перетворення мовного сигналу було запропоновано використовувати ВП. Дане перетворення має переваги в порівнянні з ПФ, та КП так, як воно володіє адаптивністю для отримання необхідного ступеня стиснення, зберігаючи при цьому необхідну якість сигналу. В результаті першого кроку ВП часове розрішення зменшується в два рази, так як лише половина відліків характеризує весь сигнал. Однак частотне розрішення подвоюється, так як сигнал займає тепер половинну смугу частот і невизначеність зменшується. Найбільш значимі частоти вхідного сигналу будуть відображатися, як великі амплітуди вейвлет-коефіцієнтів, «відповідають» за відповідний частотний діапазон. Малі значення вейвлет-коефіцієнтів означають низьку енергетику відповідних частотних смуг в сигналі. Ці коефіцієнти можуть бути прирівняні нулю без істотного спотворення сигналу.

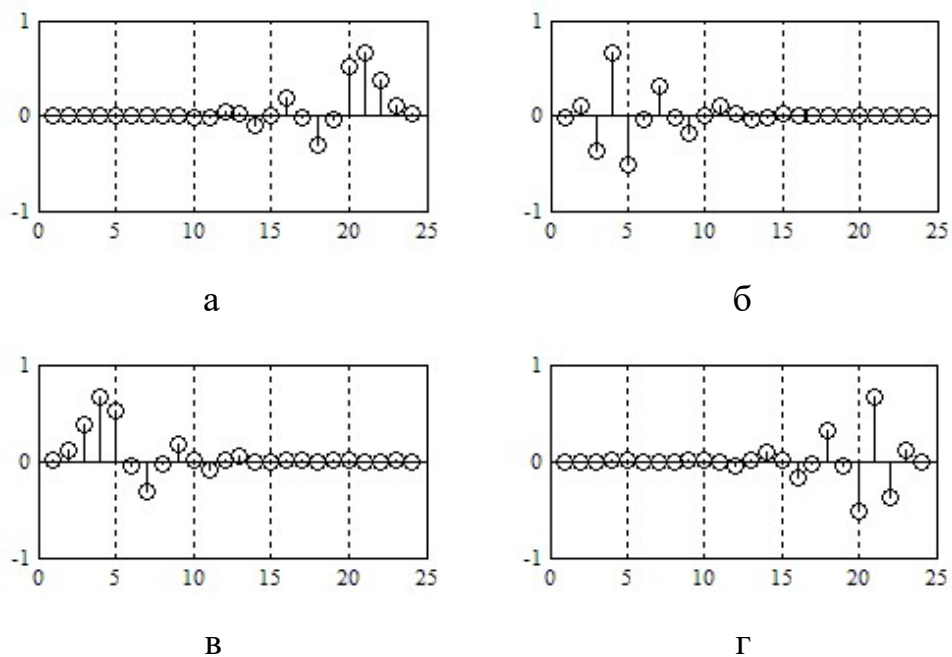


Рис. 4.6. Ортогональні фільтри Добеші 12-го порядку
 а – фільтр декомпозиції нижніх частот, б – фільтр декомпозиції високих частот, в – фільтр реконструкції нижніх частот, г – фільтр реконструкції високих частот

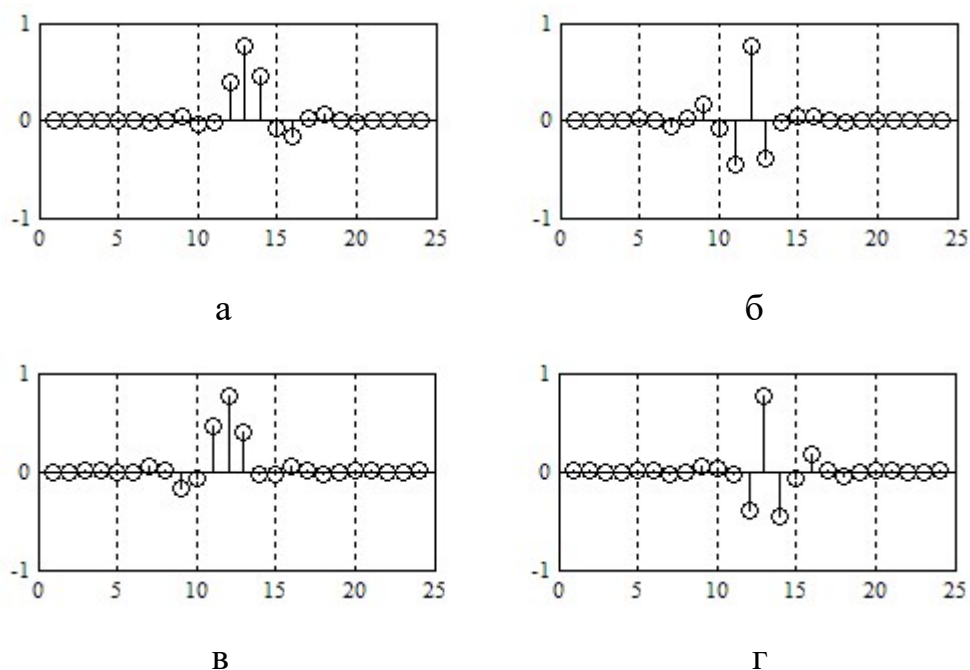


Рис. 4.7. Ортогональні фільтри «Сімлет» 12-го порядку
 а – фільтр декомпозиції нижніх частот, б – фільтр декомпозиції високих частот, в – фільтр реконструкції нижніх частот, г – фільтр реконструкції високих частот

У завданнях обробки мови і звуку в силу особливостей слухового сприйняття людини, вимоги, що накладаються на форму спектрів вейвлет-функції, досить високі, що обумовлює використання великого числа нульових моментів (10 - 15 нульових моментів). Вейвлети Добеші довжиною L мають $M = L/2$ нульових моментів. Однак слід пам'ятати, що число нульових моментів визначає тривалість вейвлет-функцій і значить, швидкість дії алгоритму обчислення вейвлет-перетворення. У класичній конструкції Добеші довжина фільтрів $L = 2M$, де M – кількість нульових моментів. Вейвлети сімейства Добеші позначаються як dbp , де $p = M$ – порядок. Всі вейвлет-функції Добеші мають компактний носій. Прикладами вейвлетів Добеші з одним і двома нульовими моментами є функції Добеші 1-го порядку $db1$ ($M = 1$) та 2-го порядку $db2$ ($M = 2$). Серед вейвлетів з компактным носієм властивістю симетричності володіють лише функція Хаара (функція Добеші – $db1$). Неважко помітити, що гладкість вейвлетів зростає в міру збільшення їх порядку. Одночасно зростає і частота осциляцій. Ці вейвлети мають характерну асиметрію – наростання функції розтягнуто в порівнянні зі спадом.

Основний при роботі з вейвлет-перетворенням є проблема вибору найбільш відповідного вейвлета. Вибір конкретного сімейства вейвлетів диктується прикладними завданнями та типом інформації про сигнал, який потрібно максимально розпізнати. Не існує якихось жорстких правил, але краще всього вибирати вейвлет таким чином, щоб він належав такому ж класу функцій, що й аналізований сигнал. Число нульових моментів більш важливо для досягнення більш високого коефіцієнта стиснення сигналу, який при великому числі нульових моментів збільшується. Гладкість вейвлета стає важливіше при зворотному перетворенні, коли необхідно згладити помилки, викликані стисненням (відкиданням малих вейвлет-коефіцієнтів). Для вейвлетів dbM число коефіцієнтів всіх фільтрів – $2M$. З підвищенням порядку (числа нульових моментів) підвищується гладкість функції. Таким чином, підбором порядку материнського вейвлета можна домогтися найкращого

наближення до заданого сигналу. Однак при збільшенні числа нульових моментів M зростає довжина фільтра $L = 2M$, що приводить до зростання обчислювальної складності, яка дорівнює LN , де N – довжина блоку даних.

Перший експеримент (Табл. 4.2) було проведено з урахуванням сталого критерію якості, в ролі якого виступав КК оригінального та відновленого сигналів. При сталому КК – 0.95, ПФ показує найгірші результати, де КС – 4.4305, ВСШ – 10.1173 дБ, ПВСШ – 29.9180 дБ, СКП – 0.3120. КП з тим же КК – 0.95 показує вже кращі результати, якщо ВСШ – 10.1115 дБ, ПВСШ – 29.9123 дБ та СКП – 0.3122 приблизно однакові з ПФ, то КС – 5.6917, що робить КП більш ефективним в стисненні сигналів.

Таблиця 4.2

Відношення залежності КС від КК, ВСШ, ПВСШ, СКП

Досліджувана технологія	КК	ВСШ (дБ)	ПВСШ (дБ)	СКП	КС
FFT	0.95	10.1173	29.9180	0.3120	4.4305
DCT	0.95	10.1115	29.9123	0.3122	5.6917
DWT «Haar»	0.95	10.1888	30.1481	0.3094	6.3883
DWT «Coif1»	0.95	10.1240	30.0434	0.3117	9.3149
DWT «Db4»	0.95	10.1262	29.9270	0.3117	10.7401
DWT «Sym4»	0.95	10.1207	29.9215	0.3119	10.8083
DWT «Coif3»	0.95	10.1193	30.2496	0.3119	11.7169
DWT «Db8»	0.95	10.1366	29.9374	0.3113	11.8531
DWT «Sym8»	0.95	10.1289	30.0188	0.3116	11.8708
DWT «Sym12»	0.95	10.1164	29.9171	0.3120	12.2648
DWT «Db12»	0.95	10.1079	29.9087	0.3123	12.2813
DWT «Coif5»	0.95	10.1268	30.0265	0.3116	12.4350
DWT «Dmey»	0.95	10.1259	29.9347	0.3117	13.2129

Примітка: FFT – перетворення Фур'є; DCT – косинусне перетворення; DWT – вейвлет-перетворення на базі вейвлет-функцій: Haar (Хаара); Db4, 8, 12 (Добеші 4, 8, 12 -го порядку); Sym4, 8, 12 («Сімлет» 4, 8, 12 -го порядку); Coif1, 3, 5 («Коіфлет» 1, 3, 5 -го порядку); Dmey (дискретна функція Мейера).

При порівняльному аналізі ВП треба враховувати той факт, що ефективність ВП напряму залежить від вибраної вейвлет-функції. ВП на базі вейвлет-функції Haar (Хаара) показує найгірші результати: КС – 6.3883, ВСШ – 10.1888 дБ, ПВСШ – 30.1481 дБ та СКП – 0.3094, з усіх досліджуваних

вейвлет-функцій, але в той же час ВП на базі вейвлет-функції Haar показує кращі результати від ПФ та КП. Дуже важливим є той факт, що ПФ є найменш затратним в обчислювальній потужності, але головна наша задача полягає в знаходженні оптимального варіанту найкращого КС при сталій якості сигналу.

ВП на базі вейвлет-функції Sym12 («Сімлет» 12-го порядку) та Db12 (Добеші 12-го порядку) мають приблизно однакові результати: КС – 12.2648 (12.2813), ВСШ – 10.1164 дБ (10.1079 дБ), ПВСШ – 29.9171 дБ (29.9087 дБ), СКП – 0.3120 (0.3123). З результатів явно видно що ВП на базі вейвлет-функцій Sym12 та Db12 в 2.8 рази більше стискає досліджувальний сигнал аніж ПФ при тій же самій якості.

Найкращі результати показали вейвлет-функції Coif5 («Коіфлет» 5-го порядку) та Dmey (дискретна функція Мейера), де КС – 12.4350 (13.2129), ВСШ – 10.1268 дБ (10.1259 дБ), ПВСШ – 30.0265 дБ (29.9347 дБ), СКП – 0.3116 (0.3117), але проблема полягає в тому, що ці вейвлет-функції з достатньою обчислювальною витратністю. Результати дослідження показали, що довжина вейвлет-фільтру Dmey – 102 коефіцієнти, а довжина фільтру Coif5 – 30 коефіцієнтів, а найближчі до них фільтри Sym12 та Db12 мають довжину – 24 коефіцієнти, що робить їх найбільш привабливими у відношенні КС від витраченої розрахункової потужності на кожен вейвлет-фільтр. Як було сказано вище число нульових моментів більш важливо для досягнення більш високого коефіцієнта стиснення сигналу, який при великому числі нульових моментів збільшується. А за допомогою порядку материнського вейвлета, можна домогтися найкращого наближення до заданого сигналу. Однак при збільшенні числа нульових моментів зростає довжина фільтра, що приводить до зростання обчислювальної складності.

Друге експериментальне дослідження (Табл. 4.3) розглядається з урахуванням сталого критерію КС – 10. Результати стиснення на основі ПФ при сталому КС: КК – 0.8709, ВСШ – 6.1692 дБ, ПВСШ – 25.9700 дБ, СКП – 0.4915, роблять ПФ найменш якісним в розбірливості сигналу. КП показує

кращі результати: КК – 0.9023, ВСШ – 7.3084 дБ, ПВСШ – 27.1092 дБ, СКП – 0.4311.

Таблиця 4.3

Відношення залежності КК, ВСШ, ПВСШ, СКП від КС

Досліджувана технологія	КС	КК	ВСШ (дБ)	ПВСШ (дБ)	СКП
FFT	10	0.8709	6.1692	25.9700	0.4915
DCT	10	0.9023	7.3084	27.1092	0.4311
DWT «Haar»	10	0.9077	7.5429	27.4352	0.4196
DWT «Coif1»	10	0.9446	9.6765	29.5929	0.3282
DWT «Db4»	10	0.9553	10.5818	30.3826	0.2957
DWT «Sym4»	10	0.9557	10.6277	30.4284	0.2942
DWT «Coif3»	10	0.9609	11.1522	31.2127	0.2769
DWT «Db8»	10	0.9615	11.2235	31.0243	0.2747
DWT «Sym8»	10	0.9618	11.2584	31.1149	0.2736
DWT «Db12»	10	0.9639	11.4894	31.2901	0.2664
DWT «Sym12»	10	0.9639	11.4959	31.2966	0.2662
DWT «Coif5»	10	0.9646	11.5799	31.4368	0.2636
DWT «Dmey»	10	0.9679	11.9991	31.8351	0.2512

Примітка: FFT – перетворення Фур’є; DCT – косинусне перетворення; DWT – вейвлет-перетворення на базі вейвлет-функцій: Haar (Хаара); Db4, 8, 12 (Добеші 4, 8, 12 -го порядку); Sym4, 8, 12 («Сімлет» 4, 8, 12 -го порядку); Coif1, 3, 5 («Коіфлет» 1, 3, 5 -го порядку); Dmey (дискретна функція Мейера).

Беззаперечним лідером враховуючи основне зауваження (відношення якості стиснення від простоти розрахунків вейвлет-фільтрів) являються вейвлет-фільтри Sym12 та Db12 з результатами: КК – 0.9639 (0.9639), ВСШ – 11.4959 дБ (11.4894 дБ), ПВСШ – 31.2966 дБ (31.2901 дБ), СКП – 0.2662 (0.2664).

4.3. Результати експериментального дослідження методу семантичного кодування мовних сигналів на основі ЕВП

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.

Результати науково-експериментальних досліджень підвищення ефективності семантичного кодування мовних сигналів викладені нижче (Табл. 4.4). 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.

Таблиця 4.4

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

На Рис. 4.8 зображені семантичні ознаки голосових команд на основі ЕВП та спектрального аналізу Гілберта.

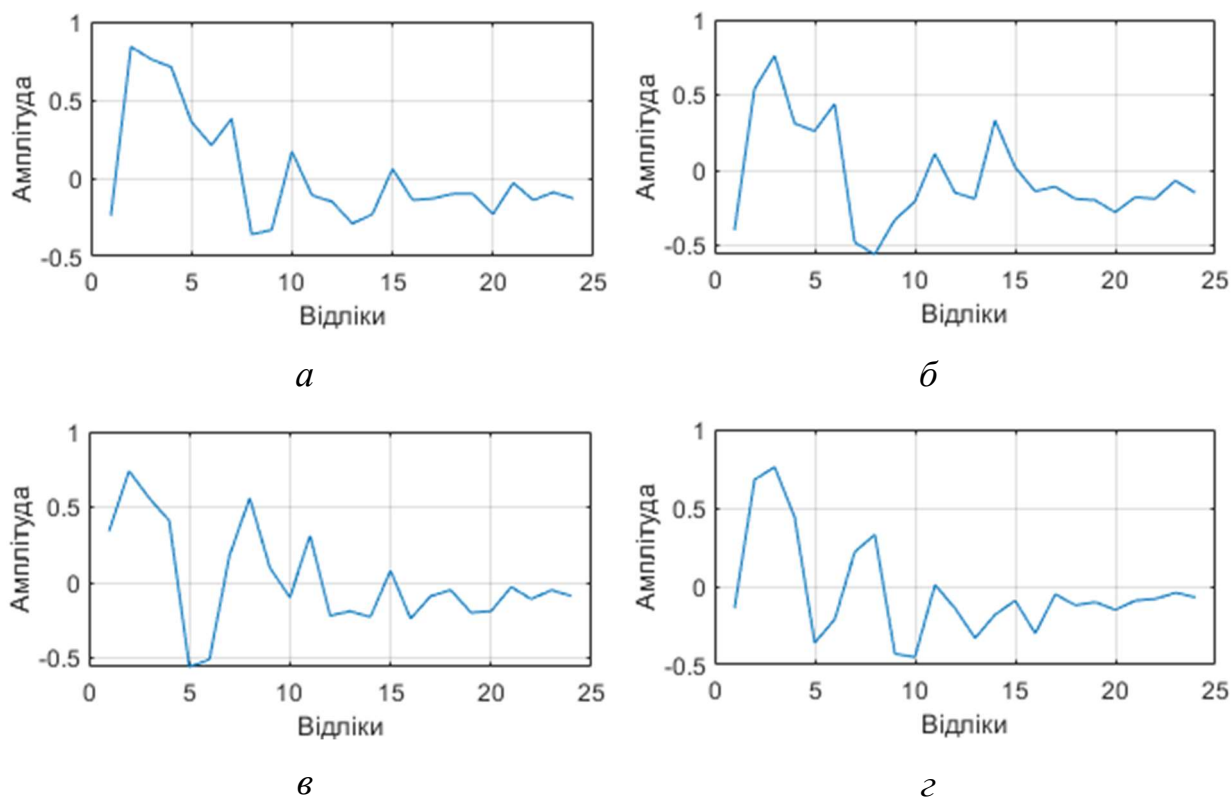


Рис. 4.8. Семантичні ознаки на основі ЕВП голосових команд: «Вгору» (а), «Вниз» (б), «Вправо» (в), «Вліво» (г)

З наведених вище результатів дослідження (Табл. 4.4) 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, а точніше експериментально підтверджує твердження, що мовні сигнали дійсно мають глибинні семантичні закономірності, визначенню та оцінці яких присвячене дане наукове дослідження.

Результати наступного науково-експериментального дослідження оцінки підвищення ефективності семантичного кодування мовних сигналів за допомогою розробленого методу викладені нижче (Табл. 4.5). 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.

Таблиця 4.5

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

Навчання	Тестування			
	«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

З наведених вище результатів дослідження (Табл. 4.5) 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. 4.9, де зображено сто реалізацій семантичних ознак знайдених на основі ЕВП та спектрального аналізу Гілберта голосових команд: «Вгору», «Вниз», «Вправо», «Вліво».

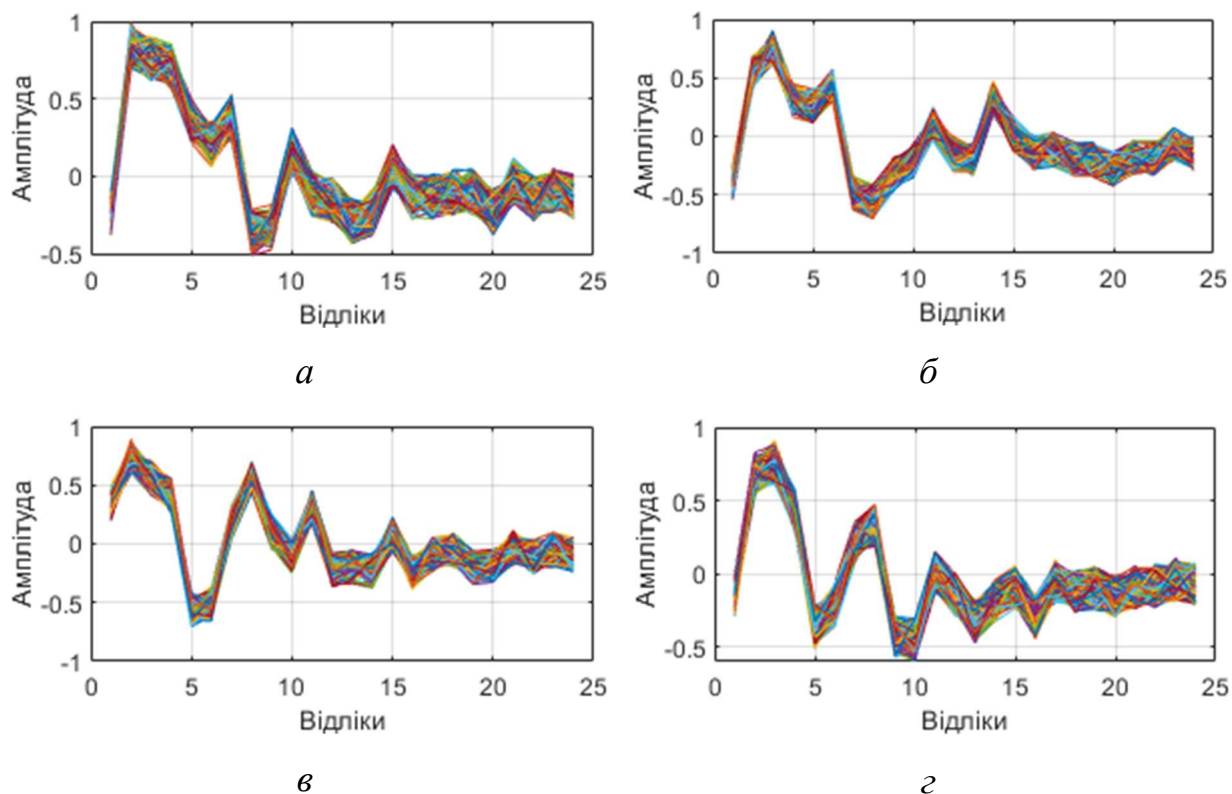


Рис. 4.9. Сто реалізацій семантичних ознак на основі ЕВП голосових команд: «Вгору» (а), «Вниз» (б), «Вправо» (в), «Вліво» (г)

Експериментальне дослідження показало (див. Рис. 4.10), 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.

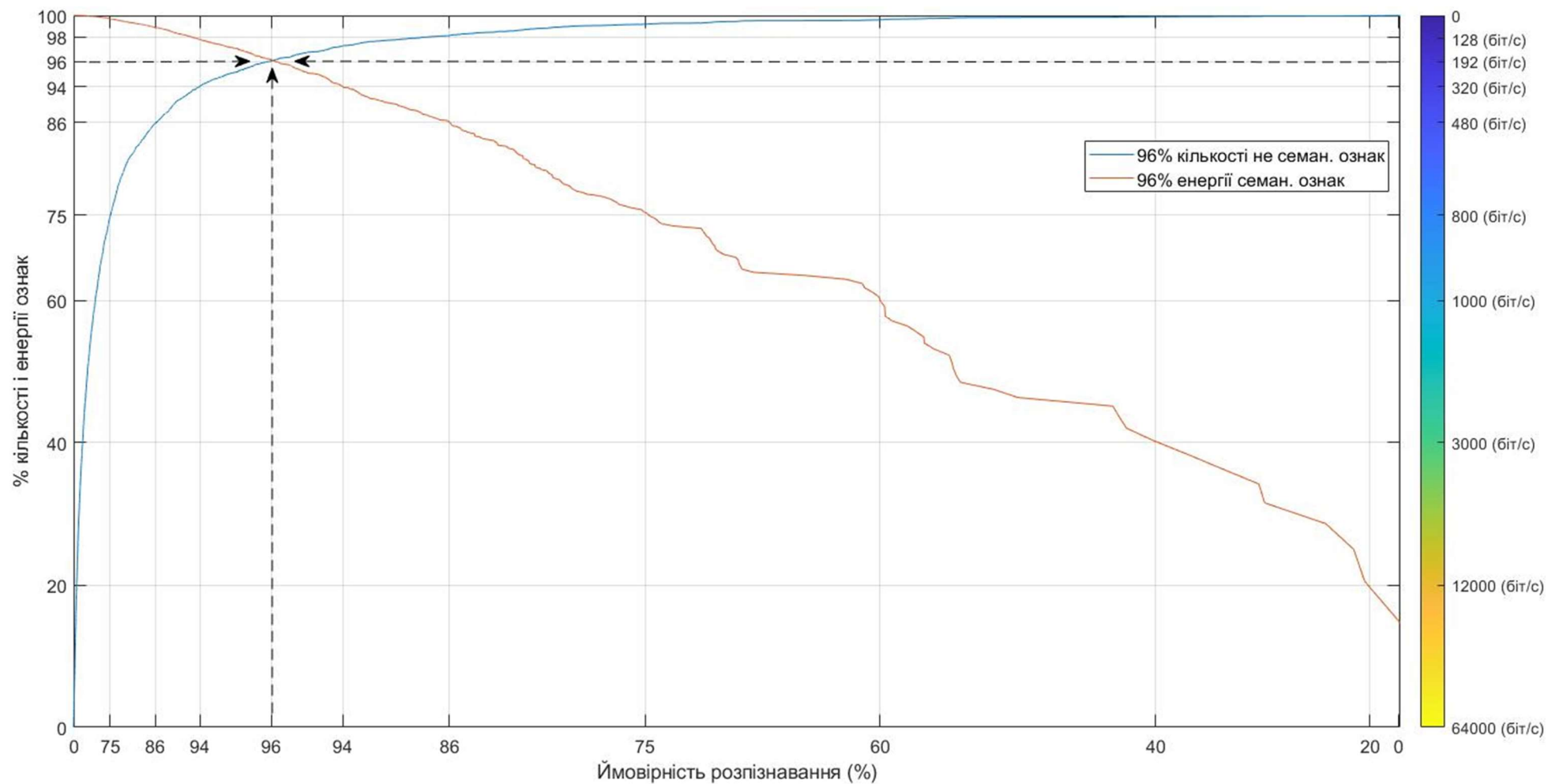


Рис. 4.10. Залежність ймовірності розпізнавання семантичних ознак від скорочення їх енергії і швидкості кодування за допомогою запропонованого методу

Висновки до четвертого розділу

1. Проведено порівняльний аналіз ПФ, КП та ВП в якості спектрального аналізу цифрових мовних сигналів в алгоритмах стиснення, де з результатів явно видно що ВП на базі вейвлет-функцій Sym12 та Db12 в 2.8 рази більше стискає досліджувальний сигнал аніж ПФ, та в 2.2 рази більше аніж КП при тій же самій якості КК – 0.95.

2. Показано, що при збільшенні числа нульових моментів зростає довжина фільтра (фільтри: Dmey – 102 коефіцієнти, Coif5 – 30 коефіцієнтів, Sym12 та Db12 – 24 коефіцієнти), що приводить до зростання обчислювальної складності.

3. Доведено, що оптимальним методом спектрального аналізу для цифрових мовних сигналів в задачах їх стиснення є ВП на базі вейвлет-функцій Sym12 та Db12 з результатами: КК – 0.9639 (0.9639), ВСШ – 11.4959 дБ (11.4894 дБ), ПВСШ – 31.2966 дБ (31.2901 дБ), СКП – 0.2662 (0.2664).

4. Отримані кількісні показники ефективності запропонованого методу семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення, в саме, знижується 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.

Список використаних джерел в четвертому розділі

[1] А. Ю. Лавриненко, Г. Ф. Конахович, Р. С. Одарченко, Д. И. Бахтияров, “Алгоритм сжатия сигналов речевых команд управления функциями беспилотного летательного аппарата,” *Авиационно-космическая техника и технология*, Т. 130, № 3, С. 51-61, 2016.

- [2] 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.
- [3] L. R. Rabiner, R. W. Schafer, "Matlab exercises in support of teaching digital speech processing," *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, May 4-9, 2014, pp. 2480-2483.
- [4] 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.
- [5] Г. Ф. Конахович, 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.
- [6] L. Rabiner, R. Schafer, *Theory and Applications of Digital Speech Processing*. Upper Saddle River, NJ: Prentice Hall, 2010.
- [7] 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.
- [8] S. M. Hughes, I. Daubechies, "Simpler alternatives to information theoretic similarity metrics for multimodal image alignment," *Proc. IEEE Int. Conf. on Image Processing*, Atlanta, Oct. 8-11, 2006, pp. 365-368.
- [9] А. Ю. Лавриненко, Ю. А. Кочергин, Г. Ф. Конахович, "Система захищеного голосового управління беспилотным летательным аппаратом," *Авиационно-космическая техника и технология*, Т. 148, № 4, С. 4-11, 2018.
- [10] P. Hill, *Audio and Speech Processing with MATLAB*. Boca Raton, FL: CRC Press, 2018.

[11] 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.

[12] A. A. de Lima *et al.*, “On the quality assessment of sound signals,” *Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS)*, Seattle, May 18-21, 2008, pp. 416-419.

[13] 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.

[14] A. B. Downey, *Think DSP: Digital Signal Processing in Python*. Needham, MA: Green Tea Press, 2012.

[15] 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.

[16] L. R. Rabiner, R. W. Schafer, “Programs for supporting the teaching of digital speech processing,” *Proc. IEEE Workshop on Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, Napa, Aug. 11-14, 2013, pp. 290-295.

[17] 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.

[18] Е. С. Вентцель, *Теория вероятностей*. Москва: Наука, 1969.

[19] А. Ю. Лавриненко, Ю. А. Кочергин, Г. Ф. Конахович, “Система распознавания стеганографически-преобразованных голосовых команд управления БПЛА,” *Радиоэлектронные и компьютерные системы*, Т. 87, № 3, С. 20-28, 2018.

[20] J. Wung, T. S. Wada, B. Juang, B. Lee, M. Trott, R. W. Schafer, “A system approach to acoustic echo cancellation in robust hands-free teleconferencing,” *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, Oct. 16-19, 2011, pp. 101-104.

[21] Г. Ф. Конахович, 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, № у 2016 12187, заявл. 01.12.2016, опубл. 25.05.2017, бюл. № 10, 5 с.

[22] О. Yu. Lavrynenko, G. F. Konakhovych, D. I. Bakhtiarov, “Compression algorithm of voice control commands of UAV based on wavelet transform,” *Electronics and Control Systems*, vol. 55, no. 1, pp. 17-22, 2018.

[23] Б. Мефферт, О. Хохмут, *Инструменты обработки сигналов – основы, примеры применения и задачи*. Берлин: Изд-во HUB, 2019.

[24] О. Lavrynenko, G. Konakhovych, D. Bakhtiarov, “Comparative analysis of speech recognition algorithms in UAV voice control system,” *Science-Based Technologies*, vol. 38, no. 2, pp. 137-142, 2018.

[25] Е. С. Вентцель, Л. А. Овчаров, *Прикладные задачи теории вероятностей*. Москва: Радио и связь, 1983.

[26] О. Yu. Lavrynenko, 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.

[27] L. O. Nunes *et al.*, “A parametric objective quality assessment tool for speech signals degraded by acoustic echo,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 8, pp. 2181-2190, 2012.

[28] Г. Ф. Конахович, Д. І. Бахтіяров, О. Ю. Лавриненко, “Комп’ютерне моделювання захищеного каналу керування безпілотним літальним апаратом,” *Наукоємні технології*, Т. 28, № 4, С. 283-290, 2015.

[29] О. Ю. Лавриненко, Г. Ф. Конахович, 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, № и 2017 03401, заявл. 10.04.2017, опубл. 25.09.2017, бюл. № 18, 6 с.

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.

ДОДАТОК А. МОДЕЛЮВАННЯ КІЛЬКІСНОЇ ОЦІНКИ СЕМАНТИЧНОЇ МОВНОЇ ІНФОРМАЦІЇ

Експериментальне підтвердження висунутих гіпотез проводиться за допомогою комп'ютерного моделювання в програмному середовищі MATLAB R2020b.

Завантажимо MAT-файл, який містить пару модельних мовних сигналів, що відповідають вимовленому слову «сильний», жінкою і чоловіком, відповідно. Сигнали дискретизовані з частотою 8 кГц.

```
S = load('strong.mat')
```

S = struct with fields:

Fs: 8000

her: [4899×1 double]

him: [4173×1 double]

Тоді задамо частоту дискретизації

```
Fs = S.Fs
```

Fs = 8000

тривалість (с) одного відліку досліджуваних сигналів

```
dt = 1/Fs
```

dt = 1.2500e-04

мовні сигнали вимовлені жіночим голосом

```
Sher = (S.her)'
```

Sher = 1×4899

-0.0011 -0.0059 0.0016 -0.0037 -0.0120 -0.0124 -0.0066 ...

та чоловічим голосом

```
Shim = (S.him)'
```

Shim = 1×4173

0.0038 -0.0571 0.0551 -0.0061 -0.0680 0.0614 -0.0661 ...

Щоб прослухати досліджувані мовні сигнали потрібно ввести наступну команду

```
soundsc(Sher,Fs), pause(1), soundsc(Shim,Fs)
```

Знайдемо тривалість за часом (с), протягом якого досліджувані сигнали існують.

```
t_Sher = dt*length(Sher)
```

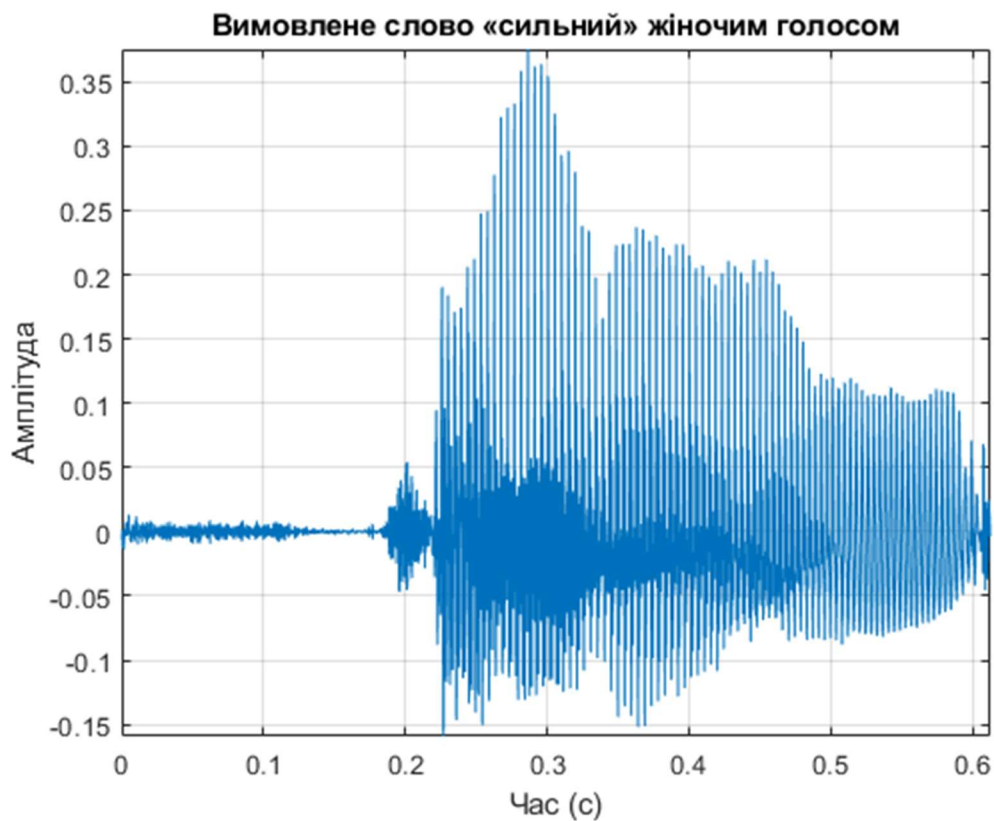
t_Sher = 0.6124

```
t_Shim = dt*length(Shim)
```

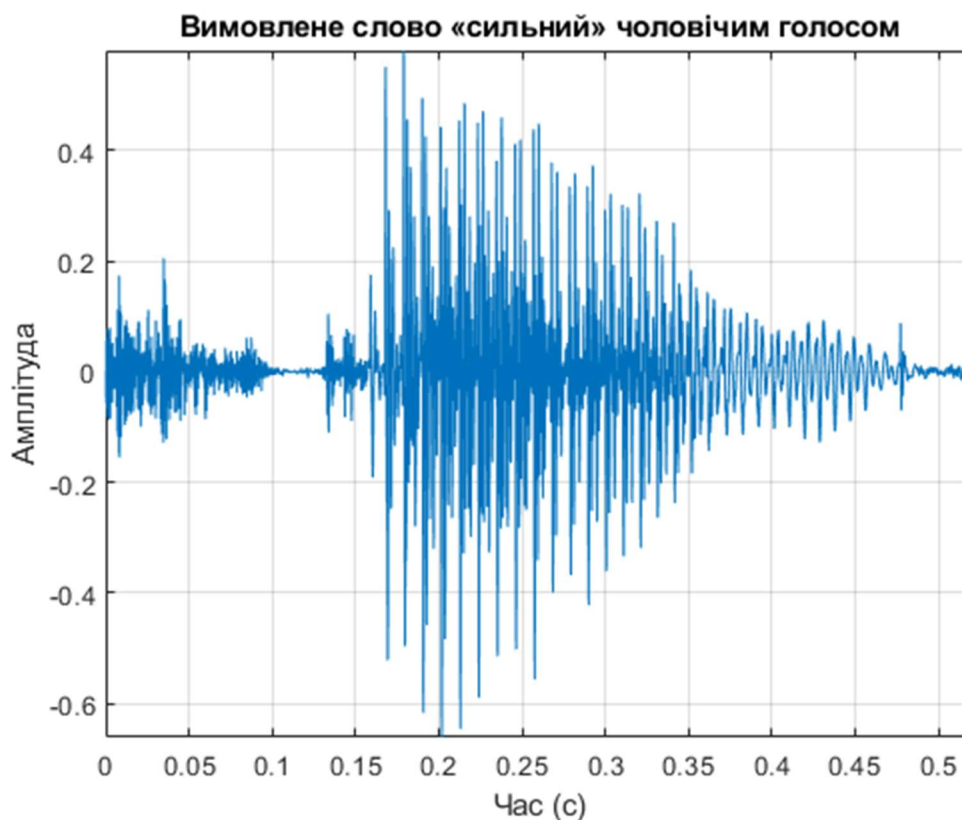
t_Shim = 0.5216

Побудуємо графіки досліджуваних мовних сигналів.

```
figure(1)
plot(dt:dt:t_Sher,Sher); axis([0 t_Sher min(Sher) max(Sher)]); grid on;
title('Вимовлене слово «сильний» жіночим голосом');
xlabel('Час (с)'); ylabel('Амплітуда');
```



```
figure(2)
plot(dt:dt:t_Shim,Shim); axis([0 t_Shim min(Shim) max(Shim)]); grid on;
title('Вимовлене слово «сильний» чоловічим голосом');
xlabel('Час (с)'); ylabel('Амплітуда');
```



Задамо кількість біт на відлік (біт/відлік).

$$b = 8$$

$$b = 8$$

Знайдемо кількість рівнів квантування.

$$L = 2^b$$

$$L = 256$$

Тоді максимальне значення рівнів квантування.

$$L_{\max} = L - 1$$

$$L_{\max} = 255$$

Знайдемо крок квантування.

$$q_Sher_n = (\max(Sher_n) - \min(Sher_n)) / L_{\max}$$

$$q_Sher_n = 0.0021$$

$$q_Shim_n = (\max(Shim_n) - \min(Shim_n)) / L_{\max}$$

$$q_Shim_n = 0.0049$$

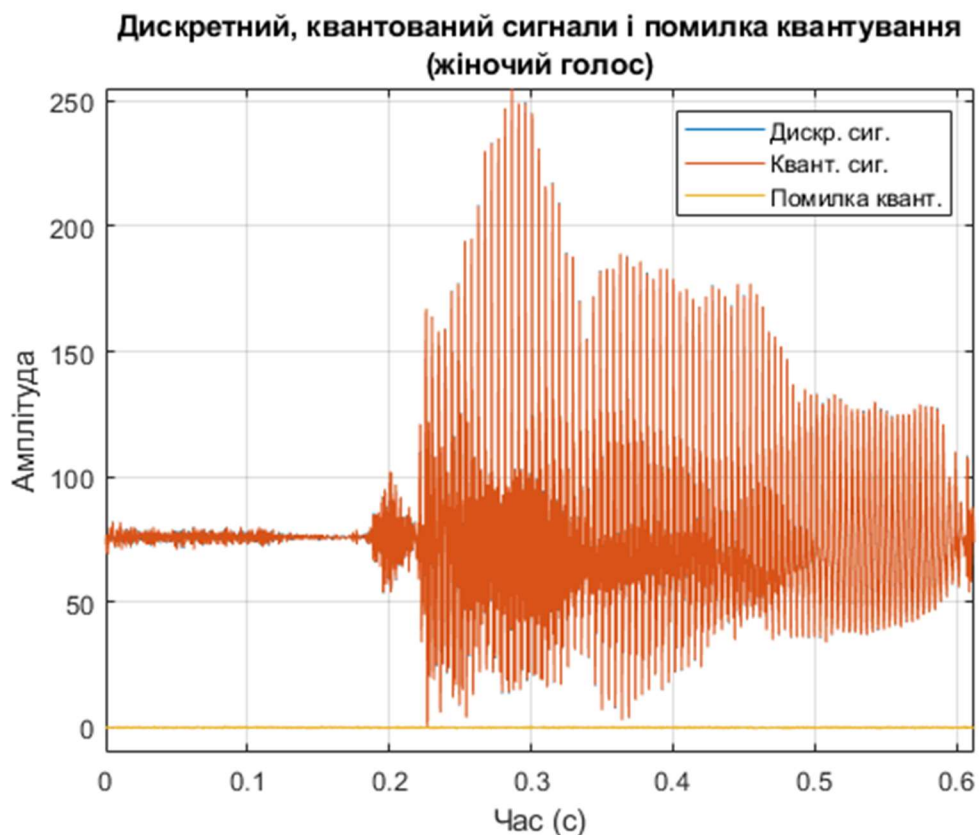
Розрахуємо відношення сигнал/шум квантування (дБ).

$$SNR_q = 6.02 * b + 1.76$$

$$SNR_q = 49.9200$$

Побудуємо графіки досліджуваних мовних сигналів до та після квантування, а також графік помилки квантування.

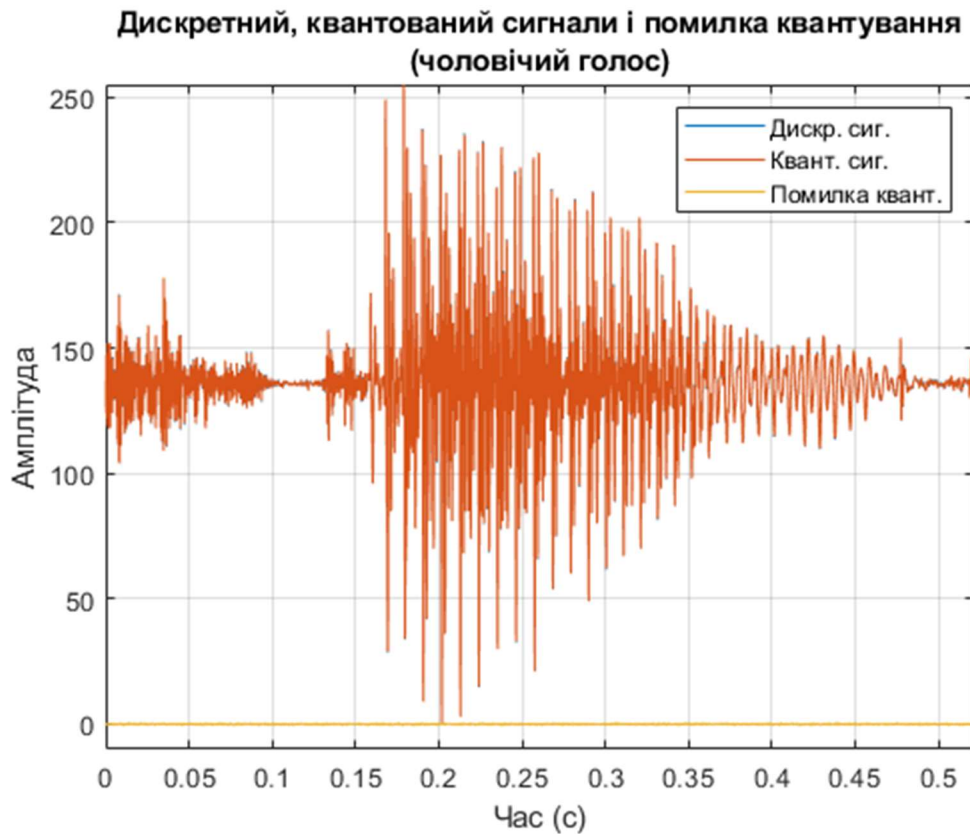
```
figure(5)
plot(dt:dt:t_Sher, Sher_n_L, dt:dt:t_Sher, Sher_q, dt:dt:t_Sher, e_Sher_q);
axis([0 t_Sher -10 max(Sher_q)]); grid on;
legend('Дискр. сиг.', 'Квант. сиг.', 'Помилка квант.', 'Location', 'northeast');
title({'Дискретний, квантований сигнали і помилка квантування'; '(жіночий голос)'});
xlabel('Час (с)'); ylabel('Амплітуда');
```



```

figure(6)
plot(dt:dt:t Shim,Shim n L,dt:dt:t Shim,Shim q,dt:dt:t Shim,e Shim q);
axis([0 t Shim -10 max(Shim q)]); grid on;
legend('Дискр. сиг.','Квант. сиг.','Помилка квант.','Location','northeast');
title({'Дискретний, квантований сигнали і помилка квантування';'(чоловічий голос)'});
xlabel('Час (с)'); ylabel('Амплітуда');

```



Для визначення середньої кількості інформації (ентропія), що припадає на одне значення квантованої амплітуди (біт/відлік) візьмемо логарифм за основою два від загального числа можливих рівнів квантування

$$H_L = \log_2(L)$$

$$H_L = 8$$

за умови, що всі квантовані амплітуди рівномірні, тобто

$$P = 1/L$$

$$P = 0.0039$$

ймовірність появи кожної амплітуди сигналу.

Тоді, також ентропію можна знайти через логарифм ймовірності появи квантованих амплітуд

$$H_P = -\log_2(P)$$

$$H_P = 8$$

Розрахуємо коефіцієнт надмірності досліджуваних сигналів.

$$r_{H_L} = 1 - (H_P/H_L)$$

$$r_{H_L} = 0$$

Знайдемо кількість відліків в досліджуваних сигналах.

$$n_{\text{Sher}} = F_s * t_{\text{Sher}}$$

$$n_{\text{Sher}} = 4899$$

$$n_{\text{Shim}} = F_s * t_{\text{Shim}}$$

$$n_{\text{Shim}} = 4173$$

Отже, визначимо загальну кількість інформації (біт), яка зберігається в сигналах через кількість рівнів квантування

$$I_{\text{Sher}_L} = n_{\text{Sher}} * \log_2(L)$$

$$I_{\text{Sher}_L} = 39192$$

$$I_{\text{Shim}_L} = n_{\text{Shim}} * \log_2(L)$$

$$I_{\text{Shim}_L} = 33384$$

або через ймовірність їх появи

$$I_{\text{Sher}_P} = -n_{\text{Sher}} * \log_2(P)$$

$$I_{\text{Sher}_P} = 39192$$

$$I_{\text{Shim}_P} = -n_{\text{Shim}} * \log_2(P)$$

$$I_{\text{Shim}_P} = 33384$$

Особливий інтерес представляє оцінка швидкості передачі інформації, що міститься в досліджуваних сигналах. Для цього потрібно знайти швидкість передачі відліків сигналів за 1 секунду (відліків/с), що можна зробити через частоту дискретизації

$$W = F_s$$

$$W = 8000$$

через тривалість (с) одного відліку сигналів

$$W_{dt} = 1/dt$$

$$W_{dt} = 8000$$

або через загальну кількість відліків поділену на тривалість (с) сигналів

$$W_{\text{Sher}} = n_{\text{Sher}}/t_{\text{Sher}}$$

$$W_{\text{Sher}} = 8000$$

$$W_{\text{Shim}} = n_{\text{Shim}}/t_{\text{Shim}}$$

$$W_{\text{Shim}} = 8000$$

Розрахуємо коефіцієнт надмірності частоти дискретизації досліджуваних сигналів.

$$r_{F_s} = 1 - (W/F_s)$$

$$r_{F_s} = 0$$

Тоді швидкість передачі інформації (біт/с), яка зберігається в кожному з досліджуваних сигналів без урахування ймовірності появи і взаємкореляції відліків визначається наступним чином

$$R_L = W * \log_2(L)$$

$$R_L = 64000$$

$$R_P = -W * \log_2(P)$$

$$R_P = 64000$$

або через ентропію

$$R_{H_L} = W * H_L$$

$$R_{H_L} = 64000$$

$$R_{H_P} = W * H_P$$

$$R_{H_P} = 64000$$

Тоді найменша можлива пропускна здатність буде дорівнюватиме

```
C_R_L = R_L
```

```
C_R_L = 64000
```

Виконаємо дискретне вейвлет-перетворення досліджуваних сигналів.

```
wname_Sher = 'db12'
```

```
wname_Sher = 'db12'
```

```
[C_dwt_Sher,L_dwt_Sher] =  
wavedec(Sher,wmaxlev(length(Sher),wname_Sher),wname_Sher)
```

```
C_dwt_Sher = 1×5058
```

```
-0.0544 -0.0620 -0.0600 -0.0635 -0.0569 -0.0597 -0.0622 ...
```

```
L_dwt_Sher = 1×9
```

```
61 61 99 175 327 632 ...
```

```
wname_Shim = 'db12'
```

```
wname_Shim = 'db12'
```

```
[C_dwt_Shim,L_dwt_Shim] =  
wavedec(Shim,wmaxlev(length(Shim),wname_Shim),wname_Shim)
```

```
C_dwt_Shim = 1×4330
```

```
-0.0166 -0.0208 -0.0234 -0.0193 -0.0204 -0.0181 -0.0212 ...
```

```
L_dwt_Shim = 1×9
```

```
55 55 87 152 282 541 ...
```

Знайдемо кількість ненульових вейвлет-коефіцієнтів, згідно заданого критерію порогової обробки, який заснований на балансі відсотків енергії і нульових коефіцієнтів досліджуваних сигналів.

```
Num_nonC0_dwt_Sher = 1;  
while  
norm(C_dwt_Sher(ind_C_dwt_Sher(1:Num_nonC0_dwt_Sher)))/norm(C_dwt_Sher)  
< ...  
    (length(C_dwt_Sher)-Num_nonC0_dwt_Sher)/length(C_dwt_Sher)  
    Num_nonC0_dwt_Sher = Num_nonC0_dwt_Sher+1;  
end  
Num_nonC0_dwt_Sher
```

```
Num_nonC0_dwt_Sher = 410
```

```
Num_nonC0_dwt_Shim = 1;  
while  
norm(C_dwt_Shim(ind_C_dwt_Shim(1:Num_nonC0_dwt_Shim)))/norm(C_dwt_Shim)  
< ...  
    (length(C_dwt_Shim)-Num_nonC0_dwt_Shim)/length(C_dwt_Shim)  
    Num_nonC0_dwt_Shim = Num_nonC0_dwt_Shim+1;  
end  
Num_nonC0_dwt_Shim
```

```
Num_nonC0_dwt_Shim = 284
```

Знайдемо кількість нульових вейвлет-коефіцієнтів.

```
Num_C0_dwt_Sher = length(C_dwt_Sher)-Num_nonC0_dwt_Sher
```

```
Num_C0_dwt_Sher = 4648
```

```
Num_C0_dwt_Shim = length(C_dwt_Shim)-Num_nonC0_dwt_Shim
```

```
Num_C0_dwt_Shim = 4046
```

Знайдемо відсоток ненульових вейвлет-коефіцієнтів.

```
Perc_nonC0_dwt_Sher = Num_nonC0_dwt_Sher/length(C_dwt_Sher)*100
```

```
Perc_nonC0_dwt_Sher = 8.1060
```

```
Perc_nonC0_dwt_Shim = Num_nonC0_dwt_Shim/length(C_dwt_Shim)*100
```

```
Perc_nonC0_dwt_Shim = 6.5589
```

Знайдемо відсоток нульових вейвлет-коефіцієнтів.

```
Perc_C0_dwt_Sher = Num_C0_dwt_Sher/length(C_dwt_Sher)*100
```

```
Perc_C0_dwt_Sher = 91.8940
```

```
Perc_C0_dwt_Shim = Num_C0_dwt_Shim/length(C_dwt_Shim)*100
```

```
Perc_C0_dwt_Shim = 93.4411
```

Знайдемо відсоток енергії ненульових вейвлет-коефіцієнтів.

```
Perc_Ener_nonC0_dwt_Sher =  
norm(C_dwt_Sher(ind_C_dwt_Sher(1:Num_nonC0_dwt_Sher)))/ ...  
norm(C_dwt_Sher)*100
```

```
Perc_Ener_nonC0_dwt_Sher = 91.9222
```

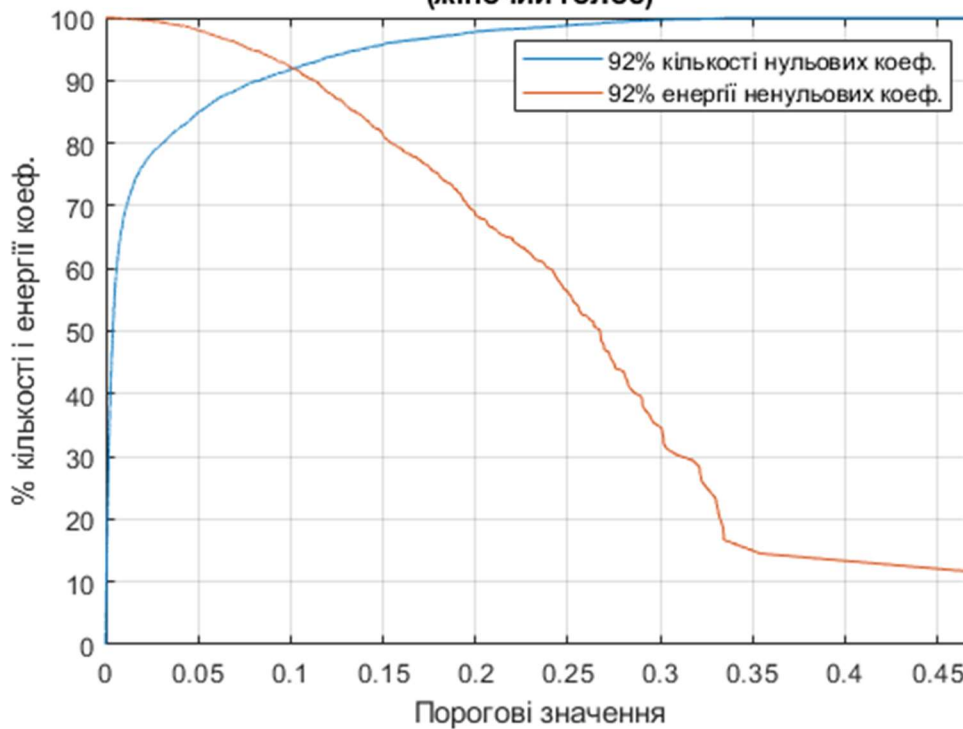
```
Perc_Ener_nonC0_dwt_Shim =  
norm(C_dwt_Shim(ind_C_dwt_Shim(1:Num_nonC0_dwt_Shim)))/ ...  
norm(C_dwt_Shim)*100
```

```
Perc_Ener_nonC0_dwt_Shim = 93.4440
```

Побудуємо залежність відсотків кількості нульових коефіцієнтів і енергії ненульових вейвлет-коефіцієнтів від порогових значень для досліджуваних мовних сигналів.

```
Perc_C0_dwt_Sher_dep = [];  
Perc_Ener_nonC0_dwt_Sher_dep = [];  
Thr_dwt_Sher_dep = [];  
i = 1;  
while i<=length(C_dwt_Sher)  
    Perc_C0_dwt_Sher_dep = [Perc_C0_dwt_Sher_dep; (length(C_dwt_Sher)-  
i)/length(C_dwt_Sher)*100];  
    Perc_Ener_nonC0_dwt_Sher_dep = [Perc_Ener_nonC0_dwt_Sher_dep; ...  
norm(C_dwt_Sher(ind_C_dwt_Sher(1:i)))/norm(C_dwt_Sher)*100];  
    Thr_dwt_Sher_dep = [Thr_dwt_Sher_dep;  
abs(C_dwt_Sher(ind_C_dwt_Sher(i)))];  
    i = i+1;  
end  
figure(17)  
plot(Thr_dwt_Sher_dep,Perc_C0_dwt_Sher_dep,Thr_dwt_Sher_dep,Perc_Ener_  
nonC0_dwt_Sher_dep);  
axis([0 max(Thr_dwt_Sher_dep) 0 100]); grid on;  
legend([int2str(Perc_C0_dwt_Sher) '% кількості нульових коеф.'], ...  
[int2str(Perc_Ener_nonC0_dwt_Sher) '% енергії ненульових коеф.']);  
title({"Залежність відсотків кількості нульових коефіцієнтів і енергії"; ...  
"ненульових вейвлет-коефіцієнтів від порогових значень"; "(жіночий  
голос)"});  
xlabel('Порогові значення'); ylabel('% кількості і енергії коеф.');
```

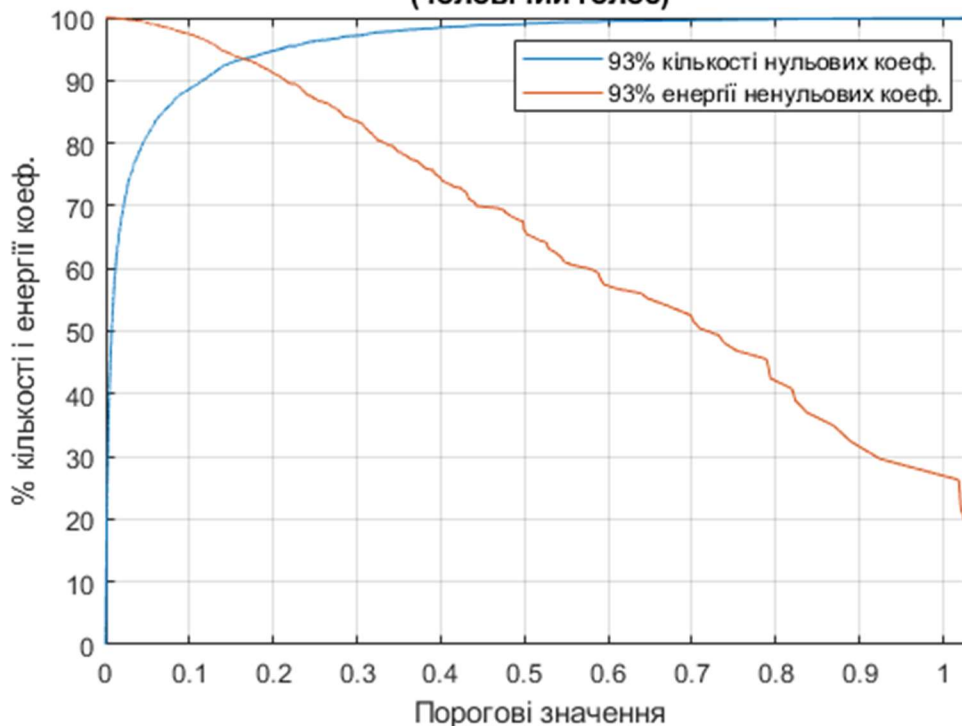
Залежність відсотків кількості нульових коефіцієнтів і енергії
ненульових вейвлет-коефіцієнтів від порогових значень
(жіночий голос)



```

Perc_C0_dwt_Shim_dep = [];
Perc_Ener_nonC0_dwt_Shim_dep = [];
Thr_dwt_Shim_dep = [];
i = 1;
while i <= length(C_dwt_Shim)
    Perc_C0_dwt_Shim_dep = [Perc_C0_dwt_Shim_dep; (length(C_dwt_Shim)-
i)/length(C_dwt_Shim)*100];
    Perc_Ener_nonC0_dwt_Shim_dep = [Perc_Ener_nonC0_dwt_Shim_dep; ...
norm(C_dwt_Shim(ind_C_dwt_Shim(1:i)))/norm(C_dwt_Shim)*100];
    Thr_dwt_Shim_dep = [Thr_dwt_Shim_dep;
abs(C_dwt_Shim(ind_C_dwt_Shim(i)))];
    i = i+1;
end
figure(18)
plot(Thr_dwt_Shim_dep, Perc_C0_dwt_Shim_dep, Thr_dwt_Shim_dep, Perc_Ener
_nonC0_dwt_Shim_dep);
axis([0 max(Thr_dwt_Shim_dep) 0 100]); grid on;
legend([int2str(Perc_C0_dwt_Shim) '% кількості нульових коеф.'], ...
[int2str(Perc_Ener_nonC0_dwt_Shim) '% енергії ненульових коеф.']);
title({'Залежність відсотків кількості нульових коефіцієнтів і енергії'; ...
"ненульових вейвлет-коефіцієнтів від порогових значень"; "(чоловічий
голос)"});
xlabel('Порогові значення'); ylabel('% кількості і енергії коеф.');
```

Залежність відсотків кількості нульових коефіцієнтів і енергії ненульових вейвлет-коефіцієнтів від порогових значень (чоловічий голос)



Знайдемо коефіцієнт зниження надмірності в процесі порогової обробки вейвлет-коефіцієнтів.

```
Red_factor_thr_C_dwt_Sher = length(C_dwt_Sher)/Num_nonC0_dwt_Sher
```

```
Red_factor_thr_C_dwt_Sher = 12.3366
```

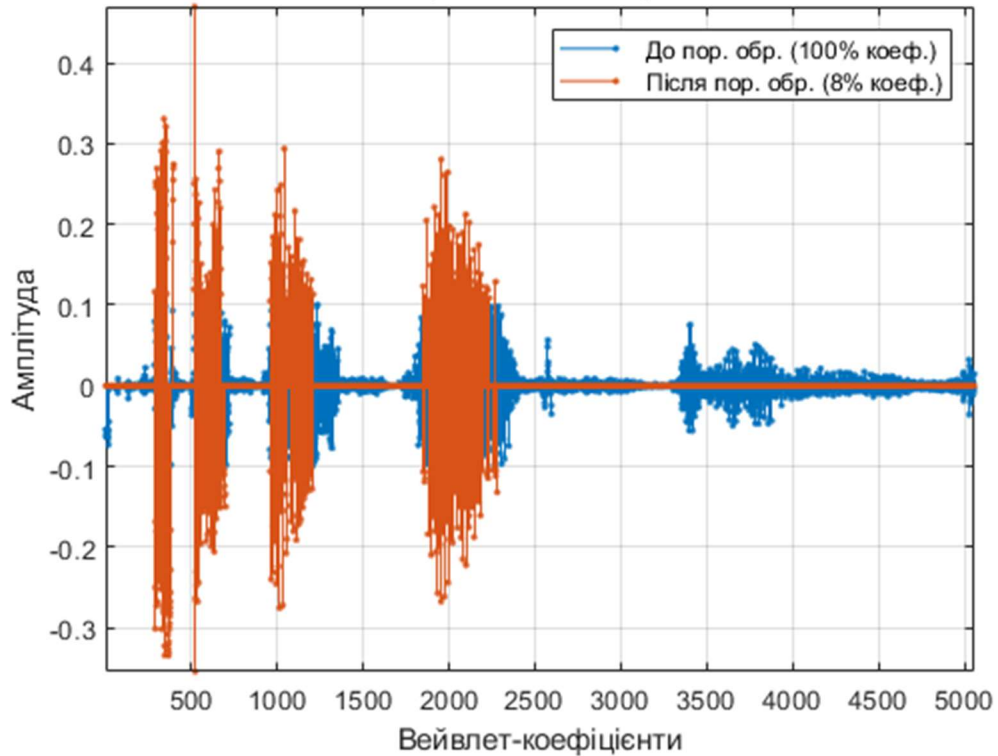
```
Red_factor_thr_C_dwt_Shim = length(C_dwt_Shim)/Num_nonC0_dwt_Shim
```

```
Red_factor_thr_C_dwt_Shim = 15.2465
```

Побудуємо графіки досліджуваних вейвлет-коефіцієнтів мовних сигналів до та після порогової обробки.

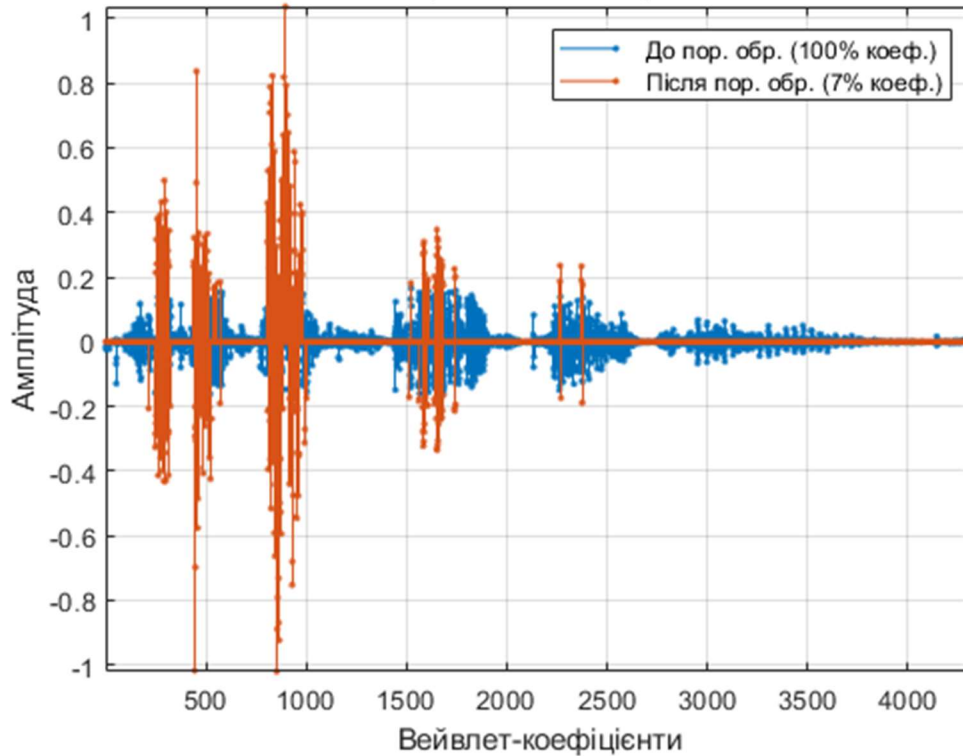
```
figure(17)
stem(C_dwt_Sher, '.'); hold on; stem(C_thr_dwt_Sher, '.'); hold off;
axis([1 length(C_dwt_Sher) min(C_dwt_Sher) max(C_dwt_Sher)]); grid on;
legend('До пор. обр. (100% коеф.)', ['Після пор. обр. ('
int2str(Perc_nonC0_dwt_Sher) '% коеф.)'), ...
'Location', 'northeast');
title({'Коефіцієнти мовного сигналу до та після порогової
обробки'; '(жіночий голос)'});
xlabel("Вейвлет-коефіцієнти"); ylabel('Амплітуда');
```

Коефіцієнти мовного сигналу до та після порогової обробки
(жіночий голос)



```
figure(18)
stem(C_dwt_Shim, '.'); hold on; stem(C_thr_dwt_Shim, '.'); hold off;
axis([1 length(C_dwt_Shim) min(C_dwt_Shim) max(C_dwt_Shim)]); grid on;
legend('До пор. обр. (100% коеф.)', ['Після пор. обр. ('
int2str(Perc_nonC0_dwt_Shim) '% коеф.)'], ...
'Location', 'northeast');
title({'Коефіцієнти мовного сигналу до та після порогової
обробки'; '(жіночий голос)'});
xlabel("Вейвлет-коефіцієнти"); ylabel('Амплітуда');
```

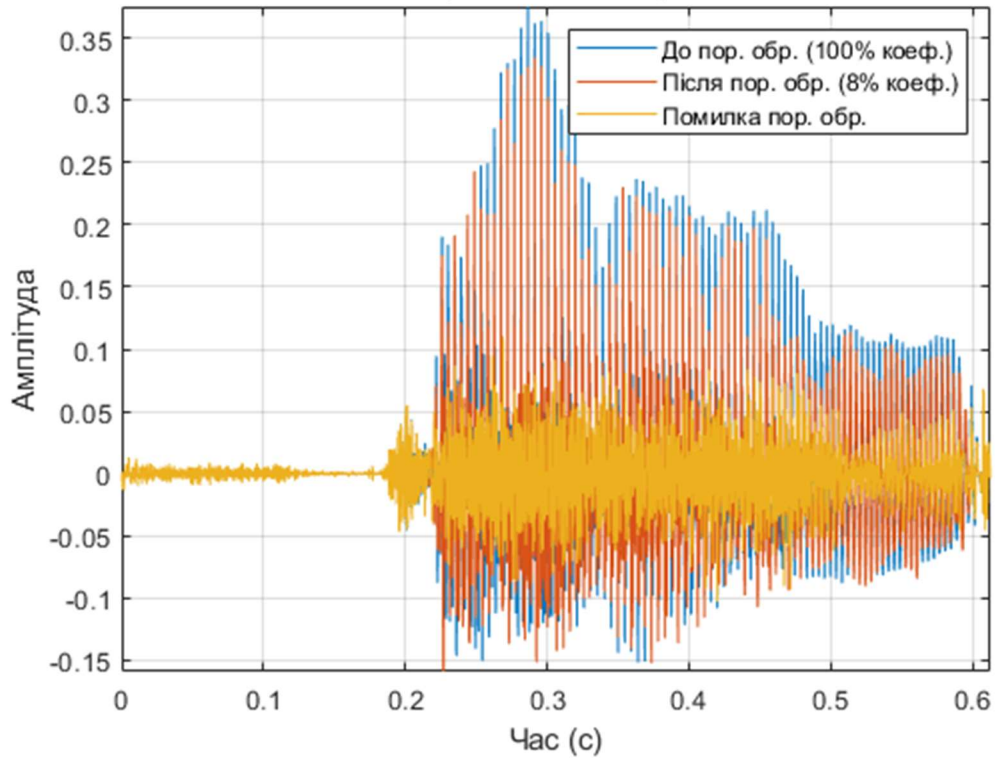
Коефіцієнти мовного сигналу до та після порогової обробки
(жіночий голос)



Побудуємо графіки досліджуваних мовних сигналів до та після порогової обробки на основі вейвлет-перетворення, а також графіки помилки порогової обробки.

```
figure(29)
plot(dt:dt:t_Sher,Sher,dt:dt:t_Sher,Sher_thr_dwt,dt:dt:t_Sher,e_Sher_thr_dwt);
axis([0 t_Sher min(Sher) max(Sher)]); grid on;
legend('До пор. обр. (100% коеф.)','Після пор. обр. ('
int2str(Perc_nonC0_dwt_Sher) '% коеф.)', ...
'Помилка пор. обр.','Location','northeast');
title({'Мовний сигнал до та після порогової обробки та її помилка!';(жіночий
голос)'});
xlabel('Час (с)'); ylabel('Амплітуда');
```

Мовний сигнал до та після порогової обробки та її помилка
(жіночий голос)

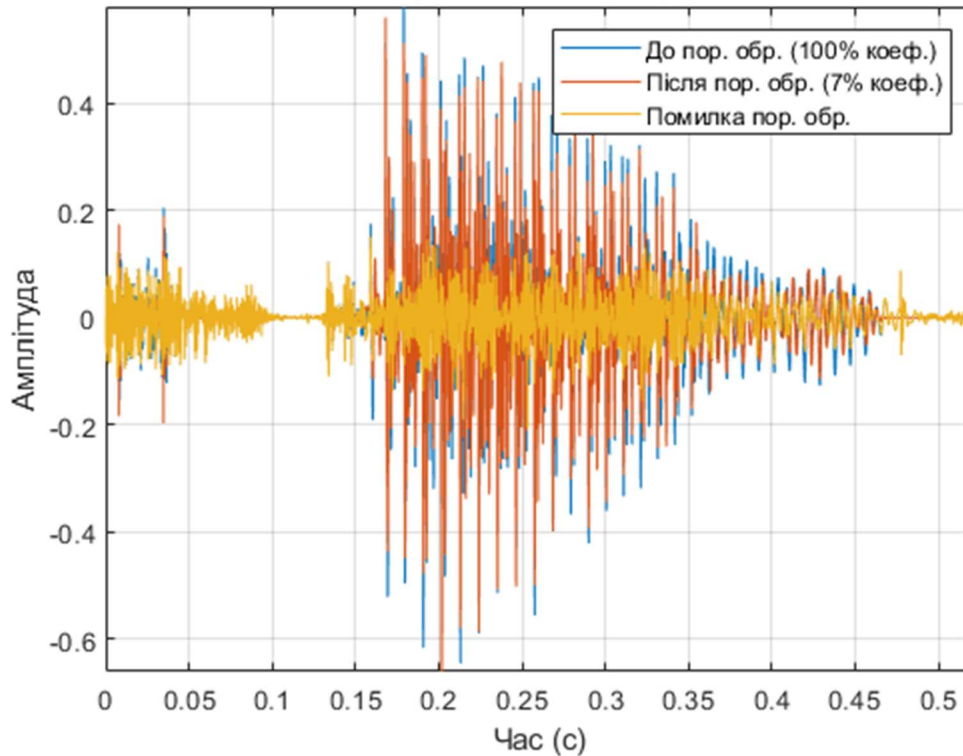


```

figure(30)
plot(dt:dt:t Shim,Shim,dt:dt:t Shim,Shim_thr_dwt,dt:dt:t Shim,e Shim_thr_dwt
);
axis([0 t Shim min(Shim) max(Shim)]); grid on;
legend('До пор. обр. (100% коеф.)','Після пор. обр. ('
int2str(Perc nonC0_dwt Shim) '% коеф.)', ...
'Помилка пор. обр.','Location','northeast');
title({'Мовний сигнал до та після порогової обробки та її
помилка';'(чоловічий голос)'});
xlabel('Час (с)'); ylabel('Амплітуда');

```


Мовний сигнал до та після порогової обробки та її помилка
(чоловічий голос)



Розрахуємо нормовану середньоквадратичну помилку досліджуваних сигналів до та після порогової обробки вейвлет-коефіцієнтів.

```
NRMSE_Sher_thr_dwt = rms(Sher-Sher_thr_dwt)./rms(Sher)
```

```
NRMSE_Sher_thr_dwt = 0.3899
```

```
NRMSE_Shim_thr_dwt = rms(Shim-Shim_thr_dwt)./rms(Shim)
```

```
NRMSE_Shim_thr_dwt = 0.3541
```

Розрахуємо коефіцієнт кореляції досліджуваних сигналів до та після порогової обробки вейвлет-коефіцієнтів.

```
R_Sher_thr_dwt = corrcoef(Sher,Sher_thr_dwt); R_Sher_thr_dwt =  
R_Sher_thr_dwt(1,2)
```

```
R_Sher_thr_dwt = 0.9209
```

```
R_Shim_thr_dwt = corrcoef(Shim,Shim_thr_dwt); R_Shim_thr_dwt =  
R_Shim_thr_dwt(1,2)
```

```
R_Shim_thr_dwt = 0.9352
```

Розрахуємо відношення сигнал/шум (дБ) досліджуваних сигналів до та після порогової обробки вейвлет-коефіцієнтів.

```
SNR_dB_Sher_thr_dwt = 20*log10(rms(Sher)./rms(Sher-Sher_thr_dwt))
```

```
SNR_dB_Sher_thr_dwt = 8.1818
```

```
SNR_dB_Shim_thr_dwt = 20*log10(rms(Shim)./rms(Shim-Shim_thr_dwt))
```

```
SNR_dB_Shim_thr_dwt = 9.0177
```

Розрахуємо пікове відношення сигнал/шум (дБ) досліджуваних сигналів до та після порогової обробки вейвлет-коефіцієнтів.

```
PSNR_dB_Sher_thr_dwt = 20*log10(max(abs(Sher))./rms(Sher-Sher_thr_dwt))
```

```
PSNR_dB_Sher_thr_dwt = 24.5171
```

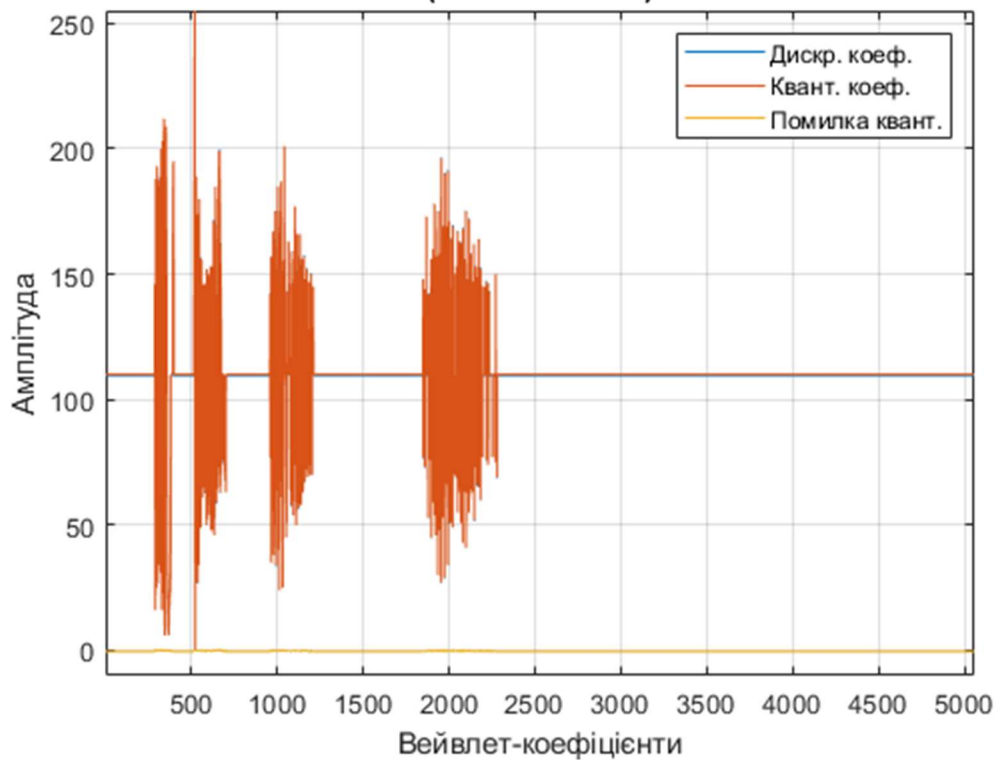
```
PSNR_dB_Shim_thr_dwt = 20*log10(max(abs(Shim))./rms(Shim-Shim_thr_dwt))
```

```
PSNR_dB_Shim_thr_dwt = 25.0310
```

Побудуємо графіки досліджуваних вейвлет-коефіцієнтів мовних, синусоїдального і випадкового сигналів до та після квантування, а також графік помилки квантування.

```
figure(41)
plot(Cthr_dwt_Sher_n_L); hold on; plot(Cthr_dwt_Sher_q);
plot(e Cthr_dwt_Sher_q); hold off;
axis([1 length(Cthr_dwt_Sher_n_L) -10 max(Cthr_dwt_Sher_n_L)]); grid on;
legend('Дискр. коеф.', 'Квант. коеф.', 'Помилка квант.', 'Location', 'northeast');
title({'Дискретні, квантовані коефіцієнти і помилка квантування'; '(жіночий голос)'});
xlabel("Вейвлет-коефіцієнти"); ylabel('Амплітуда');
```

Дискретні, квантовані коефіцієнти і помилка квантування
(жіночий голос)

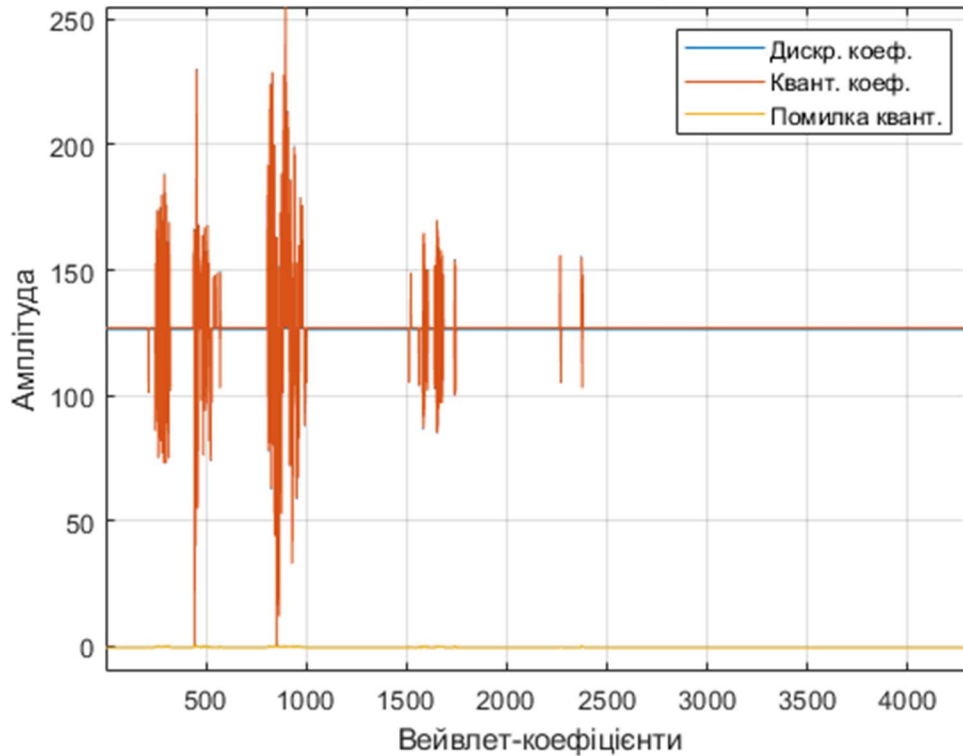


```

figure(42)
plot(Cthr_dwt_Shim_n_L); hold on; plot(Cthr_dwt_Shim_q);
plot(e_Cthr_dwt_Shim_q); hold off;
axis([1 length(Cthr_dwt_Shim_n_L) -10 max(Cthr_dwt_Shim_n_L)]); grid on;
legend('Дискр. коеф.', 'Квант. коеф.', 'Помилка квант.', 'Location', 'northeast');
title({'Дискретні, квантовані коефіцієнти і помилка квантування'; '(чоловічий  
голос)'});
xlabel("Вейвлет-коефіцієнти"); ylabel('Амплітуда');

```

Дискретні, квантовані коефіцієнти і помилка квантування
(чоловічий голос)



Знайдемо ймовірність появи вейвлет-коефіцієнтів досліджуваних сигналів після порогової обробки.

```
Prob_Cthr_dwt_Sher_q =
arrayfun(@(x)length(find(Cthr_dwt_Sher_q==x)),unique(Cthr_dwt_Sher_q)) ...
/length(Cthr_dwt_Sher_q)
```

Prob_Cthr_dwt_Sher_q = 1×126

```
0.0002 0.0004 0.0002 0.0002 0.0002 0.0004 0.0002 0.0002
0.0006 0.0004 0.0002 0.0004 0.0004 0.0004 0.0006 0.0010 0.0004
0.0002 0.0004 0.0002 0.0002 0.0002 0.0008 0.0002 0.0004 0.0002
0.0004 0.0004 0.0002 0.0004
```

```
Prob_Cthr_dwt_Shim_q =
arrayfun(@(x)length(find(Cthr_dwt_Shim_q==x)),unique(Cthr_dwt_Shim_q)) ...
/length(Cthr_dwt_Shim_q)
```

Prob_Cthr_dwt_Shim_q = 1×110

```
0.0005 0.0002 0.0002 0.0002 0.0002 0.0002 0.0002 0.0002
0.0002 0.0002 0.0005 0.0002 0.0005 0.0002 0.0002 0.0005 0.0002
```

0.0002 0.0002 0.0007 0.0007 0.0002 0.0005 0.0002 0.0002 0.0005
0.0005 0.0002 0.0007 0.0007

Тоді умовну ентропію можна знайти через ймовірність появи і кореляційний зв'язок вейвлет-коефіцієнтів досліджуваних сигналів після порогової обробки.

```
H_Prob_Cthr_dwt_Sher_q = -  
sum(Prob_Cthr_dwt_Sher_q.*log2(Prob_Cthr_dwt_Sher_q))
```

H_Prob_Cthr_dwt_Sher_q = 0.9409

```
H_Prob_Cthr_dwt_Shim_q = -  
sum(Prob_Cthr_dwt_Shim_q.*log2(Prob_Cthr_dwt_Shim_q))
```

H_Prob_Cthr_dwt_Shim_q = 0.7657

Розрахуємо коефіцієнт надмірності вейвлет-коефіцієнтів досліджуваних сигналів після порогової обробки враховуючи їх ймовірність появи і кореляційний зв'язок.

```
r_H_Prob_Cthr_dwt_Sher_q = 1-(H_Prob_Cthr_dwt_Sher_q/H_L)
```

r_H_Prob_Cthr_dwt_Sher_q = 0.8824

```
r_H_Prob_Cthr_dwt_Shim_q = 1-(H_Prob_Cthr_dwt_Shim_q/H_L)
```

r_H_Prob_Cthr_dwt_Shim_q = 0.9043

Знайдемо кількість вейвлет-коефіцієнтів досліджуваних сигналів після порогової обробки враховуючи їх ймовірність появи і кореляційний зв'язок.

```
n_uniq_Cthr_dwt_Sher_q = length(unique(Cthr_dwt_Sher_q))
```

n_uniq_Cthr_dwt_Sher_q = 126

```
n_uniq_Cthr_dwt_Shim_q = length(unique(Cthr_dwt_Shim_q))
```

n_uniq_Cthr_dwt_Shim_q = 110

Отже, визначимо загальну кількість інформації (біт), яка зберігається в вейвлет-коефіцієнтах досліджуваних сигналів після порогової обробки враховуючи їх ймовірність появи та кореляційний зв'язок.

$$I_{\text{uniq_Cthr_dwt_Sher_q}} = n_{\text{uniq_Cthr_dwt_Sher_q}} * H_{\text{Prob_Cthr_dwt_Sher_q}}$$

$$I_{\text{uniq_Cthr_dwt_Sher_q}} = 118.5489$$

$$I_{\text{uniq_Cthr_dwt_Shim_q}} = n_{\text{uniq_Cthr_dwt_Shim_q}} * H_{\text{Prob_Cthr_dwt_Shim_q}}$$

$$I_{\text{uniq_Cthr_dwt_Shim_q}} = 84.2227$$

Знайдемо швидкість передачі вейвлет-коефіцієнтів досліджуваних сигналів за 1 секунду (відліків/с), що можна зробити через загальну кількість коефіцієнтів після порогової обробки враховуючи їх ймовірність появи та кореляційний зв'язок поділену на тривалість (с) сигналів.

$$W_{\text{uniq_Cthr_dwt_Sher_q}} = n_{\text{uniq_Cthr_dwt_Sher_q}} / t_{\text{Sher}}$$

$$W_{\text{uniq_Cthr_dwt_Sher_q}} = 205.7563$$

$$W_{\text{uniq_Cthr_dwt_Shim_q}} = n_{\text{uniq_Cthr_dwt_Shim_q}} / t_{\text{Shim}}$$

$$W_{\text{uniq_Cthr_dwt_Shim_q}} = 210.8795$$

Розрахуємо коефіцієнт надмірності частоти дискретизації вейвлет-коефіцієнтів досліджуваних сигналів після порогової обробки враховуючи їх ймовірність появи і кореляційний зв'язок.

$$r_{W_{\text{uniq_Cthr_dwt_Sher_q}}} = 1 - (W_{\text{uniq_Cthr_dwt_Sher_q}} / F_s)$$

$$r_{W_{\text{uniq_Cthr_dwt_Sher_q}}} = 0.9743$$

$$r_{W_{\text{uniq_Cthr_dwt_Shim_q}}} = 1 - (W_{\text{uniq_Cthr_dwt_Shim_q}} / F_s)$$

$$r_{W_{\text{uniq_Cthr_dwt_Shim_q}}} = 0.9736$$

Тоді швидкість передачі інформації (біт/с), яка зберігається в кожному з досліджуваних сигналів з урахуванням ймовірності появи і взаємкореляції вейвлет-коефіцієнтів визначається наступним чином

$$R_{\text{uniq_Cthr_dwt_Sher_q}} = W_{\text{uniq_Cthr_dwt_Sher_q}} * H_{\text{Prob_Cthr_dwt_Sher_q}}$$

$$R_{\text{uniq_Cthr_dwt_Sher_q}} = 193.5888$$

$$R_{\text{uniq_Cthr_dwt_Shim_q}} = W_{\text{uniq_Cthr_dwt_Shim_q}} * H_{\text{Prob_Cthr_dwt_Shim_q}}$$

$$R_uniq_Prob_Cthr_dwt_Shim_q = 161.4622$$

Тоді найменша можлива пропускна здатність для вейвлет-коефіцієнтів після порогової обробки буде дорівнюватиме

$$C_R_uniq_Prob_Cthr_dwt_Sher_q = R_uniq_Prob_Cthr_dwt_Sher_q$$

$$C_R_uniq_Prob_Cthr_dwt_Sher_q = 193.5888$$

$$C_R_uniq_Prob_Cthr_dwt_Shim_q = R_uniq_Prob_Cthr_dwt_Shim_q$$

$$C_R_uniq_Prob_Cthr_dwt_Shim_q = 161.4622$$

Розрахуємо коефіцієнт стиснення досліджуваних сигналів після порогової обробки на основі вейвлет-перетворення.

$$CR_R_uniq_Prob_Cthr_dwt_Sher_q = R_L/R_uniq_Prob_Cthr_dwt_Sher_q$$

$$CR_R_uniq_Prob_Cthr_dwt_Sher_q = 330.5977$$

$$CR_R_uniq_Prob_Cthr_dwt_Shim_q = R_L/R_uniq_Prob_Cthr_dwt_Shim_q$$

$$CR_R_uniq_Prob_Cthr_dwt_Shim_q = 396.3776$$

ДОДАТОК Б. ПАТЕНТИ



УКРАЇНА



ПАТЕНТ

НА КОРИСНУ МОДЕЛЬ № 116514

**ПРИСТРІЙ ДЛЯ ГОЛОСОВОГО РАДІОУПРАВЛІННЯ
ФУНКЦІЯМИ БЕЗПЛОТНОГО ЛІТАЛЬНОГО АПАРАТА З
БУДОВАНИМ ЗАХИСТОМ ВІД НЕСАНКЦІОНОВАНОГО
ДОСТУПУ**

Видано відповідно до Закону України "Про охорону прав на винаходи і корисні моделі".

Зареєстровано в Державному реєстрі патентів України на корисні моделі **25.05.2017**.

Заступник міністра економічного розвитку і торгівлі України

М.І.Тітарчук



ДОДАТОК В. ДОКУМЕНТИ ЩОДО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОГО ДОСЛІДЖЕННЯ

ЗАТВЕРДЖУЮ
Проректор з наукової роботи
С.О. Завгородній
«24» 05 2021 р.



АКТ

впровадження результатів дисертаційної роботи
Лавриненка Олександра Юрійовича «Методи підвищення ефективності семантичного
кодування мовних сигналів»
в навчальний процес Національного авіаційного університету

Ми, що підписалися нижче, декан факультету аеронавігації, електроніки та телекомунікацій (ФАЕТ), к.т.н., Завгородній С.О., завідувач кафедри телекомунікаційних та радіоелектронних систем Факультету аеронавігації, електроніки та телекомунікацій, д.т.н., доц. Одарченко Р.С., склали цей акт про те, що результати дисертаційної роботи асистента кафедри телекомунікаційних та радіоелектронних систем, Лавриненка Олександра Юрійовича на здобуття наукового ступеня кандидата технічних наук «Методи підвищення ефективності семантичного кодування мовних сигналів» використовуються в навчальному процесі факультету аеронавігації, електроніки та телекомунікацій кафедри телекомунікаційних та радіоелектронних систем.

№	Що впроваджено	Форма впровадження	Ефект від впровадження
1	Метод семантичного кодування мовних сигналів на основі мел-частотних кепстральних коефіцієнтів з використанням адаптивного спектрального аналізу	Постановка нової лабораторної роботи з дисципліни «Цифрова обробка сигналів»	Підвищення якості підготовки бакалаврів за спеціальністю 172 «Телекомунікації та радіотехніка»
2	Метод адаптивного емпіричного вейвлет-перетворення в задачах кратномасштабного аналізу та семантичного кодування мовних сигналів	Постановка нової лабораторної роботи з дисципліни «Сигнали та процеси в телекомунікаційних системах»	Підвищення якості підготовки бакалаврів за спеціальністю 172 «Телекомунікації та радіотехніка»
3	Метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення та спектрального аналізу Гільберта	Розробка нового практичного заняття «Методи цифрової обробки мовних сигналів»	Підвищення якості підготовки магістрів за спеціальністю 172 «Телекомунікації та радіотехніка»

Декан ФАЕТ

Завідувач кафедри телекомунікаційних та радіоелектронних систем

С.О. Завгородній

Р.С. Одарченко

Україна

**НАВЧАЛЬНО-НАУКОВО-ВИРОБНИЧИЙ КОМПЛЕКС
«ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНІ СИСТЕМИ»**

04116, м. Київ, Шевченківський район, вул. Богдана Гаврилишина, 24

№177 від 25.05.2019 р.

**АКТ
впровадження результатів дисертаційної роботи
Лавриненка Олександра Юрійовича**

Цим Актом засвідчується, що наступні наукові та практичні результати досліджень, які розроблені в дисертаційній роботі на тему «Методи підвищення ефективності семантичного кодування мовних сигналів» Лавриненка Олександра Юрійовича, використовуються у Навчально-науково-виробничому комплексі «Інформаційно-комунікаційні системи»:

1. Метод семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення та спектрального аналізу Гільберта, що дозволяє підвищити ефективність безпомилкового дистанційного функціонування систем розпізнавання мови та голосового управління технічними об'єктами.

2. Метод адаптивного емпіричного вейвлет-перетворення в задачах низькошвидкісного кодування мовних сигналів, що дозволяє знизити необхідну смугу пропускання, тим самим підвищуючи пропускну здатність каналу передачі мовної інформації.

3. Алгоритм семантичного кодування мовних сигналів на основі емпіричного вейвлет-перетворення та його програмна реалізація мовою програмування MATLAB R2020b, що дозволяє на практиці використовувати запропоновані методи в задачах розпізнавання і синтезу мови, голосового управління технічними об'єктами, низькошвидкісного кодування мовної інформації, голосового перекладу з іноземних мов і т.д.

Директор Навчально-науково-
виробничого комплексу
«Інформаційно-комунікаційні
системи»



Фесенко А.О.