

**ВІННИЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

Факультет комп'ютерних систем та автоматики  
Кафедра комп'ютерних систем управління  
Спеціальність 151 Автоматизація та комп'ютерно-інтегровані технології  
Освітньо-професійна програма Інтелектуальні комп'ютерні системи

**ЗАТВЕРДЖУЮ**

Завідувач кафедри КСУ  
Дубовой В.М.

«\_\_\_» \_\_\_\_\_ 2019 року

**МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА**

Розробка ефективних методів і пристроїв представлення і передачі інформації  
в системах управління

08-01.МКР.002.00.000

Студент групи 2АКІТ-18м Воловодівська О.О.

Керівник к.т.н., доцент Ковтун В.В.

Рецензент к.т.н., доцент Козачко О.М.

Вінницький національний технічний університет

(повне найменування вищого навчального закладу)

Факультет комп'ютерних систем та автоматики

Кафедра комп'ютерних систем управління

Освітньо-кваліфікаційний рівень магістр

Спеціальність 151 – «Автоматизація та комп'ютерно-інтегровані технології»

(шифр і назва)

ЗАТВЕРДЖУЮ

Завідувач кафедри КСУ

В.М. Дубовой

“ 04 ” 09 2019 року

**З А В Д А Н Н Я**

**НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ**

Воловодвіській Ользі Олегівні

(прізвище, ім'я, по батькові)

1. Тема магістерської кваліфікаційної роботи «Нейромережева система автентифікації суб'єкта за візуальним представленням його голосу»

керівник магістерської кваліфікаційної роботи Ковтун В'ячеслав Васильович, к. т. н., доцент

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від “28” 09 2019 року № 267

2. Строк подання студентом магістерської кваліфікаційної роботи 10.12.2019 р.

3. Вихідні дані до магістерської кваліфікаційної роботи максимальний рівень шуму у приміщенні – 70 дБ; підтримка ОС – Windows; максимальний час завантаження – 5 с; автентифікація суб'єктів – так; максимальна кількість запитів до системи – до 1000 шт/с; мови графічного та голосового інтерфейсів – українська, англійська; достовірність автентифікація суб'єктів – понад 90%; використання кодування під час передавання даних – так.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити) вступ, аналіз об'єкту дослідження та варіантний огляд методів автентифікації суб'єктів за голосом, постановка задачі, аналіз стану розв'язання науково-технічної задачі, обґрунтування критеріїв для порівняння, проектування розв'язання задачі автентифікації суб'єктів за голосом, практична реалізація та дослідження ефективності вейвлет-представлення мовних сигналів при автентифікації суб'єктів за голосом, економічна частина.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень) структурна схема системи, UML діаграма варіантів використання, UML-діаграма діяльності, алгоритми попереднього оброблювання сигналу, виділення інформативних ознак та процесу автентифікації суб'єктів за голосом, UML-діаграма послідовності, UML-діаграма станів, екранні форми системи автентифікації суб'єктів за голосом.

## 6. Консультанти розділів магістерської кваліфікаційної роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1, 2, 3	к.т.н., доцент Ковтун В.В.		
4	к.т.н., доцент кафедри ЕПВМ Ратушняк О.Г.		

7. Дата видачі завдання “ 04 ” 09 2019 року

## Календарний план

№ з/п	Назва етапів роботи	Строк виконання етапів роботи	Примітка
1	Аналіз об'єкту дослідження та варіантний огляд методів автентифікації суб'єктів за голосом. Постановка задачі дослідження	12.09.2019 р.	
2	Аналіз стану розв'язання науково-технічної задачі, обґрунтування критеріїв для порівняння	22.09.2019 р.	
3	Проектування розв'язання задачі автентифікації суб'єктів за голосом	3.10.2019 р.	
4	Реалізація системи та дослідження ефективності вейвлет-представлення мовних сигналів для автентифікації суб'єктів за голосом	12.11.2019 р.	
5	Підготовка економічної частини	22.11.2019 р.	
6	Апробація результатів дослідження	25.11.2019 р.	
7	Публікації		
8	Оформлення пояснювальної записки, графічного матеріалу і презентації	02.12.2019 р.	
9	Захист МКР	10.12.2019 р.	

Студент \_\_\_\_\_  
( підпис )

Воловодівька О.С.  
(прізвище та ініціали)

Керівник магістерської кваліфікаційної роботи \_\_\_\_\_  
( підпис )

Ковтун В.В.  
(прізвище та ініціали)

## ЗМІСТ

АНОТАЦІЯ.....	6
ВСТУП.....	8
РОЗДІЛ 1 ОГЛЯД МЕТОДІВ ВІЗУАЛІЗАЦІЇ ТА ОБРОБЛЕННЯ МОВНИХ СИГНАЛІВ.....	11
1.1 Огляд і аналіз існуючих систем автентифікації суб'єктів за голосом .....	13
1.2 Цифровий аналіз та оброблення мовних сигналів.....	14
1.2.1 Спектральний аналіз .....	14
1.2.2 Вейвлет-аналіз.....	16
1.2.3 Фільтрація сигналів.....	22
1.2.4 Згортка сигналів.....	23
1.3 Оброблення сигналів в часовому просторі .....	24
1.4 Постановка задачі дослідження.....	26
РОЗДІЛ 2 РОЗРОБКА СИСТЕМИ АВТЕНТИФІКАЦІЇ СУБ'ЄКТІВ ЗА ГОЛОСОМ ІЗ ВІЗУАЛІЗАЦІЄЮ МОВНИХ СИГНАЛІВ.....	29
2.1 Аналіз основних функцій системи .....	29
2.2 Розроблення структурної схеми системи .....	29
2.3 Проектування системи з використанням UML-діаграм .....	31
2.4 Розробка алгоритму нелінійного методу часової нормалізації.....	35
2.5 Алгоритм фільтрації мовного сигналу.....	40
2.6 Алгоритм дискретного вейвлет-перетворення.....	42
РОЗДІЛ 3 ТЕСТУВАННЯ І ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ СТВОРЕНОЇ СИСТЕМИ.....	45
РОЗДІЛ 4 ЕКОНОМІЧНИЙ РОЗДІЛ.....	52
4.1 Технологічний аудит розробленої системи автентифікації суб'єктів за голосом .....	52
4.2 Прогнозування витрат на виконання даної роботи.....	57
4.3 Прогнозування комерційних ефектів від реалізації результатів	

розробки.....	62
ВИСНОВКИ.....	67
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	68
ДОДАТКИ.....	73
Додаток А (обов'язковий) Технічне завдання .....	73
Додаток Б (обов'язковий) Лістинг програми .....	76
Додаток В (обов'язковий) Перелік графічних матеріалів.....	79

## АНОТАЦІЯ

У магістрській кваліфікаційній роботі створено нейромережеву систему для аутентифікації суб'єкта за візуалізованим представленням його голосу. Для візуалізації мовних сигналів у системі використовувалося вейвлет-перетворення, а нейромережа використовувалася для аутентифікації суб'єкта.

Проаналізовано методи вилучення інформативних ознак з мовних сигналів, за допомогою яких відбувалася автентифікація суб'єкта. Проведено огляд нейромереж. Створено алгоритми, необхідні для функціонування системи аутентифікації суб'єкта за голосом. Створено алгоритми для попередньої обробки мовних сигналів, підбору інформативних для аутентифікації суб'єкта за голосовими ознак тощо. Створено UML-діаграми використання, активності, послідовності та стану. Проведено оцінювання ефективності створеної нейромережевої системи аутентифікації суб'єкта за візуальним представленням його голосу

## ABSTRACT

In the master's qualification work the neural network system of authentication of the subject on visual representation of his voice was created. Wavelet transforms were used to visualize speech signals in the system, and a convolutional neural network was used to authenticate the subject.

Methods of extracting basic informative features from speech signals are analyzed, using which can be authenticated by the subject. Neural networks are reviewed and classified. The algorithms necessary for the functioning of the subject's voice authentication system have been created. Algorithms have been created for the pre-processing of speech signals, the selection of informative ones for the authentication of a subject by voice traits, for the classification of voices, etc. UML diagrams of usage, activity, sequence, and status are created. The efficiency of the created neural network system of authentication of the subject by visual representation of his voice is conducted.

## ВСТУП

**Актуальність теми.** Методи цифрового оброблювання і аналізу мовних сигналів в даний час активно розвиваються. Ця обставина перш за все обумовлена прогресом у сфері цифрової мікросхемотехніки, завдяки чому стало реальна можливим виготовлення компактної і складної цифрової апаратури передавання і аналізу інформації, а також застосування цих цифрових пристроїв для автентифікації суб'єкта за його голосом. Перші прототипи таких пристроїв вже активно проходять лабораторні дослідження і викликають підвищений інтерес у розробників і залучають нових прихильників до досліджень у цьому напрямку для вивчення існуючих та розробки нових методів і алгоритмів цифрового оброблення мовних сигналів [1].

Автентифікація суб'єкта за голосом засновано на аналізі унікальних характеристик мовного сигналу, обумовлених анатомічними властивостями (розмір і форма рота, горла, голосових зв'язок тощо) та набутими мовленнєвими звичками (манера, гучність, швидкість мови). Перевагою цього підходу до автентифікації суб'єкта за голосом є, насамперед, невіддільність мовної інформації від людини, природність мовного спілкування тощо.

Однією з основних проблем при автентифікації суб'єкта за голосом є те, що парольна слово або фраза може вимовлятися із довільною швидкістю. Через це тривалість парольного мовного сигналу не співпадає з тривалістю відповідного еталону [2]. Отже, виникає потреба в створенні алгоритму нормалізації тривалості звучання мовних сигналів, який можна створити застосувавши методи машинного навчання для візуалізації мовного сигналу.

У нас в країні роботи в області аналізу, перетворення і синтезу мовних сигналів проводяться науковими колективами під керівництвом М.Бондаренка, Т.Вінцюка, І.Сіроджі, Ю.Василенка, Ю.Рашкевича тощо. За кордоном вагомий внесок в цю сферу внесли Е.Моулінез, Д.Малах, Т.Кват'єрі, Р.МакАулау, Р.Сузуки, Д.Роукос.



У більшості відомих напрацювань передбачено рівномірну модифікацію тривалості мовних сигналів без врахування особливостей часових змін окремих класів звуків. Також недостатньо досліджено питання зміни структури мовленого сигналу в системах автентифікації суб'єкта за голосом.

### **Зв'язок роботи з науковими програмами, планами, темами.**

Магістерську кваліфікаційну роботу виконано на кафедрі комп'ютерних систем управління Вінницького національного технічного університету. Дана робота співзвучна із основними науковими напрямами кафедри. Задачі, розглянуті у роботі, відповідають кафедральній науково-дослідній роботі №46К4 «Методи моделювання та оптимізації складних систем на основі інтелектуальних технологій».

**Мета і завдання дослідження.** Метою є розроблення нейромережевої системи автентифікації суб'єкта за візуальним представленням його голосу.

**Об'єкт дослідження** - процес мовотворення людини.

**Предметом дослідження** - методи візуалізації мовних сигналів, математичні методи оброблювання цифрових рядів, методи цифрового оброблювання сигналів, методи теорії прийняття рішень і машинного навчання.

### **Задачі дослідження:**

1. Здійснення аналітичного огляду і аналізу досягнень в галузі систем автентифікації суб'єкта за голосом;
2. Розроблення методу візуалізації мовних сигналів в задачі автентифікації суб'єкта за голосом;
3. Дослідження ефективності розробленого методу за допомогою відповідної програмної системи для автентифікації особи суб'єкта за голосом.

**Наукова новизна одержаних результатів.** Вдосконалено метод візуалізації мовних сигналів, який, на відміну від існуючих, використовує дискретне вейвлет-представлення мовних сигналів в частотних смугах, де присутні інформативні для автентифікації суб'єкта за голосом ознаки, що дозволяє проводити розпізнавання особи суб'єкта згортковою нейромережею.

**Практичне значення одержаних результатів** полягає у розробленні алгоритма візуалізації мовних сигналів в задачі автентифікації суб'єкта за голосом та створенні відповідної програмної системи.

**Апробація результатів.** За результатами магістерської кваліфікаційної роботи подано тези доповіді «Нейромережева система автентифікації суб'єкта за візуальним представленням його голосу» на Всеукраїнську науково-практичну Інтернет-конференцію студентів, аспірантів та молодих науковців «Молодь в науці: дослідження, проблеми, перспективи – 2019».

## 1 ОГЛЯД МЕТОДІВ ВІЗУАЛІЗАЦІЇ ТА ОБРОБЛЕННЯ МОВНИХ СИГНАЛІВ

### 1.1 Огляд і аналіз існуючих систем автентифікації суб'єктів за голосом

Автентифікації суб'єктів за голосом за деяким його висловленням називають іще розпізнаванням мовця. Розпізнавання мовця за голосом часто використовується для підвищення безпеки процедури доступу до системи. Ця технологія активно використовується у роботі банків та інших профільних організацій. Основною перевагою такого підходу до розпізнавання особи є можливість дистанційної перевірки користувачів на право доступу до інформації. Процес починається з вимовлення суб'єктом паролі чи фрази. Таким чином, перетворюючи мовний сигнал створюються акустичні характеристики голосу суб'єкта, які реєструються у базі даних-еталонів. А згодом при автентифікації достатньо вимовити пароль, щоб система розпізнала голос користувача, порівнявши із вже записаними зразками у базі даних. Процес автентифікації не займає багато часу і практично співпадає із часом вимовлення паролю. Недоліком системи є невисока гнучкість, оскільки, наприклад, хрипота чи ларингіт можуть суттєво спотворити параметри голосу і розпізнавання стане неможливим [3].

Останні закордонні дослідження у галузі мовних технологій спрямовано на удосконалення автоматичних систем для автентифікації суб'єктів за голосом, що пов'язано із потребою забезпечення безпеки суспільства та розширенням сфери комерційного застосування мовних технологій. Розробкою систем для автентифікації суб'єктів за голосом займаються десятки організацій, як крупні корпорації, так і невеликі фірми [4]. У Європі з 1994 року діє спеціалізована програма COST 250 «Автоматичне розпізнавання голосу в телефонних мережах», головним виконавцем якої є IDIAP - один з найвідоміших колективів у напрямку розробки мовних технологій у світі. В США і Канаді також є декілька розробників систем для автентифікації суб'єктів за голосом. В цій сфері працюють такі корпорації як Nortel, ІТТ, АТТ, Google та ін. З-поміж сильних дослідницьких груп можна відхачити колективи Bell Lab., BBN MIT. Відомі також групи і у Росії, де

ведуться дослідження у галузі розпізнавання голосу, наприклад, Акустичний інститут РАН (автоматичне розпізнавання голосу в телефонному каналі), Московський лінгвістичний університет (суб'єктивний аналіз відеограм), "Одітек-РД" (розробка програмного забезпечення для автоматичної автентифікації суб'єктів за голосом (зокрема у телефонному каналі), створено мовні бази даних для тестування і навчання систем автентифікації суб'єктів за голосом).

Наведемо декілька прикладів систем автентифікації суб'єктів за голосом, де застосовується нормалізація мовних сигналів.

Система фірми Texas Instruments (TI) була запропонована Доддингтоном для автоматизації функціонування контрольно-пропускного пункту (КПП). Мовний сигнал через мікрофон та підсилювач подається до нормалізатора рівня сигналу. Роль нормалізатора здійснює схема автоматичного регулювання підсилення, яка дозволяє працювати у динамічному діапазоні частотою 40 дБ. Нормалізований сигнал далі надходить на 16-канальний банк смугових фільтрів, які рівномірно перекривають частотний діапазон 300-3000 Гц. Смуга пропускання фільтрів дорівнює 220 Гц. Після детектування мовної активності, вихідні сигнали надходять на фільтри низьких частот і згладжуються. В результаті такої обробки кожна фраза описується  $16 \times 6 \times 7 = 672$  бітами інформації, що вважається еталонним описом мовного сигналу.

Система фірми Bell Laboratories засновується на ідеях, запропонованих Доддингтоном і вдосконалених Розенбергом, Ламмісом та Сембуром. Особливістю даної системи є використання для автентифікації голосу параметрів мовного сигналу, які залишаються стійкими при низькому рівні співвідношення сигнал/шум, зокрема, при передачі мовних сигналів до ЕОМ телефонним каналом зв'язку. Після запису паролльної фрази, проводиться її аналіз і виділення індивідуальних параметрів. Потім відбувається нормалізація тривалості звучання паролльної фрази методом нелінійного ущільнення та розтягування для можливості коректного співставлення із еталоном. Порівняння контурів інтенсивності дох воляє створити деформуючу функцію, яка використовується для порівняння також і інших контурів. Обчислена за всіма контурами розбіжність порівнюється із пороговим значенням та

використовується для прийняття остаточного рішення.

Відомі технології VoiceKey і VoiceNet, які розроблено компанією «Центр мовних технологій» для розпізнання або верифікації мовця використовують паролльні фрази тривалістю до 5 секунд кожна або фрагменти спонтанного мовного сигналу тривалістю понад 16 сек. У обох технологіях в якості індивідуальних характеристик голосу використовуються положення резонансних максимумів у спектрі мовного сигналу, що забезпечує надійну роботу у шумному оточенні і при незначних змінах емоційного стану мовця [5].

Основні недоліками аналогів – суворі вимоги до апаратних ресурсів складових системи та, відповідно, висока ціна.

## 1.2 Цифровий аналіз та оброблення мовних сигналів

До цифрового аналізу і оброблення сигналів відносять задачі [6]:

- лінійної фільтрації, або селекції сигналу у частотній області; синтез фільтрів, узгоджених із сигналом;
- спектрального аналізу - оброблення мовних або звукових сигналів; розпізнавання образів;
- частотно-часового аналізу, зокрема, компресії сигналів, різноманітні задачі виявлення сигналів;
- адаптивну фільтрацію - оброблення мовних сигналів, зображень, розпізнавання образів, пригнічення шумів;
- нелінійне оброблення - обчислення кореляцій, медіанну фільтрацію тощо;
- швидкісне оброблення - інтерполяція (збільшення) і децимація (зменшення) частоти дискретизації у швидкісних каналах телекомунікації, аудіосистемах тощо;
- згортки (лінійної і циклічної).

Розглянемо детальніше деякі з цих задач, які в згодом будуть використовуватися у системі, що створюється.

### 1.2.1 Спектральний аналіз

Для автентифікації суб'єктів за голосом мовний сигнал потрібно представити у цифровій формі.

Частотний діапазон мови лежить у межах 70-7000 Гц. При оцінюванні рівня гучності звуку як еталон звукового тиску  $P_0$  використовується мінімальне значення його енергії на частоті 1 кГц, при якому звук вже можна почути:  $P_0 = 2 \times 10^{-5} \text{ Н / м}^2$ .

Рівень звукового тиску описується співвідношенням:

$$L = 20 \lg \frac{P}{P_0} \text{ [Дб]}, \quad (1.1)$$

де  $P$  - значення звукового тиску.

Під динамічним діапазоном розуміють різницю між максимальним і мінімальним рівнями енергії сигналів. Динамічний діапазон мови - 35-45 Дб.

Цифрова система звукозапису представляє аналоговий мовний сигнал у цифровому вигляді. В результаті аналого-цифрового перетворення (АЦП) аналоговий сигнал переводиться в послідовність дискретних відліків  $sk$ , кожен з яких представляє собою дійсне число, яке характеризує аналоговий сигнал у цій точці часу з певною точністю. Точність представлення залежить від ширини діапазону одержуваних чисел, тобто від розрядності АЦП. Процес розбиття сигналу на відліки називають дискретизацією. Кількість відліків в секунду називають частотою дискретизації. Згідно теореми Котельникова, частота дискретизації  $\nu D$  повинна бути, принаймні, вдвічі більшою від максимальної частоти претворюваного аналогового сигналу. Справа у том, що спектр сигналу, перетвореного в цифрову форму за допомогою АЦП, має періодичний характер. Спектр сигналу - це його відображення, що характеризує інтенсивність і розподіл частотних складових сигналу у часі [7].

Процес оцінювання сигналу із округленням до розряду АЦП називають квантуванням. Зважаючи на потрібний динамічний діапазон цифрової системи звукозапису, необхідну кількість розрядів квантування можна визначити як

$$D = 6n + 1.8, \quad (1.2)$$

де  $D$  - динамічний діапазон, Дб,  $n$  - кількість двійкових розрядів. Звідки отримаємо, що для запису мови необхідно виділяти не менше восьми біт на кожен відлік.

Вочевидь, ефективність автентифікації суб'єктів за голосом залежить від обраної паролльної фрази. тому є необхідність проведення досліджень залежності імовірності правильної автентифікації суб'єктів за голосом від тривалості паролльної фрази. Паролльна фраза повина по можливості бути короткою, щоб не витратити надмірну кількість ресурсів машинного часу для її аналізу, мати фонетичний склад, який дозволяє суб'єкту проявити індивідуальні особливості голосу для одержання вектору інформативних ознак, легко запам'ятовуватися і вимовлятися і т.ін. Також ефективність автентифікації суб'єктів за голосом залежить від емоційного та фізичного стану мовця під час вимовляння паролльної фрази і від рівня шумів навколишнього середовища, які спотворюють мовний сигнал, за яким проводиться автентифікація. Отже, впливає необхідність дослідження залежності імовірності правильної автентифікації суб'єктів за голосом від тривалості паролльної фрази.

Для виділення інформативних ознак необхідно провести аналіз мовного сигналу. Найрозповсюдженішим методом аналізу мовних сигналів вважають короткочасний спектральний аналіз [8], заснований на припущенні, що у межах відносно короткого інтервалу часу голос може вважатися стаціонарним процесом. Це припущення дозволяє досліджувати актуальні часові зміни у мовному сигналі. Короткочасний спектр сигналу  $f(t)$  математично визначається через циклічну частоту  $\omega$  і час  $t$  таким виразом:

$$F(\omega, t) = \int_{-\infty}^t f(\tau)h(t - \tau)e^{-j\omega\tau} d\tau, \quad (1.3)$$

де  $h(t)$  - вагова віконна функція, яка описує лише частину сигналу, який відповідає аналізованому моменту часу. На практиці ефективна тривалість функції  $h(t)$  складає 20-30 мс. Часто поряд із короткочасним спектром обчислюють амплітудний спектр  $|F(\omega, t)|$  і енергетичний спектр  $|F(\omega, t)|^2$  тощо.

Енергетичний спектр визначається так:

$$|F(\omega, t)|^2 = a^2(\omega, t) + b^2(\omega, t). \quad (1.4)$$

Автентифікувати суб'єктів за голосом можна порівнявши енергетичні спектри на окремому сегменті еталонного мовного сигналу та парольного. Приклади енергетичних спектрів представлено на рисунку 1.1.

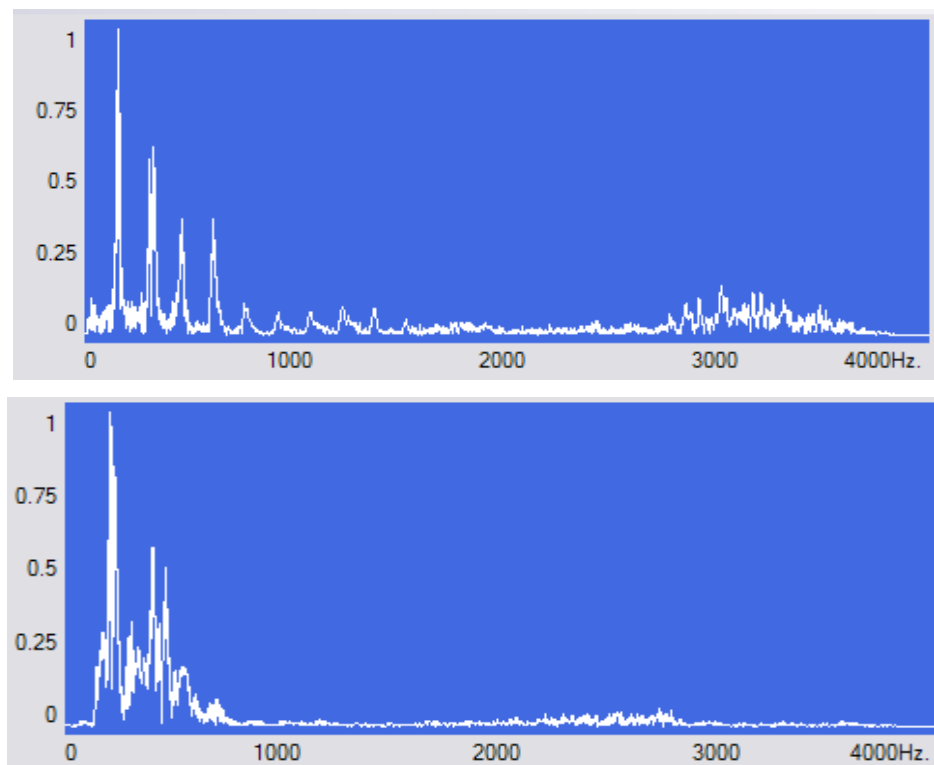


Рисунок 1.1 – Енергетичні спектри мовних сигналів

### 1.2.2 Вейвлет-аналіз



Вейвлети (wavelets) - це узагальнена множина часових функцій, які мають вигляд хвильових наборів певної форми, локалізованих незалежною часовою змінною  $t$  або  $x$  і здатних до масштабування і зсуву по ній. Вейвлети створюються за допомогою специфічних базових функцій-прототипів, які задають форму і властивості вейвлету.

На сьогоднішній день вейвлет-аналіз є однією із найбільш перспективних технологій аналізу даних, його інструменти застосовують у самих різних сферах інтелектуальної діяльності людей.

При дослідженні і оброблюванні сигналу багато інформації можна отримати аналізуючи його частотні характеристики. Оскільки вейвлети мають гнучку частотно-часову адаптацію, вони могли б стати зручним інструментом для опису частотних характеристик сигналу. Крім того, зв'язок між масштабом вейвлет-перетворення і його частотною областю інтуїтивно зрозумілий: чим крупніший масштаб, тим більш низькі частоти аналізуються. Однак, необхідно встановити відповідність між масштабом вейвлета і його частотними характеристиками [9] для практичного застосування вейвлетів в спектральному аналізі сигналів.

Завдяки прекрасному поданню локальних властивостей сигналів, принципово не властивому Фур'є-аналізу, і безлічі форм, вейвлети практично застосовують для аналізу особливостей, стиснення та очищення від шуму складних сигналів. У Фур'є-аналізі в якості базисних функцій використовуються синусоїди. Вони чітко локалізовані у частотній області, але дуже погано локалізовані (фактично, взагалі не локалізовані) у часовій області. Протилежним прикладом слугує імпульсна базисна функція, яка чітко локалізована у часовій області, а отже ідеально підходить для опису розривів у сигналі. Але ця базисна функція не дає інформації про частоту сигналу і тому не пристосована для представлення сигналів, особливо періодичних, на заданому часовому відрізку.

Вейвлети беруть еаще від розглянутих крайніх випадків (синусоїдою і імпульсною функцією). Різні функції, в тому числі, модульованні імпульсами синусоїди, функції зі стрибками рівня і т.д. можуть бути базисними функціями вейвлетів. Це забезпечує просте представлення сигналів із локальними стрибками та

розривами наборами вейвлетів відповідного типу. Не маючи аналітичного представлення у вигляді однієї формули, вейвлети можуть даватися ітераційними виразами.

Перша похідна від функції Гаусса дає WAVE-вейвлет:

$$\psi_w(t) = -\frac{d}{dt} \left[ \exp\left(-\frac{t^2}{2}\right) \right]. \quad (1.5)$$

Для wave-вейвлета спектр Фур'є опишемо як:

$$\psi_w(k) = jk \exp\left(-\frac{k^2}{2}\right). \quad (1.6)$$

Друга похідна від функції Гаусса дає Mhat-вейвлет:

$$\psi_M(t) = \frac{d^2}{dt^2} \left[ \exp\left(-\frac{t^2}{2}\right) \right]. \quad (1.7)$$

Для Mhat -вейвлета спектр Фур'є опишемо як:

$$\psi_M(k) = 2(jk)^2 \exp\left(-\frac{k^2}{2}\right). \quad (1.8)$$

Побудова базису вейвлет-перетворення відбувається на сонові однієї функції - материнського вейвлету. На рисунку 1.2 представлено приклади часто вживаних вейвлетів і відповідних їм Фур'є-аналогів [10].

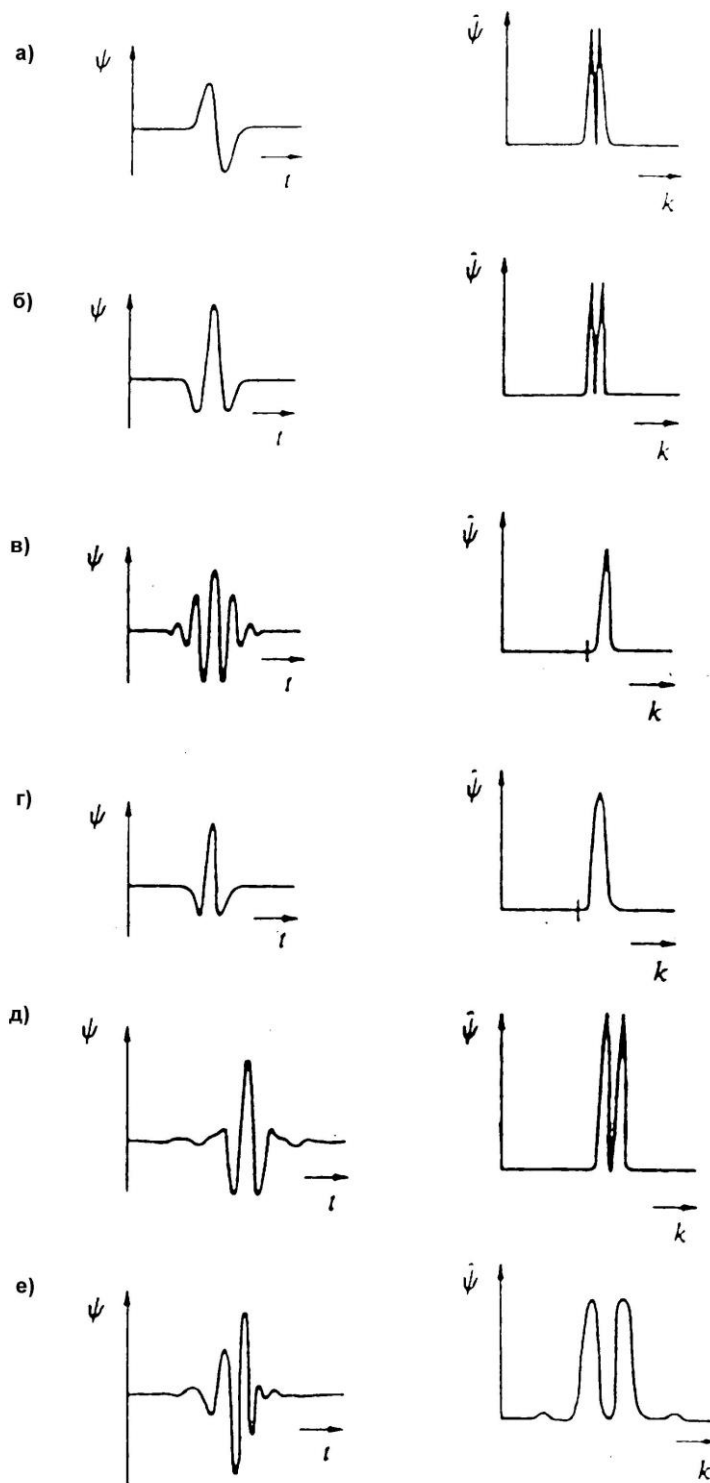


Рисунок 1.2 – Вейвлети: а) WAVE б) МНАТ в) Морле г) Пауля д) LMB  
е) Добеші

Один із найвідомих ортогональних вейвлетів називають вейвлетом Хаара (haar). Зображені на рисунку 1.3 вейвлети Хаара добре локалізовані у часовому просторі, але не так добре локалізовано у частотній області, оскільки меандр функції містить широкий спектр частот. В загальному вигляді для одновимірного сигналу перетворення Хаара виглядає так:

$$a_j = \frac{s_{2j} + s_{2j+1}}{2},$$

$$d_j = \frac{s_{2j} - s_{2j+1}}{2}. \quad (1.9)$$

Прикладі перетворення Хаара добре характеризує структуру вейвлет-перетворення дискретного сигналу. На кожному кроці перетворення сигнал розбивається на дві складові: наближення з більш низькою роздільною здатністю - апроксимацію і деталізуючу складову.

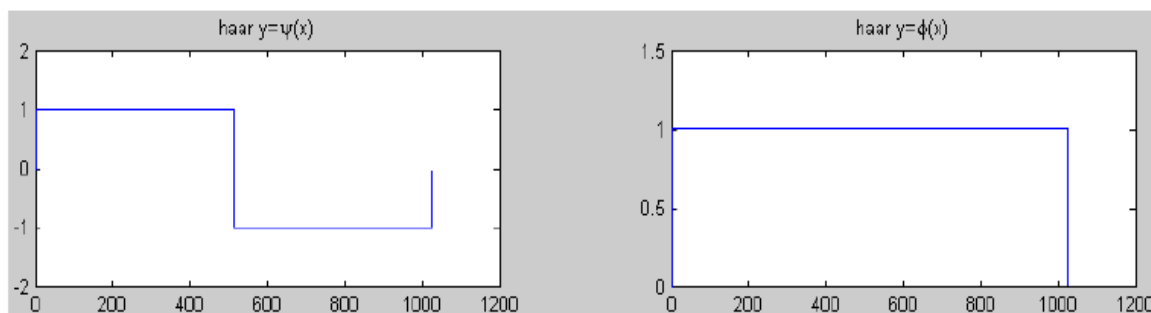


Рисунок 1.3 – Вейвлет Хаара

Дослідимо спектри чотирьох похідних представників сімейства вейвлетів Добеши, зображених на рисунку 1.4.

На спектрограмах є помітний пік в області низьких частот, а от амплітуда інших складових спектру практично рівна нулю. Отже, вейвлети Добеши є вузькосмуговими фільтрами і можуть виділяти з сигналу складові, частота яких відповідає їх піку (цю частоту називають центральною частотою вейвлету, на

риснку показано ті ж чотири вейвлети і синусоїди, що відповідають їх центральним частотам) [11].

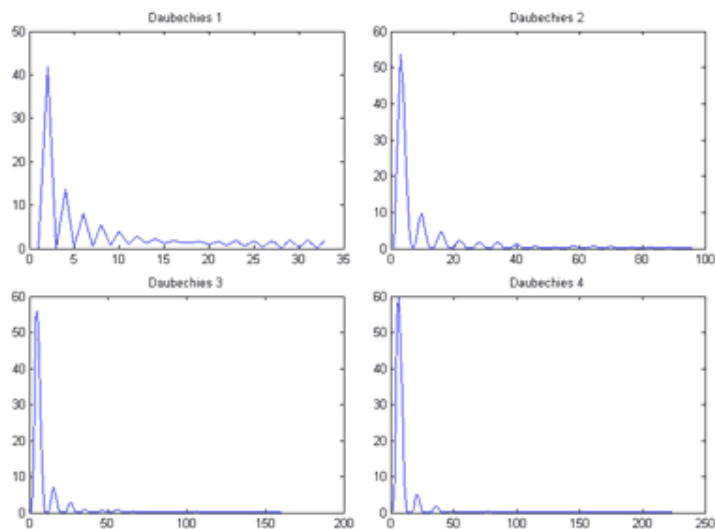


Рисунок 1.4 – Вейвлети Добеши

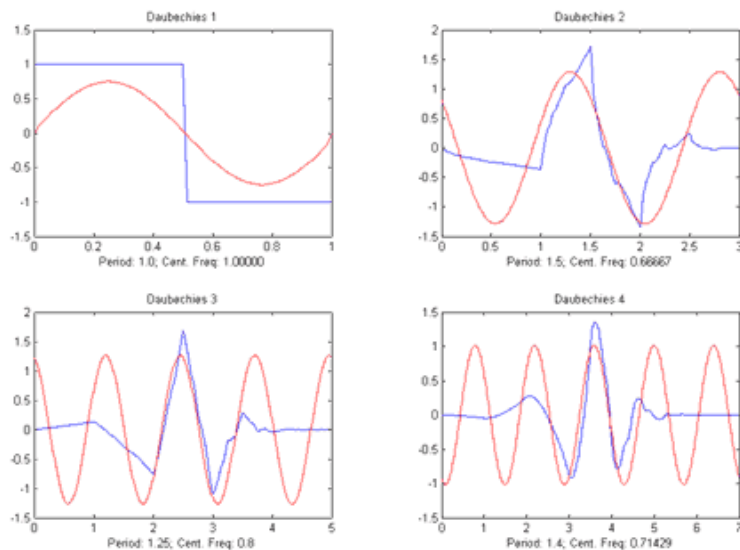


Рисунок 1.5 – Застосування вейвлетів Добеши

Відзначимо, що змінюючи масштаб вейвлету можна пропорційно змінювати його центральну частоту:

$$F_a = \frac{\Delta \cdot F_c}{a} \quad (1.10)$$

Таким чином, можна перекрити весь частотний діапазон досліджуваного сигналу змінюючи масштаб вейвлет-перетворення. Одночасно, за рахунок точної масштабованості функції вейвлет-перетворення у часовій області можна отримати спектральну інформацію практично у кожній точці сигналу.

### 1.2.3 Фільтрація сигналів

Смуговим називають фільтр, смуга пропускання якого обмежується відрізком частот від  $f_{c1}$  до  $f_{c2}$ .

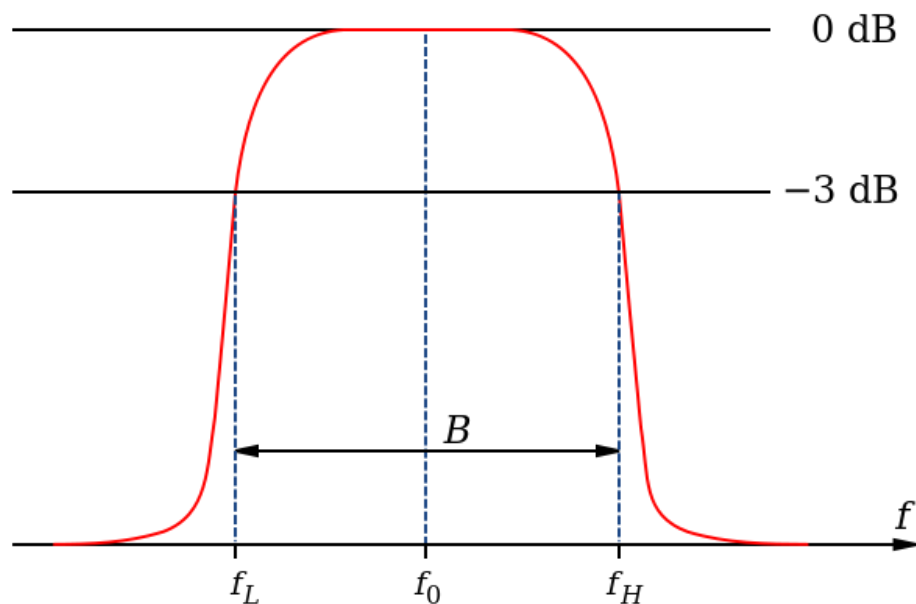


Рисунок 1.5 – Пропускна здатність смугового фільтра

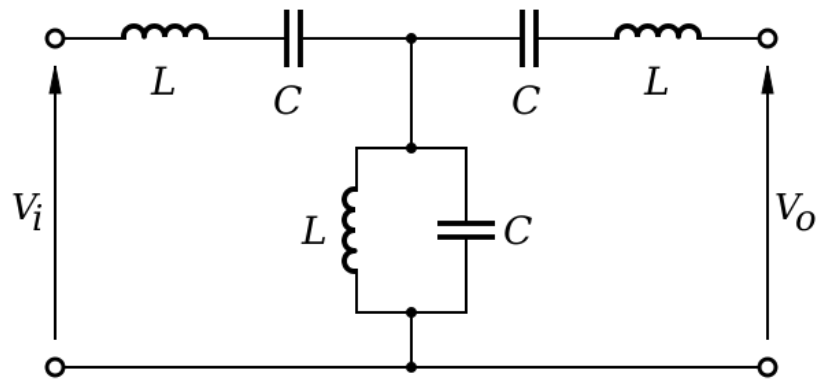


Рисунок 1.6 – Типова схема смугового фільтра

Передавальні функції смугових фільтрів можуть бути знайдені, якщо застосувати до передавальним функціям фільтрів-прототипів нижніх частот перетворення частоти. Відповідна перетворююча функція в області нижніх частот повинна вести себе так само, як і комплексна змінна  $p$  фільтра верхніх частот, тобто  $p \neq 0$ , якщо  $p = 0$ , а в області верхніх частот - як комплексна змінна фільтра нижніх частот, тобто якщо  $p \neq 0$ , то  $p = 0$ . В межах же смуги пропускання перетвореного фільтра комплексна змінна  $p$  повинна змінюватися так само, як і змінна в смузі пропускання фільтра-прототипу нижніх частот [12].

Найпростіша перетворююча функція, яка задовольняє цим вимогам, являє собою суму комплексних змінних фільтрів нижніх і верхніх частот.

$$p = p' + (w' 0')^2 / p. \quad (1.10)$$

Якщо фільтром-прототипом нижніх частот служить LC - фільтр, то перетворення переводить кожен індуктивність фільтра нижніх частот в реактивний двухполюсник, тобто в послідовний коливальний контур без втрат з тією ж індуктивністю, ємністю і резонансною частотою, а кожен ємність C1 в двухполюсник з провідністю Y1, тобто в паралельний коливальний контур без втрат з тією ж ємністю C1, індуктивністю і резонансною частотою. Таким чином, вихідна ланцюг перетвориться в інший, також фізично реалізований LC –ланцюг [13].

#### 1.2.4 Згортання сигналів

Згортання сигналу може бути представлене формулою:

$$s(t) = x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau) \cdot h(t - \tau) d\tau . \quad (1.11)$$

Згортка дозволяє розрахувати сигнал  $s(t)$  на виході лінійного фільтра з імпульсною характеристикою  $h(t)$ , при вхідному сигналі  $x(t)$ .

В дискретному випадку розрізняють два види згорток: лінійну (або аперіодичну) і циклічну. Циклічну згортку ще часто називають круговий або періодичною [14].

Розглянемо лінійну згортку. Нехай є два дискретних сигнали  $a(n)$ ,  $n=0\dots N-1$  та  $b(n)$ ,  $n=0\dots M-1$ . У загальному випадку довжини цих сигналів можуть відрізнятись. Лінійної згорткою сигналів  $a$  і  $b$  буде сигнал виду:

$$s(n) = a * b = \sum_{m=0}^n a(m) \cdot b(n - m),$$

$$n = 0 \dots N + M - 2. \quad (1.12)$$

Для обчислення лінійної згортки сигнали  $a(n)$  і  $b(n)$  зрушують відносно один одного почленно перемножують і складають. При цьому передбачається, що  $a(n)=0$  при  $n<0$  та  $n>N$ , а також  $b(n)=0$  при  $n<0$  та  $n>M$ .

У разі циклічної згортки передбачається, що дискретні сигнали  $a(n)$  і  $b(n)$  - періодичні з однаковим періодом  $N$  відліків. Тоді круговим згортанням називається сигнал виду:

$$s(n) = \sum_{m=0}^{N-1} a(m) \cdot b(n - m). \quad (1.13)$$

### 1.3 Оброблення сигналів в часовому просторі



При вирішенні багатьох задач оброблення мовних сигналів інтерес представляють їх часові характеристики. Оскільки мова - це нестационарний процес, то її прийнято аналізувати на коротких інтервалах (10-30 мс), де спектрально-кореляційні характеристики залишаються приблизно сталими [15]. Властивості мовного сигналу змінюються у часі, наприклад, характер збудження на невокалізованих і вокалізованих відрізаках, пікова амплітуда сигналу, частота основного тону на вокалізованих сегментах сигналу. Методи оброблення сигналу у часовій області повинні забезпечувати достовірний опис таких вагомих характеристик сигналу, як потужність, основний тон, характер збудження, навіть такі параметри голосового тракту, як форматні частоти.

Енергія є одним із важливих параметрів мовного сигналу і може слугувати мірою відмінності вокалізованих від невокалізованих сегментів мови. Енергія невокалізованих сегментів мови набагато менша, ніж енергія вокалізованих сегментів. Інший спосіб виявлення вокалізованих і невокалізованих сегментів мови засновано на вимірюванні середньої кількості переходів мовного сигналу через нуль. Ця характеристика є грубою оцінкою частотного вмісту мовного сигналу. Відомо, що енергія фрикативних звуків в основному зосереджена на частотах вище 3 кГц, а енергія вокалізованих звуків концентрується у частотному діапазоні нижче 3 кГц. Отже, якщо середня кількість переходів через нуль велика, то це свідчить про невокалізований характер сегменту мови, і навпаки [16].

Важливу задачею аналізу мовних сигналів у часовій області є оцінювання періоду основного тону. Період основного тону можна визначити як часовий інтервал між помітними піками вокалізованого сегменту мовного сигналу. Однак, основна проблема тут полягає у тім, що навіть на коротких інтервалах часу мовний сигнал не має сталої періодичної структури. Іншим способом оцінювання періоду основного тону є обчислення функції короткочасної авто кореляційної функції.

В основі більшості методів оброблення мови лежить припущення, що властивості мовного сигналу із часом повільно змінюються [17]. Це припущення лежить в основі методів короткочасового аналізу, при якому сегменти мовного

сигналу виділяються і оброблюються так, як неби вони є короткими сегментами окремих звуків із відмінними властивостями. Процедуру повторюють так часто, як потрібно. Сегменти, які іноді називають інтервалами (фреймами) аналізу, зазвичай, перекриваються. Результатом оброблення на кожному фреймі є єильість або послідовність чисел. Тобто, подібне оброблення призводить до нової, залежної від часу послідовності, яка може слугувати характеристикою мовного сигналу. Більшість методів короткочасного аналізу можга описати виразом:

$$Q_n = \sum_{m=-\infty}^{+\infty} T[x(m)]w(n-m). \quad (1.14)$$

Після обмеження частотного діапазону в наслідок лінійної фільтрації, мовний сигнал піддається лінійному чи нелінійному перетворенню  $T^*$ , яке може залежати від деякого керуючого параметру або їх сукупності. Згодом результуюча послідовність множиться на послідовність значень часового вікна і вагову функцію, розташовану у часі відповідно до індексу  $n$ . Результати множення потім підсумовуються для всіх ненульовим значень. Зазвичай послідовність значень часового вікна скінченна. Значення  $Q$  з виразу (1.14) являє собою, таким чином, зважене середнє значення послідовності  $T$ .

#### 1.4 Постановка задачі дослідження

Задачу автентифікації суб'єкта за голосом можна розкласти на чотири основні частини: одержання мовних даних, попереднє оброблення, виявлення інформативних ознак, автентифікація суб'єкта. Автентифікувати людину за допомогою аналізу унікальних характеристик її голоса, обумовлених її анатомічними властивостями і набутими звичками. Для одержання голосових даних суб'єкт-користувач повинен промовити парольну фразу принаймні тричі. На етапі попереднього оброблення потрібно здійснити wavelet-перетворення сигналу. Для отримання інформативних ознак для автентифікації потрібно провести спектральний аналіз сигналу, розділивши його на шум та інформативну складову.

Інформативна складова мовного сигналу фільтрується смуговим фільтром у діапазоні 100-1000 Гц. Після цього здійснюється часова нормалізація сигналу відповідно до еталону з бази даних. Для створення еталону суб'єкт-користувач має промовити парольну фразу щонайменше три рази, щоб розрахувати середню тривалість звучання сигналу. Відповідно до отриманого еталону проводитиметься стиснення або розтягування сигналу у часі при автентифікації мовця. Інформативною ознакою для автентифікації виступить спектральна характеристика - енергія сигналу, яку ми розрахуємо з інформативної складової. Автентифікація проводиться за допомогою нейромережі.

До запису мовного сигналу висуваються вимоги, узагальнені у таблиці 1.1.

Таблиця 1.1 – Вимоги до запису мовного сигналу

Назва характеристики	Значення
Частота дискретизації	8000 Гц
Глибина біт	16 біт
Кількість каналів	моно
Кількість записів	не менше 3

До нейромережі висуваються вимоги, наведені в таблиці 1.2.

Таблиця 1.2 – Вимоги до нейромережі

Назва характеристики	Значення
Тип нейромережі	Багатошаровий перцептрон – глибока нейромережа
Кількість нейронів вхідного шару	100
Кількість нейронів вихідного шару	Дорівнює кільк. мовців

Продовження таблиці 1.2

Кількість прихованих шарів	2
Кількість нейронів першого прихованого шару	50
Кількість нейронів другого прихованого шару	50

Тестування системи проводитимемо за допомогою бази даних із записами мовних сигналів 20 мовців, які будуть автентифіковані до та після нормалізації парольних сигналів. Для визначення доцільності застосування нормалізації обрахуємо та порівняємо точність автентифікації.

## 2 РОЗРОБКА СИСТЕМИ АВТЕНТИФІКАЦІЇ СУБ'ЄКТІВ ЗА ГОЛОСОМ ІЗ ВІЗУАЛІЗАЦІЄЮ МОВНИХ СИГНАЛІВ

### 2.1 Аналіз основних функцій системи

Вхідними даними для роботи системи є оцифрований запис мовного сигналу (формат wav, 16 кГц, моно, 8 біт). У системі необхідно реалізувати такі функції:

- зчитування вхідні дані – мовний сигнал з файлу;
- вейвлет-перетворення вхідних даних;
- фільтрація вхідних даних;
- згортка вхідних даних.
- візуалізація результатів роботи.

### 2.2 Розробка структурної схеми системи

В даній системі буде реалізовано алгоритм швидкого wavelet-перетворення, що передбавчає фільтрацію вихідного сигналу НЧ- і ВЧ-фільтрами, та подальшого спектрального аналізу відфільтрованого сигналу для виділення інформативних для автентифікації відрізків та шуму. Результати цих перетворень будуть представлені у формі dat-файлу та у часовій карті. Структурну схему системи наведено на рисунку 2.1.

Таку систему планується розробити у програмному середовищі MATLAB 2017. Для розробки потрібно застосувати такі набори інструментів, як Signal Processing Toolbox (для оброблення цифрових сигналів), Filter design toolbox (для фільтрації сигналів), Wavelet toolbox (для вейвлет-перетворення сигналів).

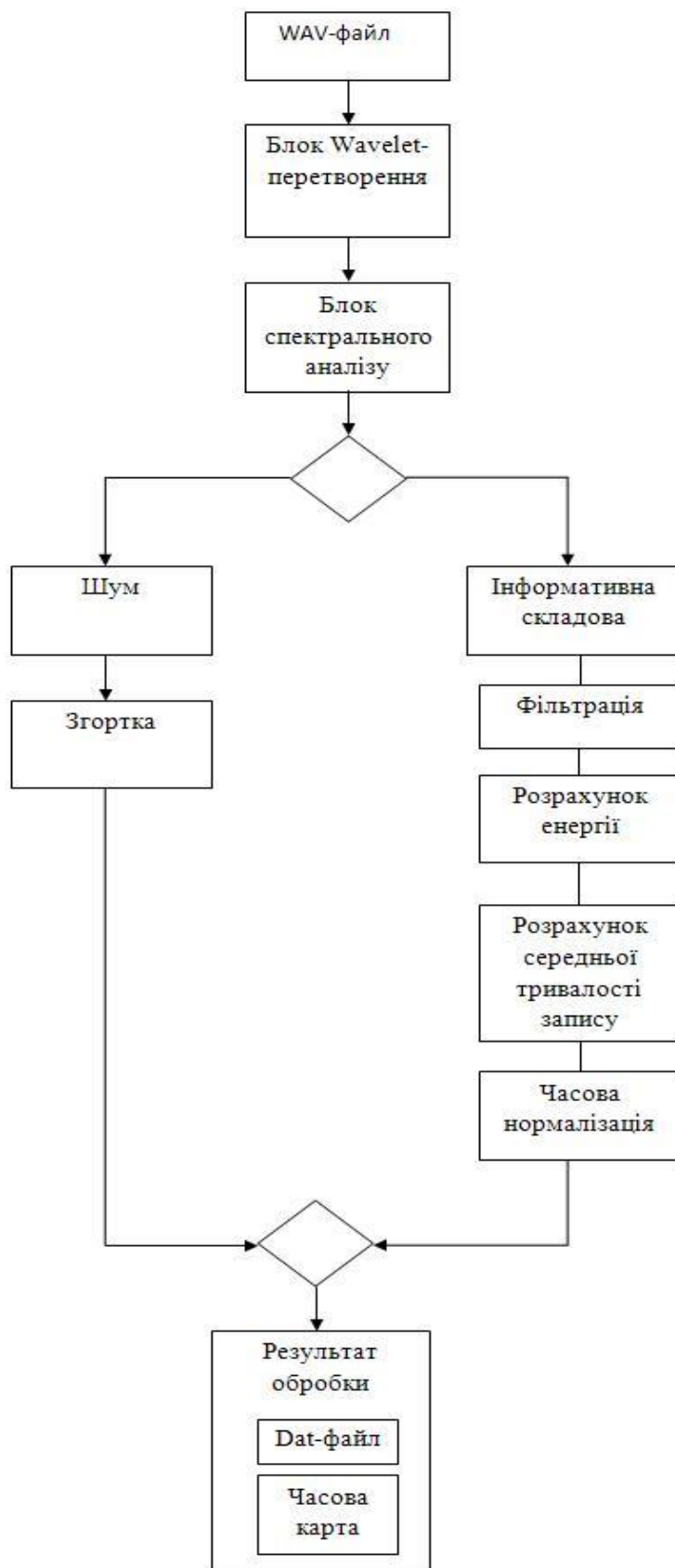


Рисунок 2.1 - Структурна схема системи

## 2.3 Проектування системи з використанням UML-діаграм

Представимо проєкт системи за допомогою чотирьох типів UML-діаграм: діаграма варіантів використання, діаграма діяльності, діаграма послідовності та діаграма розгортання.

Діаграма варіантів використання - це вихідний концептуальний опис або концептуальна модель створеної системи на етапі розроблення і впровадження.

Призначення цієї діаграми полягає в тому, що проєктована програмна система представляється у вигляді так званих варіантів використання, з якими взаємодіють зовнішні сутності, названі акторами. При цьому дійовою особою або актором називається будь-який суб'єкт, об'єкт або система, що взаємодіють із модельованою системою ззовні. Це може бути технічний пристрій, програма, людина або будь-яка інша система, ступінь впливу на модельовану систему якої визначає розробник. Варіант використання служить для опису сервісів, які проєктована система надає акторові. Іншими словами, кожен варіант використання визначає множину дій, який чиняться системою у процесі діалогу із актором. При цьому нічого не говориться про те, яким чином буде реалізовано взаємодію акторів із системою і як власне підтримуватимуться варіанти використання [18].

На рисунку 2.2 представлено діаграму варіантів використання, де показано взаємодію між всіма елементами проєктованої системи.

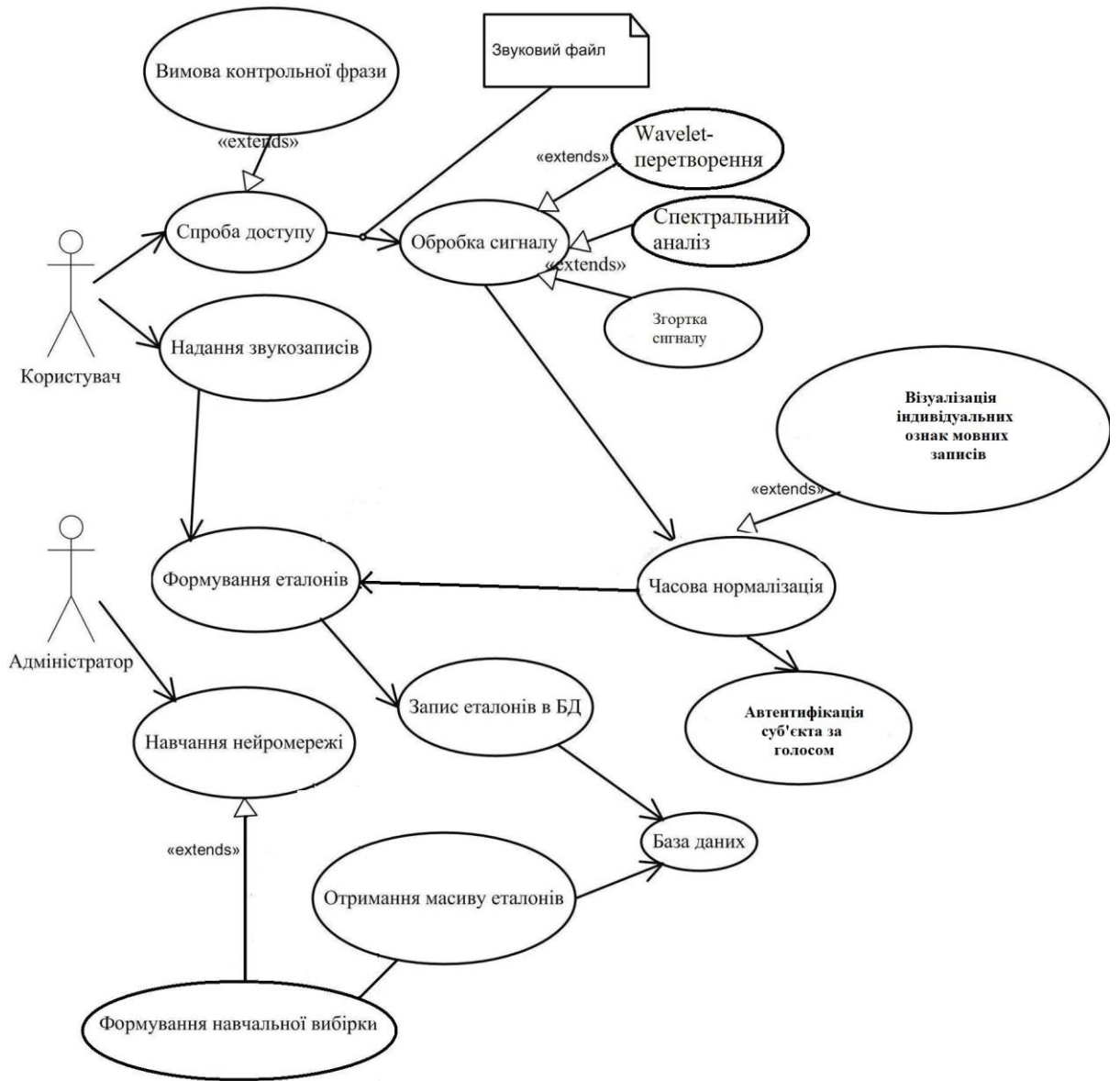


Рисунок 2.2 – UML-діаграма варіантів використання

На рисунку 2.3 представлено UML-діаграма послідовності, яка графічно відображає послідовність реалізації процесів у модельованій системі під час її функціонування.



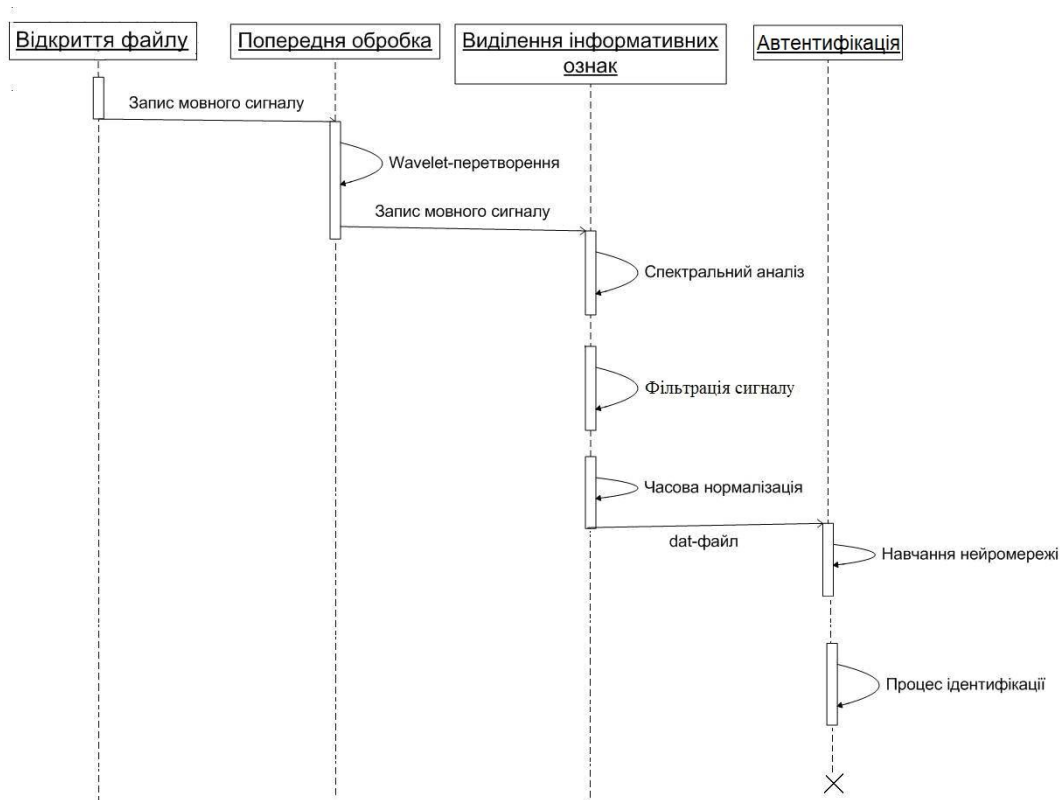


Рисунок 2.3 - UML-діаграма послідовності

Кожен стан на діаграмі діяльності відповідає виконанню певної операції, а перехід у наступний стан відбувається тільки після завершення виконання поточної операції. Діаграма діяльності представляється у формі графу діяльності, вершинами якого є стани дій, а дугами - переходи від одного стану дії до іншого [19]. На рисунку 2.4 представлено діаграму діяльності модельованої системи.

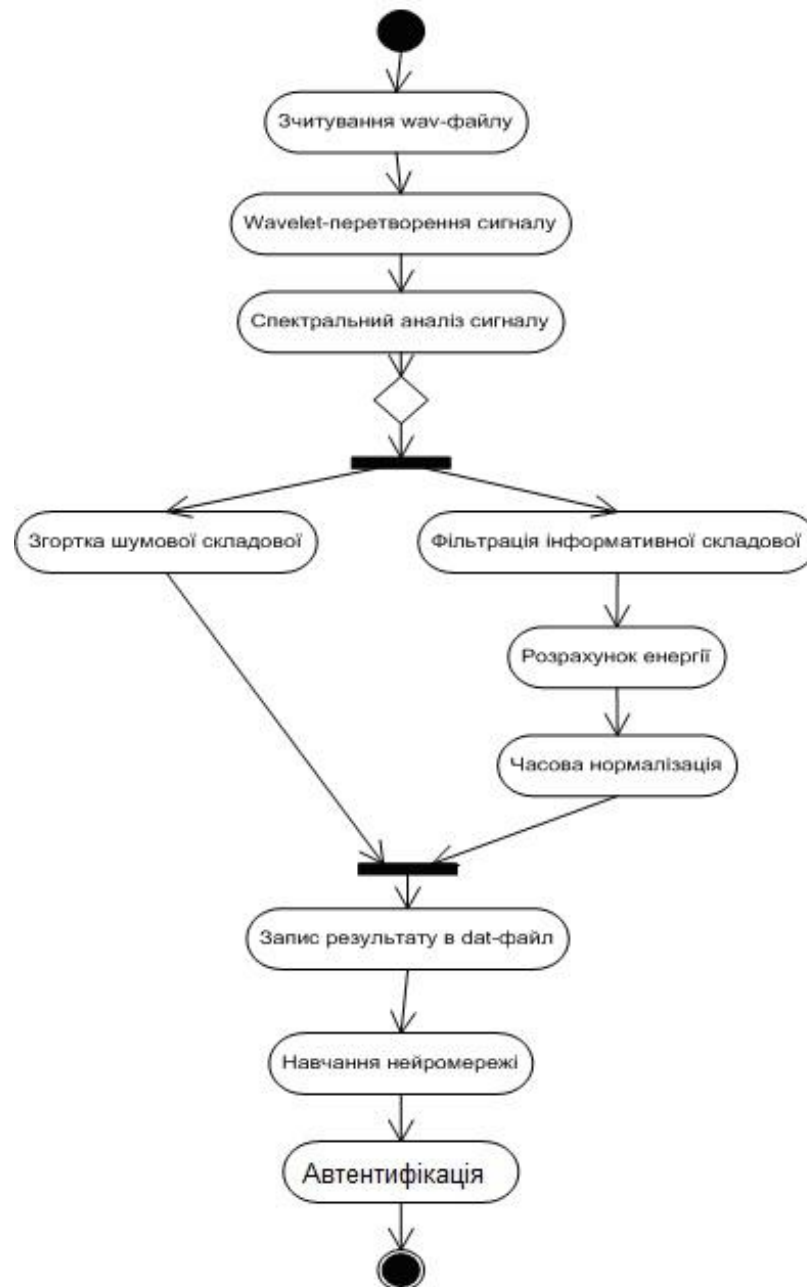


Рисунок 2.4 – UML-діаграма діяльності блоку часової нормалізації системи автентифікації суб'єктів за голосом із візуалізацією мовних сигналів

UML-діаграма розгортання - це діаграма, на якій відображаються обчислювальні вузли, задіювані під час роботи програми, її компоненти та об'єкти, що виконуються на цих вузлах. Компоненти відповідають представленню класів – робочих екземплярів одиниць коду [19]. UML-діаграма розгортання модельованої системи представлено на рисунку 2.5.

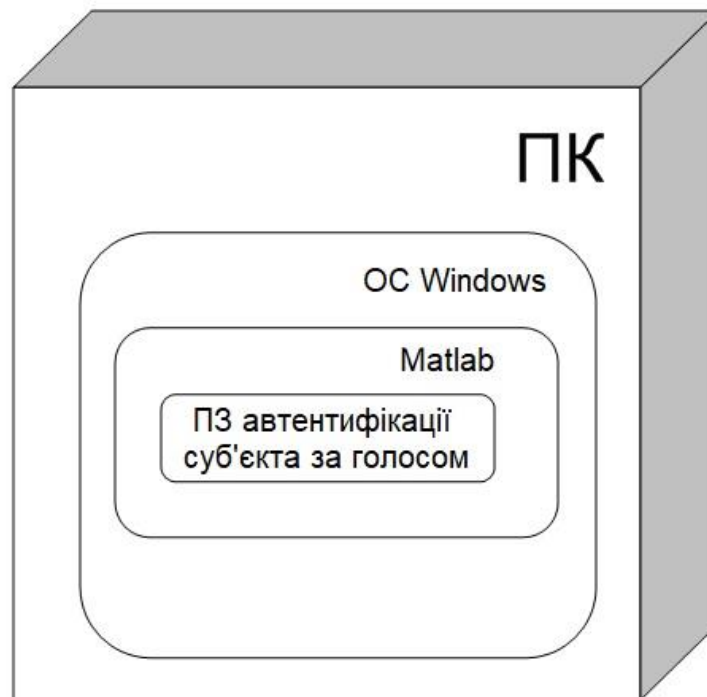


Рисунок 2.5 - UML-діаграма розгортання системи автентифікації суб'єктів за голосом із візуалізацією мовних сигналів

#### 2.4 Розробка алгоритму нелінійної часової нормалізації мовного сигналу

Нехай відомо дві реалізації набору:  $X^{(0)}, \dots, X^{(i)}, \dots, X^{(m)}$  і  $Y^{(0)}, \dots, Y^{(j)}, \dots, Y^{(n)}$ .

Перша реалізація набору вважається еталонною, друга - новою.

Перш ніж порівнювати їх між собою необхідно зліїснити їх часову нормалізацію, тобто привести реалізації наборів до однакової довжини. Лінійний стиснення або розтягнення однієї реалізації набору до величини іншої не вирішує питання внаслідок важливої особливості мовного сигналу - нерівномірності його перебігу у часі. Ця властивість мовлення виражається у складно контрольованій залежності часу утворення і звучання елементів мовного сигналу від контексту, темпу, діалектну та індивідуальних особливостей мовця. Тому порівняння має спиратися на нелінійну короткочасну нормалізацію [20].

Для цього обчислюється деформуюча функція, застосування якої мінімізує розбіжність між еталонною і новою реалізаціями наборів. Точніше обчислюються дві функції:

$$\omega_X : \{1, \dots, l\} \rightarrow \{1, \dots, m\}, \quad (2.1)$$

$$\omega_Y : \{1, \dots, l\} \rightarrow \{1, \dots, n\}. \quad (2.2)$$

Такі, щоб:

$$\begin{aligned} \omega_X(1)=1, \omega_Y(1)=1, \omega_X(l)=m, \omega_Y(l)=n, \\ \omega_X(i+1)=\omega_X(i) \text{ або } \omega_X(i)+1, \forall i=1, \dots, m-1, \\ \omega_Y(j+1)=\omega_Y(j) \text{ або } \omega_Y(j)+1, \forall j=1, \dots, n-1. \end{aligned} \quad (2.3)$$

Сегментуюча функція повинна характеризувати сумарну зміну використовуваних у ній параметрів мовного сигналу і залежить від двох фреймів: поточного і попереднього. В якості параметрів мовного сигналу будемо використовувати розподіл енергії сигналу по частотним проміжкам. Опишемо процедуру знаходження контуру сегментів функції  $SX(1), \dots, SX(i), \dots, SX(m)$  для еталонної реалізації набору [21].

Для кожного фрейму  $X(i)$  знаходимо розподіл енергії сигналу за частотними групами:  $p_1(i), \dots, p_{20}(i)$ ;  $i = 0, 1, \dots, l$  обчислюємо модулі кінцевих різниць:

$$\Delta_k(i) = |p_k(i) - p_k(i-1)|; i = 1, \dots, m; k = 1, \dots, 20. \quad (2.4)$$

Згодом обчислюються середні різниці:

$$K = 1, \dots, 20. \quad (2.5)$$

Далі обчислюються середньозважені різниці:

$$\bar{\Delta}_k = \frac{1}{m} \sum_{i=1}^m \Delta_k^{(i)} ; i=1, \dots, m; k=1, \dots, 20. \quad (2.6)$$

І, врешті, контур сегменту функції  $S_X$ :

$$\delta_k^{(i)} = \frac{\Delta_k^{(i)}}{\bar{\Delta}_k} ; i=1, \dots, m. \quad (2.7)$$

Аналогічно обчислюється контур сегментуючої функції  $S_Y(1), \dots, S_Y(j), \dots, S_Y(n)$  для нової реалізації набору.

Обчислення деформуючих функцій  $\omega_X$ ,  $\omega_Y$  здійснюється за методом динамічного програмування і дозволяє зробити внутрішнє нелінійне вирівнювання реалізацій наборів із часом [22].

Спочатку будується матриця відстаней  $R = \{ \rho_{i,j} \}$  розміром  $(m \times n)$ . На її основі потім розраховується матриця  $D = \{ d_{i,j} \}$  того ж розміру:

- $d_{m,n} = \rho_{m,n}$ ;
- $d_{i,n} = \rho_{i,n} + d_{i+1,n}$ ,  $i = m-1, \dots, 1$ ;
- $d_{m,j} = \rho_{m,j} + d_{m,j+1}$ ,  $j = n-1, \dots, 1$ ;
- $d_{i,j} = \rho_{i,j} + \min\{ d_{i+1,j+1}, d_{i+1,j}, d_{i,j+1} \}$ ,  $i = m-1, \dots, 1$ ;  $j = n-1, \dots, 1$ .

У свою чергу, матриця  $D$  використовується для обчислення функцій  $\omega_X$ ,  $\omega_Y$ . Спочатку присвоюються значення  $\omega_X(1)=1$ ,  $\omega_Y(1)=1$ . Далі на  $k$ -ому кроці знаходять значення  $\omega_X(k+1)$  і  $\omega_Y(k+1)$ . Можливі чотири випадки:

1. Якщо  $\omega_X(k)=m$  і  $\omega_Y(k)=n$ , то деформуючі функції знайдено;
2. Якщо  $\omega_X(k)=m$ , а  $\omega_Y(k)<n$ , то присвоюємо:  $\omega_X(k+1)=m$ ,  $\omega_Y(k+1)=\omega_Y(k)+1$ ;

3. Якщо  $\omega_X(k) < m$ , але  $\omega_Y(k) = n$ , то присвоюємо:  $\omega_X(k+1) = \omega_X(k) + 1$ ,  $\omega_Y(k+1) = n$ ;
4. Якщо  $\omega_X(k) < m$  та  $\omega_Y(k) < n$ , то порівнюємо  $d_{i_1, j_1}$ ,  $d_{i_2, j_2}$ ,  $d_{i_3, j_3}$  для знаходження серед них мінімального, із віловідними індексами  $i_{\min}$ ,  $j_{\min}$ .

Зважаємо, що  $i_1 = i_2 = \omega_X(k) + 1$ ,  $i_3 = \omega_X(k)$ ,  $j_1 = j_3 = \omega_Y(k) + 1$ ,  $j_2 = \omega_Y(k)$ .

Потім присвоюємо:  $\omega_X(k+1) = i_{\min}$ ,  $\omega_Y(k+1) = j_{\min}$ .

Розрахувавши деформуючі функції  $\omega_X$ ,  $\omega_Y$  можна для будь-якого сегменту еталонної реалізації набору знайти відповідний йому сегмент у новій реалізації. Застосуємо це для поділу нової реалізації набору на звукові одиниці. Звукова оцінювача це перехідний процес від однієї фонемі до іншої, що відображає перехідний процес у апараті артикуляції. На відміну від реалізацій фонемі, реалізація звукової лрицниці значно менше зазнає впливу контексту і відображає взаємозв'язок сусідніх фонем у мовному сигналі. Границями звукових одниць є центри квазістаціонарних сегментів фонем [23].

Таким чином, звукова одиниця складається з другої половини першої фонемі і першої половини другої фонемі.

Еталонна реалізація наблру ділиться на звукові одиниці вручну: визначаються номери  $a_0, \dots, a_L$  кадрів, які є центрами квазістаціонарних сегментів фонем. Потім вибираються точки  $n_l$ ,  $l = 0, \dots, L$  такі, що  $\omega_X(n_l) = a_l$ . Далі за допомогою функції  $\omega_Y$  можна визначити номери  $b_0, \dots, b_L$  фреймів, які є центрами квазістаціонарних сегментів фонем у новій реалізації набору:  $b_l = \omega_Y(n_l)$ ,  $l = 0, \dots, L$ . Наведений алгоритм дозволяє перейти від порівняння реалізацій наборів – послідовностей звуків до порівняння реалізацій звукових одиниць [24].

Алгоритм нелінійного методу часової нормалізації зображено на рисунку 2.7.

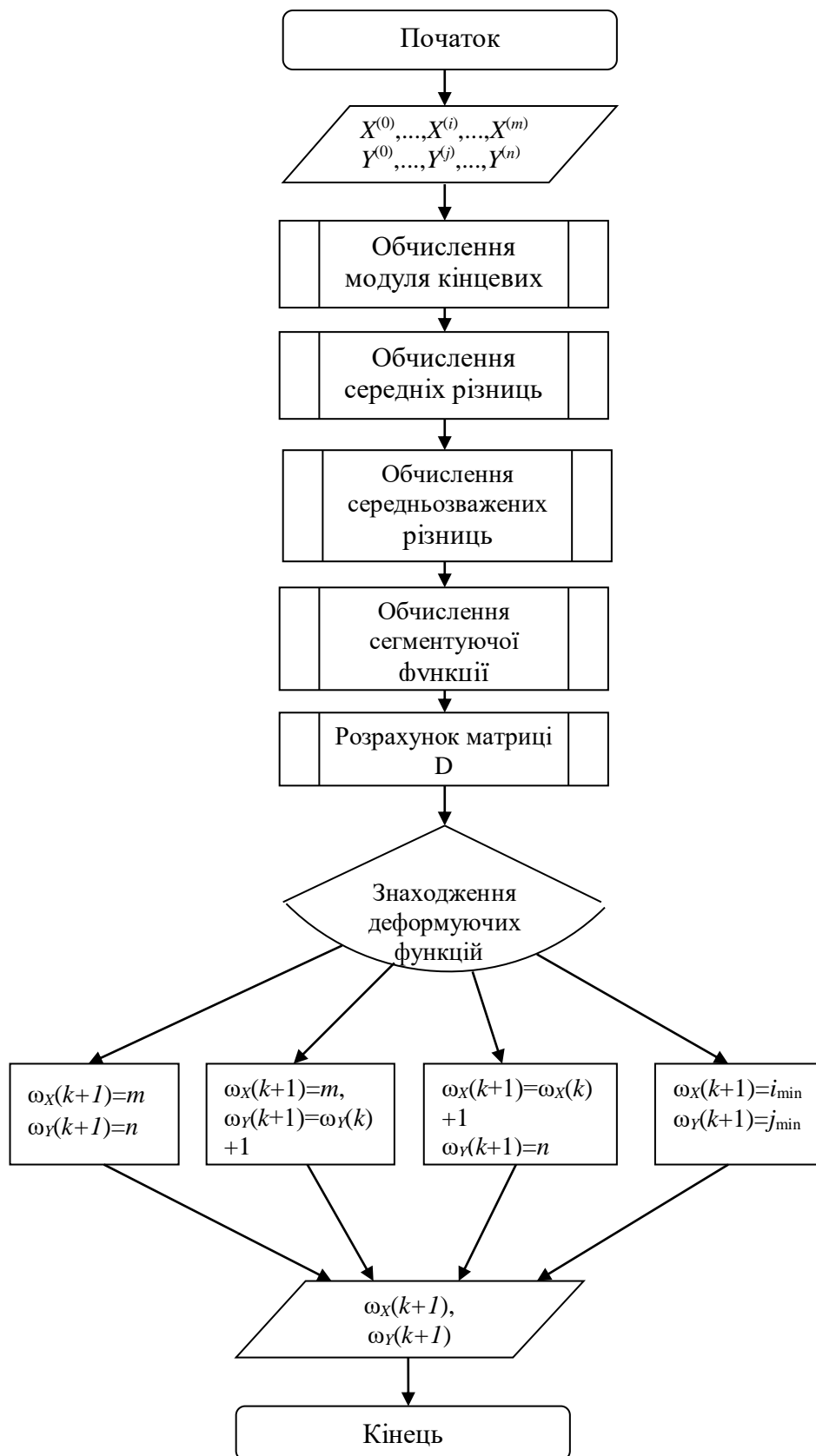


Рисунок 2.7 – Алгоритм нелінійного методу часової нормалізації

## 2.5 Алгоритм фільтрації цифрового сигналу

В загальному випадку терміном цифровим фільтром (ЦФ) називають апаратну або програмну реалізацію математичного алгоритму, на вхід якого подається цифровий сигнал, а на виході отримується інший цифровий сигнал із певним чином модифікованою формою, амплітудною і фазовою характеристикою. Класифікація цифрових фільтрів зазвичай спирається на функціональні ознаки алгоритми цифрової фільтрації, згідно із якими ЦФ узагальнюються у 4 групи:

- фільтри частотної селекції;
- оптимальні (квазіоптимальний);
- адаптивні;
- евристичні.

Найбільше вивченими і вживаними на практиці є ЦФ частотної селекції.

У одновимірній дискретній лінійній системі зв'язок між входом і виходом задається лінійним оператором перетворення TL:

$$y(k\Delta t) = TL\{x(k\Delta t)\}. \quad (2.8)$$

Цей вираз є фактично коротким записом лінійного різницевого рівняння:

$$\sum_{m=0}^M a_m y(k\Delta t - m\Delta t) = \sum_{n=0}^N b_n x(k\Delta t - n\Delta t), \quad (2.9)$$

де  $k = 0, 1, 2, \dots$  - порядковий номер відліків;

$\Delta t$  - інтервал дискретизації сигналу;

$a_m$  і  $b_n$  - дійсні або комплексні коефіцієнти.

Покладемо  $a_0 = 1$ , чого завжди можна досягти виконавши відповідне нормування рівняння (2.8), і, приймаючи в подальшому  $\Delta t = 1$ , тобто застосовуючи



числову нумерацію цифрових послідовностей значень сигналів, приведемо його лпис до виду:

$$y(k) = \sum_{n=0}^N b_n x(k-n) - \sum_{m=1}^M a_m y(k-m). \quad (2.10)$$

При  $k < n$  і  $m$  здійснення фільтрації можливе тільки при завданні початкових умов для точок  $x(-k)$ ,  $k = 1, 2, \dots, N$ , і  $y(-k)$ ,  $k = 1, 2, \dots, M$ . Як правило, в якості початкових умов використовуються або нульові значення, або здійснюється інтерполяція відліків вхідних сигналів або його тренду за негативними значеннями аргументу.

Оператор, представлений у правій частині рівняння (2.9), названо цифровим фільтром, а виконувана їм операція - цифровою фільтрацією даних (інформації, сигналіу тощо). Якщо хоча б один із коефіцієнтів  $a_m$  або  $b_n$  залежить від змінної  $k$ , то фільтр називають параметричним, тобто зі змінюваними параметрами.

Цифрові фільтри розбтвають на два основні типи: фільтри із кінцевою імпульсною характеристикою (КІХ-фільтри) та фільтри із нескінченною імпульсною характеристикою (БІХ-фільтри). До БІХ-фільтрів відносять фільтр Баттерворта, фільтр Чебишева, фільтр Бесселя, фільтр Калмана.

Розрахуємо для смугового фільтру Баттерворта 6-го порядку для полоси пропускання 100-600Hz коефіцієнти  $a_m$  і  $b_n$  за допомогою математичного середвища MATLAB. Функція `butter` дозволяє синтезувати аналогові і дискретні фільтри Баттерворта нижніх частот, верхніх частот, смугові і режекторні. Фільтри Баттерворта характеризуються амплітудно-частотною характеристикою, максимально плоскою у смузі пропускання і монотонно спадаючою по-за її границями. Амплітудно-частотну характеристику синтезованого фільтра представлено на рисунку 2.8.

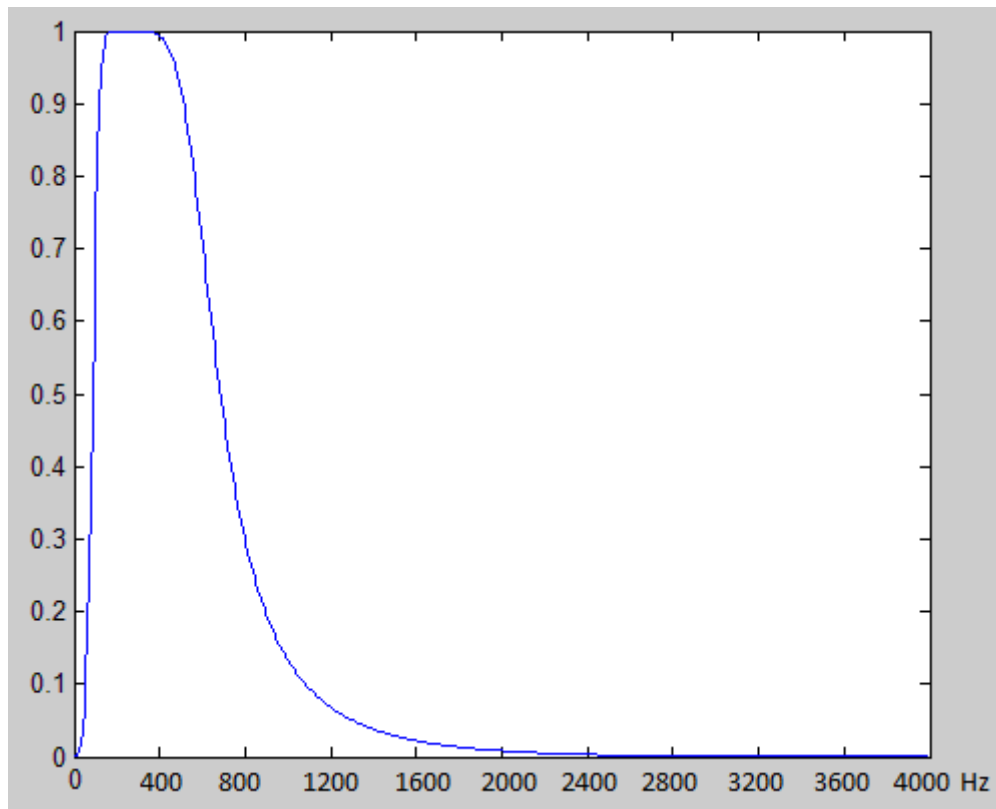


Рисунок 2.8 – Амплітудно-частотна характеристика синтезованого смугового фільтру Баттерворта

## 2.6 Алгоритм дискретного вейвлет-перетворення

Для побудови алгоритму дискретного вейвлет-перетворення використаємо деякі лінійні перетворення. По-перше, визначимо суму чисел по модулю  $s$  поклавши, що є певний вектор у якому  $s$  - парне. Із врахуванням цього вейвлет-перетворення мають вигляд:

$$L^s z(k) = \sum_l h_l z((l + 2k) \bmod s), \quad (2.11)$$

$$H^s z(k) = \sum_l g_l z((l + 2k) \bmod s). \quad (2.12)$$

Очевидно, що дані вирази являють собою аналоги високочастотного і низькочастотного фільтрів із урахуванням періодичного доповнення даних за допомогою підсумовування по модулю два. Ясно, що перетворення  $H^s$  реалізує поділ вихідного вектора тривалістю  $s$  на два вектора половинної довжини.

Отже, алгоритм вейвлет-перетворення зводиться до реалізації ітеративної процедури перетворення, що вживається до всіх елементів вектору  $z = (z(0), z(1), \dots, z(2^K - 1))$ .

Результатом такого перетворення є вектори коефіцієнтів апроксимації та деталізації. Інакше кажучи, рекурсивно даний алгоритм описується такими виразами:

$$a(j, k) = L^{2^{j+1}} a(j+1, k) = \sum_l h_l a(j+1, (l+2k) \bmod 2^{j+1}), \quad (2.13)$$

$$d(j, k) = H^{2^{j+1}} d(j+1, k) = \sum_l g_l d(j+1, (l+2k) \bmod 2^{j+1}). \quad (2.14)$$

Відзначимо, що введені позначення для коефіцієнтів розкладу досить схожі із позначенням коефіцієнтів  $a_{j,k}, d_{j,k}$ , тоді як рекурсії (2.13), (2.14) схожі із каскадним алгоритмом. Справа в тім, що алгоритм дискретного вейвлет-перетворення повністю заснований на теорії дискретного математичного аналізу рядів у базисі вейвлет-функцій. Основна відмінність - у статистичних додатках коефіцієнти лише наближено відповідають коефіцієнтам розкладу.

Рекурсії (2.13), (2.14) можуть, зокрема, із успіхом застосовуватися для розрахунку коефіцієнтів апроксимації та деталізації також і для випадків  $k \in \mathbb{Z}$ . Це досягається за рахунок періодичними отриманих доповнених послідовностей:

$$a(j, k) = a(j, k + 2^j), \quad (2.15)$$

$$d(j, k) = d(j, k + 2^j). \quad (2.16)$$

Алгоритм обереного дискретного вейвлет-перетворення представимо виразом:

$$\hat{a}_{j+1,n} = \sum_k h_{n-2k} \hat{a}_{j,k}. \quad (2.17)$$

Алгоритм починається відновленням вектора:

$$\{a(j_0, k) | k = \overline{0.2^{j_0} - 1}\}. \quad (2.18)$$

І триває до його осточного відновлення:

$$\{a(j, s) | s = \overline{0.2^j - 1}\}, \quad (2.19)$$

поки рівність  $j=K-1$  не виконається. В цьому випадку рекурсивний вираз для відновлення даних опишемо виразом:

$$a(j+1, s) = a(j, s + 2^{j+1}). \quad (2.20)$$

## РОЗДІЛ 3

### ТЕСТУВАННЯ І ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ СТВОРЕНОЇ СИСТЕМИ

Головний інтерфейс програми прелставлено вікном із системним меню. Вигляд головного вікна програми показано на рисунку 3.1.

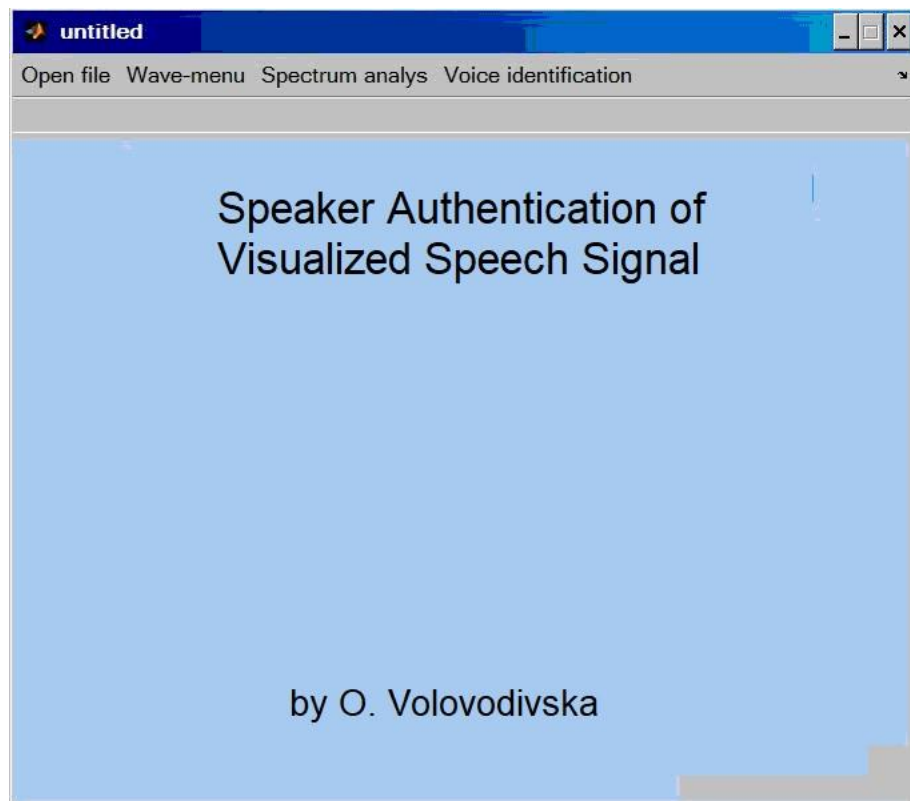


Рисунок 3.1 - Вигляд головного вікна програми

Головне меню містить такі функції:

- Open file (відкриття файлу типу \*.wav);
- Wave-menu (відкриття меню роботи з вейвлетами);
- Spectrum analys (спектральний аналіз);
- Voice identification (відкриття програми ідентифікації мовців).

Підменю Spectrum analys має два підпункти:

- Spectrogram (виводить на екран спектрограму сигналу);

- Normalized signal (виводить на екран нормалізований сигнал).

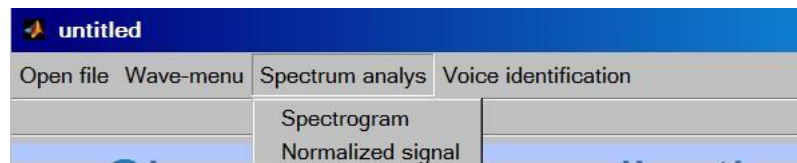


Рисунок 3.2 – Підменю Spectrum analys

В програмному середовищі MATLAB за допомогою команди «wavemenu» запускається меню роботи з вейвлетами, яке зображено на рисунку 3.3

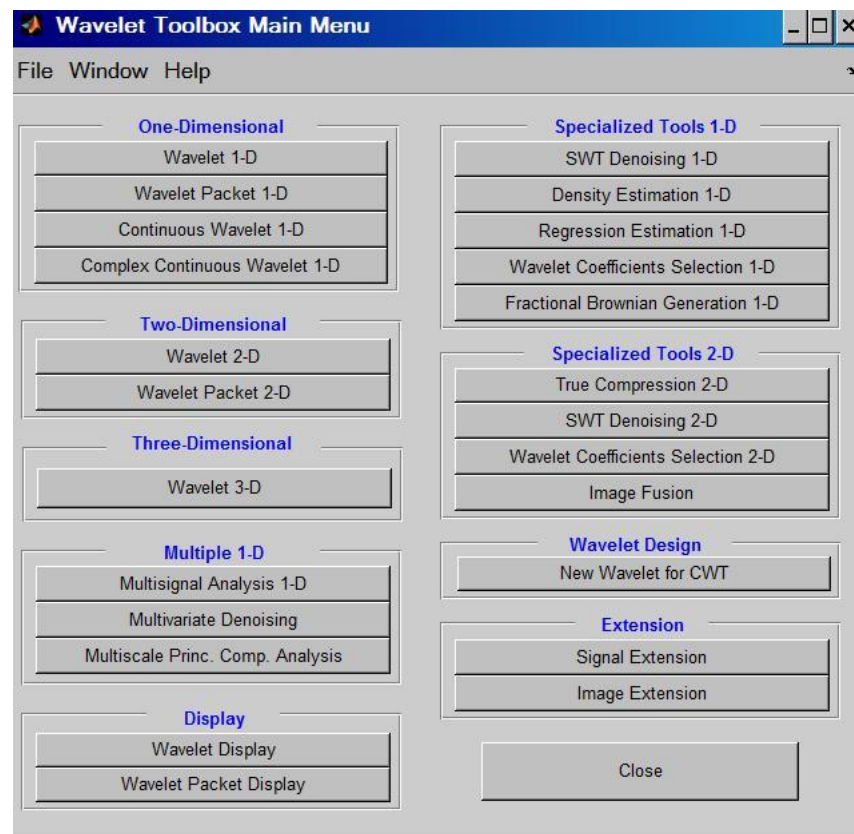


Рисунок 3.3 – Вигляд Wavemenu

Важливим моментом є вибір необхідного вейвлету. Інформацію про вейвлети можна отримати виконавши команду «Wavelet Display». У створеній системі використаємо вейвлет Хаара, який задається командою «Haar». Далі потрібно виконати такі дії:

- розкладання сигналу (обираємо кількість рівнів розкладання  $J$ . Обчислюється wavelet-перетворення сигналу  $s$  до рівня  $J$ );

- вибір порогу для wavelet-коефіцієнтів (для кожного рівня від 1 до  $J$  вибирається поріг і проводиться модифікація коефіцієнтів за обраним правилом «hard» або «soft»).

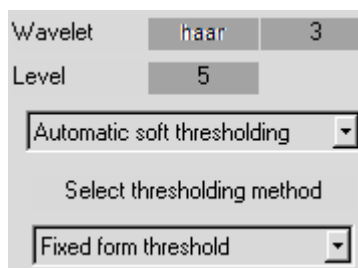


Рисунок 3.4 – Вибір параметрів процедур

- здійснюється wavelet-перетворення вихідного мовного сигналу, після чого запам'ятовуються тільки значущі коефіцієнти, тобто ті, значення яких перевищує задане порогове значення.

Розкладання сигналу за допомогою вейвлету Хаара представлено на рисунку 3.5.

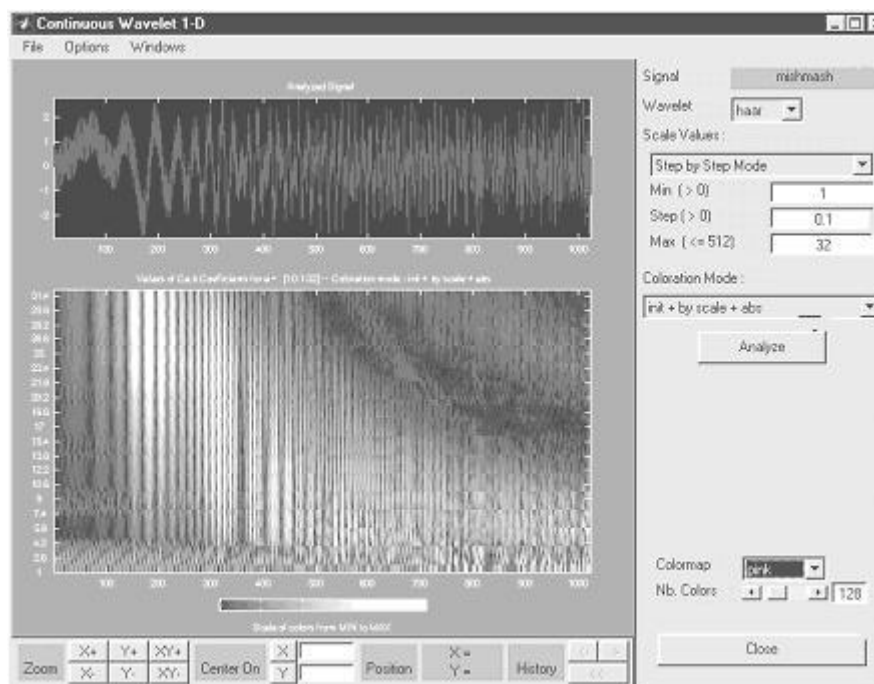


Рисунок 3.5 - Розклад сигналу за допомогою вейвлету Хаара

Спектрограму звукового сигналу представлено на рисунку 3.6.

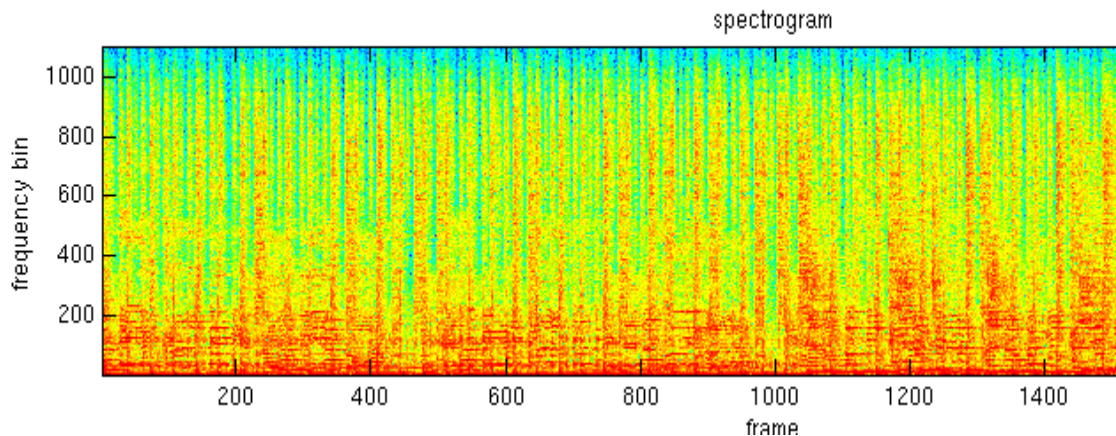


Рисунок 3.6 – Спектрограма звукового сигналу

Спектральний аналіз вживаємо для роздення сигналу на дві складові: шум та інформативну складову. До шумові складової застосуємо операцію згортки. В Signal Processing Toolbox згортка двох сигналів еквівалентна множенню Фур'є-перетворень цих сигналів і реалізуються за допомогою команди  $\text{conv}(x, y) = \text{ifft}(X.*Y)$ .

Інформативну складову спочатку фільтруватимемо. У Filter design toolbox смуговий фільтр можна синтезувати за допомогою команди  $[n, Wn] = \text{buttord}(Wp, Ws, Rp, Rs)$ , де  $Wp$  - границя смуги пропускання,  $Wp$ , або частота зрізання - скаляр або двоелементний вектор, елементи якого являють собою нормовані частоти і належать діапазону від 0 до 1;  $Ws$  - границя смуги затримки, скаляр або двоелементний вектор, елементи якого є нормовані частоти і належать діапазону від 0 до 1;  $Rp$  - пульсації АЧХ у смугі пропускання (у децибелах). Цей параметр задає максимально допустимі втрати сигналу у смугі пропускання;  $Rs$  - згасання АЧХ у смугі затримки (у децибелах). Ця характеристика задає мінімально необхідне пригнічення сигналу в смугі затримки.

Якщо необхідно синтезувати смуговий фільтр із неоднаковим пригніченням сигналу у двох смугах затримки, потрібно синтезувати два фільтри - ФНЧ і ФВЧ із необхідними параметрами, а потім з'єднати їх послідовно.



Після фільтрації розраховуємо енергію сигналу, середню тривалість його звучання та здійснюємо його часову нормалізацію. Приклади реалізації часової нормалізації зображено на рисунках 3.7-3.10.

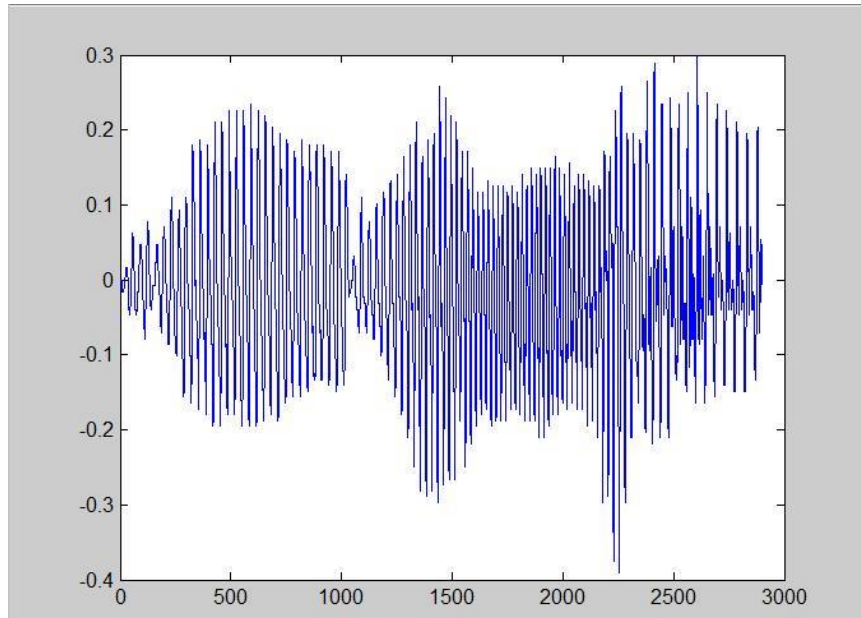


Рисунок 3.7 – Мовний сигнал до нормалізації

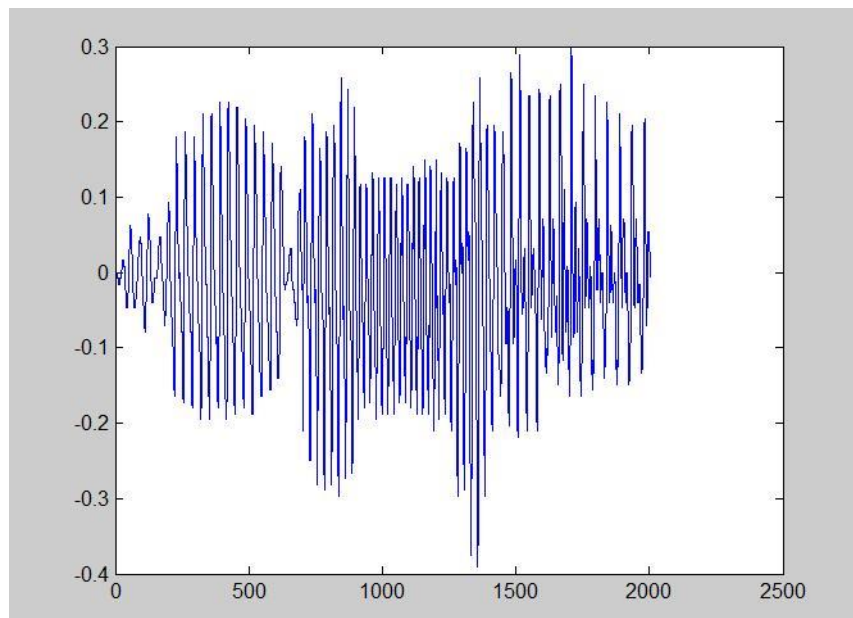


Рисунок 3.8 – Мовний сигнал після нормалізації (стиснення)

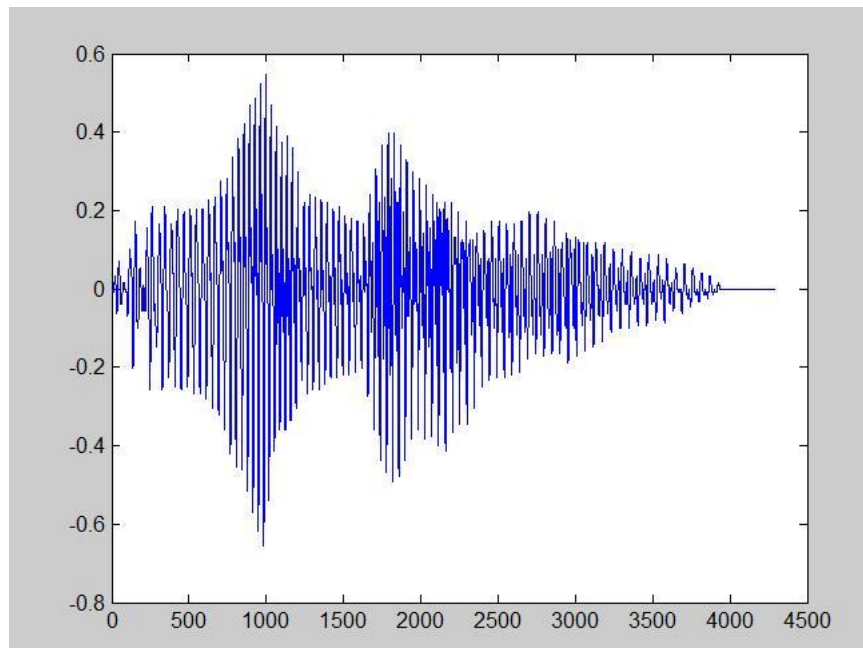


Рисунок 3.9 - Мовний сигнал до нормалізації

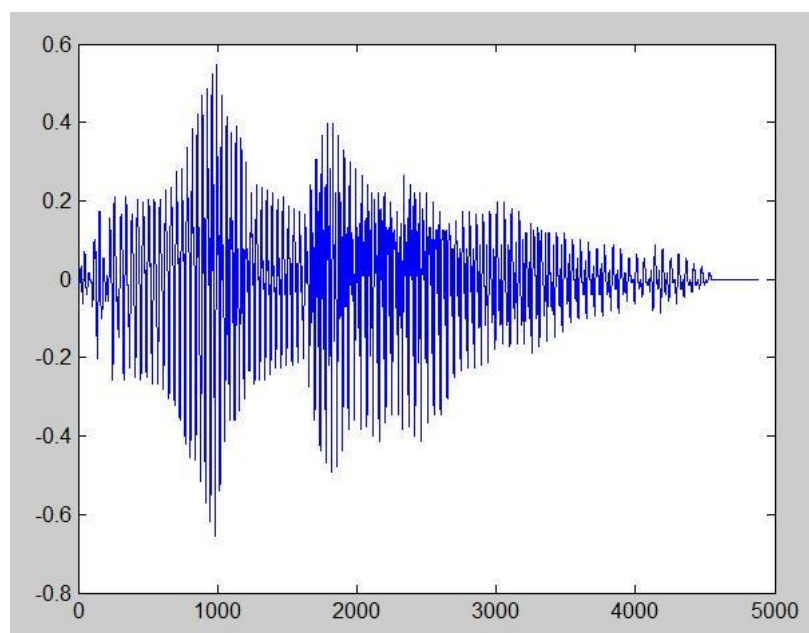


Рисунок 3.10 – Мовний сигнал після нормалізації (розтягування)

Проведемо дослідження доцільності застосування нормалізації тривалості звучання мовних сигналів в задачі автентифікації суб'єктів за голосом. Для цього використаємо базу даних із мовними матеріалами 20 мовців. За допомогою програмного забезпечення проведемо їх автентифікацію до та після застосування нормалізації. Порівняємо точність автентифікації в обох випадках. Якщо точність

автентифікації зросла, то застосування нормалізації можна вважати доцільним. Результати дослідження наведено у таблиці 3.1.

Таблиця 3.1 – Результати дослідження доцільності застосування нормалізації

Запис мовного сигналу	Ідентифікація	
	До нормалізації	Після нормалізації
speaker1	Вірно	Вірно
speaker2	Вірно	Вірно
speaker3	Не вірно	Не вірно
speaker4	Вірно	Вірно
speaker5	Вірно	Вірно
speaker6	Вірно	Вірно
speaker7	Вірно	Вірно
speaker8	Не вірно	Вірно
speaker9	Вірно	Вірно
speaker10	Вірно	Вірно
speaker11	Вірно	Вірно
speaker12	Вірно	Вірно
speaker13	Вірно	Вірно
speaker14	Не вірно	Вірно
speaker15	Вірно	Вірно
speaker16	Вірно	Вірно
speaker17	Вірно	Вірно
speaker18	Вірно	Вірно
speaker19	Вірно	Вірно
speaker20	Вірно	Вірно
Точність	85%	95%

Як ми бачимо з результатів дослідження, застосування нормалізації тривалості записів мовних сигналів дозволило підвищити точність автентифікації на 10%, тобто використання цієї операції можна вважати доцільним.

## 4 ЕКОНОМІЧНИЙ РОЗДІЛ

### 4.1 Технологічний аудит розробки

Відомо, що сучасні комп'ютерні програми дозволяють реалізовувати найскладніші алгоритми управління. Але програми без технічних засобів мертві і не можуть нічого зробити. Саме у поєднанні програмного забезпечення з апаратним і народжуються високоефективні системи управління.

З метою дослідження ефективності вейвлет-аналізу мовних сигналів для розпізнавання мовців згортальними нейронними мережами нами була розроблена та змодельована система для розпізнавання мовців в системі Matlab.

Проведемо технологічний аудит виконаної нами роботи. Технологічний аудит передбачає оцінювання технічного рівня даної розробки та визначення потенційних можливостей її комерційного використання. Для проведення технологічного аудиту скористаємося експертним методом. Для цього запросимо 3-х експертів Вінницького національного технічного університету кафедри «Комп'ютерні системи управління»: к.т.н., доцента Ковтуна В.В., к.т.н., доцента Юхимчук М.С. та к.т.н., доцента Грищук Т.В.

Запрошені експерти є знаними фахівцями у даній галузі науки і техніки, мають значний досвід роботи, тривалий час займаються вивченням даної проблеми, мають вагомі наукові здобутки у даній галузі тощо.

Оцінювання комерційного потенціалу розробки будемо здійснювати за дванадцятьма критеріями, рекомендованими Державним комітетом України з питань науки, інновацій та інформатики (2010 р.), які наведені в таблиці 4.1 [57].

Щоб визначити загальний рівень комерційного потенціалу нашої розробки, скористаємося даними таблиці 4.2, в якій наведено рекомендації щодо можливих рівнів комерційного потенціалу будь-якої розробки.

Таблиця 4.1 – Критерії оцінювання комерційного потенціалу розробки та їх бальна оцінка

Критерії оцінювання та бали (за 5-ти бальною шкалою)					
Кри-терій	0	1	2	3	4
Технічна здійсненність концепції:					
1	Достовірність концепції не підтверджена	Концепція підтверджена експертними висновками	Концепція підтверджена розрахунками	Концепція перевірена на практиці	Перевірено роботоздатність продукту в реальних умовах
Ринкові переваги (недоліки):					
2	Багато аналогів на малому ринку	Мало аналогів на малому ринку	Кілька аналогів на великому ринку	Один аналог на великому ринку	Продукт не має аналогів на великому ринку
3	Ціна продукту значно вища за ціни аналогів	Ціна продукту дещо вища за ціни аналогів	Ціна продукту приблизно дорівнює цінам аналогів	Ціна продукту дещо нижче за ціни аналогів	Ціна продукту значно нижче за ціни аналогів
4	Технічні та споживчі властивості продукту значно гірші, ніж в аналогів	Технічні та споживчі властивості продукту трохи гірші, ніж в аналогів	Технічні та споживчі властивості продукту на рівні аналогів	Технічні та споживчі властивості продукту трохи кращі, ніж в аналогів	Технічні та споживчі властивості продукту значно кращі, ніж в аналогів
Ринкові перспективи					
5	Експлуатаційні витрати значно вищі, ніж в аналогів	Експлуатаційні витрати дещо вищі, ніж в аналогів	Експлуатаційні витрати на рівні експлуатаційних витрат аналогів	Експлуатаційні витрати трохи нижчі, ніж в аналогів	Експлуатаційні витрати значно нижчі, ніж в аналогів
6	Ринок малий і не має позитивної динаміки	Ринок малий, але має позитивну динаміку	Середній ринок з позитивною динамікою	Великий стабільний ринок	Великий ринок з позитивною динамікою
7	Активна конкуренція великих компаній на ринку	Активна конкуренція	Помірна конкуренція	Незначна Конкуренція	Конкуренція немає

Продовження таблиці 4.1

Кри- терій	0	1	2	3	4
Практична здійсненність					
8	Відсутні фахівці як з технічної, так і з комерційної реалізації ідеї	Необхідно наймати фахівців або витратити значні кошти та час на навчання наявних фахівців	Необхідне незначне навчання фахівців та збільшення їх штату	Необхідне незначне навчання фахівців	Є фахівці з питань як з технічної, так і з комерційної реалізації ідеї
Критерії оцінювання та бали (за 5-ти бальною шкалою)					
Кри- тер.	0	1	2	3	4
9	Потрібні значні фінансові ресурси, які відсутні. Джерела фінансування ідеї відсутні	Потрібні незначні фінансові ресурси. Джерела фінансування відсутні	Потрібні значні фінансові ресурси. Джерела фінансування є	Потрібні незначні фінансові ресурси. Джерела фінансування є	Не потребує додаткового фінансування
10	Необхідна розробка нових матеріалів	Потрібні матеріали, що використовуються у військово-промисловому комплексі	Потрібні дорогі матеріали	Потрібні досяжні та дешеві матеріали	Всі матеріали для реалізації ідеї відомі та давно використовуються у виробництві
11	Термін реалізації ідеї більший за 10 років	Термін реалізації ідеї більший за 5 років. Термін окупності інвестицій більше 10-ти років	Термін реалізації ідеї від 3-х до 5-ти років. Термін окупності інвестицій більше 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій від 3-х до 5-ти років	Термін реалізації ідеї менше 3-х років. Термін окупності інвестицій менше 3-х років

## Продовження таблиці 4.1

Кри- терій	0	1	2	3	4
12	Необхідна розробка регламентних документів та отримання великої кількості дозвільних документів на виробництво та реалізацію продукту	Необхідно отримання великої кількості дозвільних документів на виробництво та реалізацію продукту, що вимагає значних коштів та часу	Процедура отримання дозвільних документів для виробництва та реалізації продукту вимагає незначних коштів та часу	Необхідно тільки повідомлення відповідним органам про виробництво та реалізацію продукту	Відсутні будь-які регламентні обмеження на виробництво та реалізацію продукту

Таблиця 4.2 – Рівні комерційного потенціалу розробки

Середньоарифметична сума балів $\overline{СБ}$ , розрахована на основі висновків експертів	Рівень комерційного потенціалу розробки
0 – 10	Низький
11 – 20	Нижче середнього
21 – 30	Середній
31 – 40	Вище середнього
41 – 48	Високий

Результати оцінювання комерційного потенціалу нашої розробки, зробленого запрошеними експертами, зведемо в таблицю 4.3.

Таблиця 4.3 – Результати оцінювання комерційного потенціалу розробки

Критерії	Прізвище, ініціали експерта		
	Ковтун В.В.	Юхимчук М.С.	Грищук Т.В.
	Бали, виставлені експертами:		
1	2	3	2
2	2	2	2
3	3	4	3
4	3	2	3
5	3	4	4
6	4	3	4
7	2	3	2
8	2	3	2
9	1	2	1
10	4	3	4
11	3	4	4
12	2	3	3
Сума балів	СБ <sub>1</sub> = 31	СБ <sub>2</sub> = 36	СБ <sub>3</sub> = 34
Середньоарифметична сума балів $\overline{СБ}$	$\overline{СБ} = \frac{\sum_{i=1}^3 СБ_i}{3} = \frac{31+36+34}{3} = 34$		

Оскільки середньоарифметична сума балів, що їх виставили експерти, складає 34 бали, то, керуючись даними таблиці 4.2, можна зробити висновок, що розроблена нами система розпізнавання мовців має рівень комерційного потенціалу, який вважається «вище середнього».

Високий технологічний рівень нашої розробки пояснюється тим, що розроблена нами система розпізнавання мовців має значно ширші функціональні можливості та забезпечує вищу якість розпізнавання мовців.



## 4.2 Прогнозування витрат на виконання роботи

Прогнозування витрат на виконання нашої наукової роботи складається з таких етапів [58]:

- 1) розрахунок витрат, які безпосередньо стосуються виконавців даного розділу роботи.
- 2) розрахунок загальних витрат на виконання всієї роботи.
- 3) прогнозування загальних витрат на виконання та впровадження результатів всієї роботи.

1-й етап:

1) Основна заробітна плата розробників (дослідників)  $Z_o$ , які працюють в наукових установах бюджетної сфери розраховується за формулою:

$$Z_o = \frac{M}{T_p} \cdot t \text{ [грн]}, \quad (4.1)$$

де  $M$  – місячний посадовий оклад конкретного розробника, грн. Величини окладів знаходиться в межах (8 000...12 000) грн. за місяць;

$T_p$  – число робочих днів в місяці; прийmemo  $T_p = 20$  днів;

$t$  – число робочих днів роботи розробника (дослідника).

Зроблені розрахунки зведемо до таблиці 4.4:

Таблиця 4.4 – Розрахунок основної заробітної плати

Найменування посади виконавця	Місячний посадовий оклад, грн.	Оплата за робочий день, грн.	Число днів роботи	Витрати на оплату праці, грн.	Примітка
1. Керівник роботи	10 000	500	5	2 500	
2. Студент	4 200	140	30	4 200	
3. Консультант	5 000	250	5	1 250	
Всього				$Z_o = 7 950$	

2) Додаткова заробітна плата  $Z_d$  всіх розробників розраховується як  $(10...12)\%$  від величини основної заробітної плати, тобто:

$$Z_d = (0,1 \dots 0,12) \cdot Z_o \quad (4.2)$$

Для нашого випадку:

$$Z_d = 0,12 \cdot 7950 = 954(\text{грн})$$

3) Нарахування на заробітну плату  $H_{зп}$  фахівців, які беруть участь у виконанні наукової роботи, розраховуються за формулою:

$$H_{зп} = (Z_o + Z_d) \cdot \frac{\beta}{100}, \quad (4.3)$$

де  $\beta$  – ставка єдиного внеску на загальнообов'язкове державне соціальне страхування, %.

Ставка єдиного внеску на загальнообов'язкове державне соціальне страхування для бюджетних установ визначена в 22%. Тоді:

$$H_{зп} = (7950 + 954) \cdot 0,22 \approx 1\,958 \text{ (грн)}$$

4) Амортизація обладнання, комп'ютерів та приміщень  $A$ , які використовувались під час виконання даної роботи.

У спрощеному вигляді амортизаційні відрахування  $A$  в цілому можуть бути розраховані за формулою:

$$A = \frac{Ц}{T_{кор}} \cdot \frac{T}{12} \text{ [грн]}, \quad (4.4)$$

де  $\Pi$  – загальна балансова вартість всього обладнання, комп'ютерів, приміщень тощо, що використовувались для виконання даної роботи, грн.;

$T_{\text{кор}}$  – термін корисного використання, роки.

$T$  – термін, використання обладнання, приміщень тощо, місяці.

Зроблені розрахунки зведемо до таблиці 4.5:

Таблиця 4.5 – Розрахунок амортизації обладнання, приміщень.

Найменування обладнання, приміщень тощо	Балансова вартість, грн.	Термін корисного використання, роки	Термін використання, міс.	Величина амортизаційних відрахувань, грн.
1.Персональні комп'ютери, принтери, інші пристрої	12000	2	1	500
2.Приміщення університету	180 000	20	1	750
Всього				A = 1 250

- 4) Витрати на матеріали  $M$ , що були використані під час виконання роботи, розраховуються по кожному виду матеріалів за формулою:

$$M = \sum_1^n H_i \cdot \Pi_i \cdot K_i - \sum_1^n V_i \cdot \Pi_v \text{ [грн]}, \quad (4.5)$$

де  $H_i$  – витрати матеріалу  $i$ -го найменування, кг;

$\Pi_i$  – вартість матеріалу  $i$ -го найменування, грн./кг.;

$K_i$  – коефіцієнт транспортних витрат,  $K_i = (1,1 \dots 1,15)$ ;

$V_i$  – маса відходів матеріалу  $i$ -го найменування, кг;

$\Pi_v$  – ціна відходів матеріалу  $i$ -го найменування, грн/кг;

$n$  – кількість видів матеріалів.

Таблиця 4.6 - Витрати на матеріали

Матеріали та комплектуючі	Ціна, грн	Витрати, шт	Загальна вартість
Папір	120	1	120
Тонер	500	0,05	25
Флешка	150	1	150
Ручка	10	1	10
CD-диск	15	1	15
Всього з урахуванням транспортних витрат			352

б) Витрати на силову електроенергію  $V_e$  розраховуються за формулою:

$$V_e = V \cdot P \cdot \Phi \cdot K_n \text{ [грн]}, \quad (4.6)$$

де  $V$  – вартість 1 кВт-год. електроенергії, в 2019 р.  $V \approx 8,44$  грн./кВт;

$P$  – установлена потужність обладнання, кВт;

$\Phi$  – фактична кількість годин роботи обладнання, годин. Прийmemo, що  $\Phi = 150$  годин;

$K_n$  – коефіцієнт використання потужності;  $K_n < 1 = 0,85$ .

У виконанні роботи були задіяні силові блоки потужністю 1,0 кВт.

Тоді витрати на електроенергію складуть:

$$V_e = 8,44 * 0,5 * 150 * 0,85 = 1076,1 \text{ (грн)}$$

7) Інші витрати  $V_{ін}$  можна прийняти як (100...300)% від суми основної заробітної плати розробників, які виконували дану роботу, тобто:

$$V_{ін} = (1 \dots 3) \cdot Z_o. \quad (4.7)$$

Для нашого випадку:

$$B_{in}=1*7950=7950(\text{грн})$$

8) Сума всіх попередніх статей витрат дає витрати на виконання даної роботи безпосередньо самим розробником – В.

Для нашого випадку:

$$B = 7950+954+1958+1250+352+7950+538=20952(\text{грн})$$

2-й етап:

Передбачає розрахунок загальних витрат виконання даної роботи всіма виконавцями.

Загальна вартість даної роботи визначається за  $B_{zag}$  формулою:

$$B_{zag} = \frac{B}{\alpha}, \quad (4.8)$$

де  $\alpha$  – частка витрат, які безпосередньо здійснює виконавець даного етапу роботи, у відн. одиницях. Для нашого випадку  $\alpha = 0,95$ .

$$B_{zag} = \frac{20952}{0,95} = 22054 \text{ (грн)}.$$

3-й етап:

Передбачає прогнозування загальних витрат на розробку та впровадження результатів виконаної роботи.

Прогнозування загальних витрат здійснюється за формулою:

$$ЗВ = \frac{B_{zag}}{\beta}, \quad (4.9)$$

де  $\beta$  – коефіцієнт, який характеризує етап (стадію) виконання даної роботи.

Оскільки розробка знаходиться на стадії впровадження, то  $\beta \approx 0,8-0,9$ .

Тоді:

$$ЗВ = \frac{20952}{0,9} = 23\ 280 \text{ (грн)}.$$

Тобто прогнозовані витрати на розробку системи розпізнавання мовців, дослідження та впровадження результатів даної роботи становлять приблизно 23 тис. грн.

#### 4.3 Прогнозування комерційних ефектів від реалізації результатів розробки

Аналіз місткості ринку даної продукції показує, що в даний час в Україні кількість потенційних користувачів подібних систем складає щороку приблизно 5 юридичних осіб. Середня ціна подібної розробки становить в даний час приблизно 1250 тис. грн. Розроблений нами метод має значно кращі функціональні можливості, тому його можна буде реалізовувати на ринку дорожче, приблизно за 8000 грн або на 500 грн більше. Розроблений нами метод практично може бути впроваджений з 1 січня 2020 року.

Прогноз попиту на нашу розробку є таким (по роках):

1-й рік після впровадження (2020 р.) – приблизно 500 шт;

2-й рік після впровадження (2021 р.) – приблизно 550 шт;

3-й рік після впровадження (2022 р.) – приблизно 600 шт.

Очікуване збільшення чистого прибутку  $\Delta\Pi_i$ , що його може отримати потенційний інвестор від впровадження результатів нашої розробки, для кожного із років, починаючи з першого року впровадження становитиме:

$$\Delta\Pi_i = (\Delta C_0 \cdot N + C_0 \cdot \Delta N)_i \cdot \lambda \cdot \rho \cdot \left(1 - \frac{\nu}{100}\right), \quad (4.10)$$

де  $\Delta C_0$  – зміна ціни реалізації розробленого методу відносно ціни аналога;

$N$  – попит на аналогічну розробку в даний час;  $N = 5$ ;

$\Delta N$  – збільшення попиту на розроблений нами метод, який по роках становитиме:  
+1; +2; +2; +1;

$C_0$  – ціна можливої реалізації розробленого нами методу;  $C_0 = 1500$  тис. грн;

$\lambda$  – коефіцієнт, що враховує сплату ПДВ; коефіцієнт  $\lambda = 0,8333$ ;

$\rho$  – коефіцієнт, який враховує рентабельність виробництва.

Рекомендується приймати  $\rho = (0,23...0,5)$ ; візьмемо  $\rho = 0,35$ ;

$\nu$  – чинна ставка податку на прибуток;  $\nu = 18\%$ .

Тоді, збільшення чистого прибутку  $\Delta \Pi_1$  для потенційного інвестора протягом першого року від реалізації нашої розробки (2020) становитиме:

$$\Delta \Pi_1 = [500 * 1 + (8000 + 500) * (500)] \cdot 0,8333 \cdot 0,25 \cdot \left(1 - \frac{18}{100}\right) = 723222 \text{ грн.}$$

Збільшення чистого прибутку  $\Delta \Pi_2$  для потенційного інвестора від реалізації нашої розробки протягом другого (2021 р.) року складе:

$$\Delta \Pi_2 = [500 * 1 + (8000 + 500) * (500 + 550)] \cdot 0,8333 \cdot 0,25 \cdot \left(1 - \frac{18}{100}\right) = 1518673 \text{ грн.}$$

Збільшення чистого прибутку  $\Delta \Pi_3$  для потенційного інвестора протягом третього (2022) року становитиме:

$$\Delta \Pi_3 = [500 * 1 + (8000 + 500) * (500 + 550 + 600)] \cdot 0,8333 \cdot 0,25 \cdot \left(1 - \frac{18}{100}\right) = 2386438 \text{ грн.}$$

Теперішня вартість інвестицій  $PV$ , що можуть бути вкладені у комерціалізацію нашої розробки, становлять:

$$PV = 3B \quad (4.11)$$

де  $3B$  – загальні витрати на виконання даної роботи,  $3B = 56$  тис. грн;

$$PV = 23 \text{ тис. грн.}$$

Абсолютний економічний ефект, що його може отримати потенційний інвестор від вкладених інвестицій  $E_{\text{абс}}$ :

$$E_{\text{абс}} = \text{ПП} - PV, \quad (4.12)$$

де  $\text{ПП}$  – приведена вартість всіх чистих прибутків, що їх може отримати потенційний інвестор, грн;

$PV$  – теперішня вартість інвестицій  $PV = 23$  тис. грн.

Приведена вартість всіх чистих прибутків ПП розраховується за формулою:

$$ПП = \sum_1^t \frac{\Delta\Pi_i}{(1 + \tau)^t}, \quad (4.13)$$

де  $\Delta\Pi_i$  – збільшення чистого прибутку у кожному із років, протягом яких виявляються результати впровадженої роботи, грн;

$t$  – період часу, протягом якого виявляються результати впровадженої роботи, роки;  $t = 4$  роки;

$\tau$  – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні. Прийmemo, що  $\tau = 0,10$ ;

$t$  – період часу (в роках) від моменту можливого отримання чистих прибутків до точки «0» на осі координат (див. рис. 4.1).

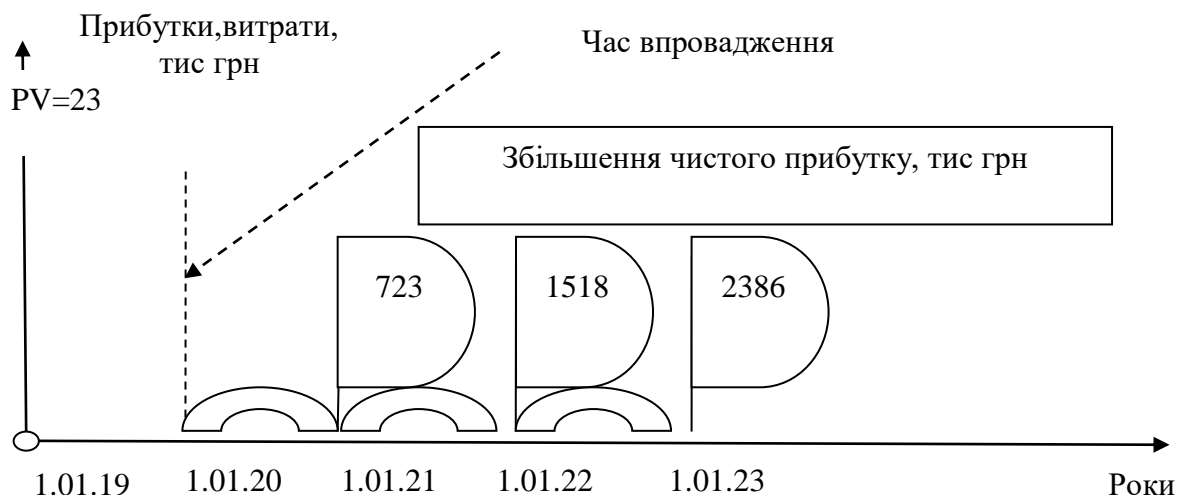


Рисунок 4.1 – Вісь часу з фіксацією платежів (у грн)

Тоді приведена вартість всіх чистих прибутків ПП, що їх може отримати потенційний інвестор від впровадження нашої розробки, складе:

$$ПП = \frac{723222}{(1 + 0,1)^3} + \frac{1518673}{(1 + 0,1)^4} + \frac{2386438}{(1 + 0,1)^5} = 543367 + 1037274 + 1481790 = 3062431$$

$\approx 3\,063$  тис. грн

Абсолютний ефект від реалізації нашої розробки за 4 роки становитиме:

$$E_{\text{абс}} = 3063 - 23 = 3040 \text{ тис. грн.}$$



Відносна ефективність  $E_B$  коштів, вкладених у впровадження нашої розробки:

$$E_B = T_{ж} \sqrt{1 + \frac{E_{абс}}{PV}} - 1, \quad (4.14)$$

де  $E_{абс}$  – абсолютний ефект вкладених коштів,  $E_{абс} = 2290$  тис грн;

$PV$  – теперішня вартість інвестицій  $PV = 120$  тис. грн;

$T_{ж}$  – життєвий цикл розробки, роки;  $T_{ж} = 5$  років.

Для нашого випадку отримаємо:

$$E_B = 5 \sqrt{1 + \frac{3040}{23}} - 1 = 1,65 \approx 165\%$$

Мінімальна дохідність  $\tau_{мін}$ , нижче за яку кошти у впровадження нашої розробки вкладатися не будуть, становитиме:

$$\tau = d + f, \quad (4.15)$$

де  $d$  – середньозважена ставка за депозитними операціями в комерційних банках; в 2018 році в Україні  $d = (0,13...0,20)$ ;

$f$  – показник, що характеризує ризикованість вкладень; зазвичай, величина  $f = (0,2...0,5)$ .

Для нашого випадку отримаємо:

$$\tau_{мін} = 0,20 + 0,50 = 0,70 \text{ або } \tau_{мін} = 70\%.$$

Оскільки величина  $E_B = 165\% > \tau_{мін} = 70\%$ , то інвестор може бути зацікавлений у фінансуванні нашої розробки.

Розрахуємо термін окупності  $T_{ок}$  коштів, вкладених у впровадження розробленого нами методу класифікації зображень:

$$T_{ок} = \frac{1}{E_B}. \quad (4.16)$$

комерціалізації даної розробки.

Для нашого випадку термін окупності коштів  $T_{ок}$  становитиме:

$$T_{ок} = \frac{1}{1,65} = 0,6 < 3\text{-х років},$$

Якщо  $T_{ок} < 3$ -х років, то це свідчить про доцільність та ефективність що свідчить про потенційну доцільність фінансування впровадження розробленого нами методу класифікації зображень.

Отже, нами була розроблена та змодельована система для розпізнавання мовців в системі Matlab, яка буде цікавою для інвестування, так як:

- витрати на розробку методу не високі - 23 тис. грн
- абсолютний економічний ефект розробки становить 2361 (за 4 роки) або по 590,25 тис. грн щорічно
- дуже хороший показник внутрішньої норми дохідності, а саме 165%
- термін окупності вкладених коштів становить приблизно 0,6 років.

## ВИСНОВКИ

У магістерській кваліфікаційній роботі нейромережеву систему автентифікації суб'єкта за візуальним представленням його голосу. Така система дозволяє підвиити інформаційну безпеку доступу до баз даних або інших інформаційних сервісів.

Основні результати досягнуті при виконанні кваліфікаційної роботи такі:

1. Розглянуто і проаналізовано за обраними критеріями основні методи аналізу та оброблення мовних сигналів
2. Здійснено порівняння і аналіз актуальних існуючих систем автентифікації суб'єктів за голосом.
3. Викладена суть проблеми автентифікації суб'єктів за голосом та отримано узагальнену постановку задачі синтезу нейромережевої системи автентифікації суб'єкта за візуальним представленням його голосу.
4. Розроблено структурну схему та UML-діаграми нейромережевої системи автентифікації суб'єкта за візуальним представленням його голосу.
5. Викладено математику процедур вейвлет-перетворення мовних сигналів, їх фільтрації та часової нормалізації тривалості їх звування, в т.ч. алгоритму нелінійної часової нормалізації.
6. У програмному середовищі MATLAB розроблено програмне забезпечення для автентифікації суб'єкта за візуальним представленням його голосу.
7. Проведено дослідження доцільності використання часової нормалізації мовних сигналів у задачі автентифікації суб'єктів за голосом та встановлено, що використання процедури нормалізації є доцільним.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Голосовая идентификация [Электронный ресурс]/ Режим доступа: [http://www.uchet-rabochego-vremeni.ru/uchet\\_03.htm](http://www.uchet-rabochego-vremeni.ru/uchet_03.htm)
2. Шпак З.Я. Модель сповільнення темпу подання мовної інформації // Вісник Державного університету "Львівська політехніка": Комп'ютерна інженерія та інформаційні технології. - Львів, 1998. - №351. - С. 98-106.
3. Рамишвили Г.С. Речевой сигнал и индивидуальность голоса / Г.С. Рамишвили. – Тбилиси: МЕЦНИЕРЕБА, 1976. – 183 с.
4. Речевые технологи [Электронный ресурс]/ Режим доступа: [www.mtu-net.ru/pinetar/dm/](http://www.mtu-net.ru/pinetar/dm/)
5. Идентификация по голосу [Электронный ресурс]/ Режим доступа: <http://www.des-crypto.ru/itsecur/voice/>
6. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу / Г.С. Рамишвили. – М.: Радио и связь, 1981. – 224 с.
7. Рабинер Л.Р. Цифровая обработка речевых сигналов. Пер. с англ./ Л.Р. Рабинер, Р.В. Шафер. -М.:Радио и связь,1981.
8. Oppenheim A.V. Digital Signal Processing / A.V. Oppenheim, R.W. Shafer. – New Jersey: Prentice-Hall, 1975. – 273 p.
9. Бахвалов Н.С. Численные методы / Н.С. Бахвалов, Н.П. Жидков. - М.:Наука,1987.
10. Краскевич В. Е. Численные методы в инженерных исследованиях / В.Е. Краскевич, К.Х. Зеленський, В.И. Гречко. – К.: Высшая шк., 1986. – 263 с.
11. Вентцель Е.С. Исследование операций / Е.С. Вентцель. - М.:Наука,1988.
12. Теория нейронных сетей. Учеб. пособие для вузов / Общая ред. А. И. Галушкина. - М.: ИПРЖР. 2000р. – 416 с.
13. Kussul E. Rosenblatt Perceptrons for Handwritten Digit Recognition / / IEEE. - 2001. - P. 1516-1520.

14. Перцептрон [Электронный ресурс]/ Режим доступу:  
<http://znaimo.com.ua/%D0%9F%D0%B5%D1%80%D1%86%D0%B5%D0%BF%D1%82%D1%80%D0%BE%D0%BD>
15. Ли У. Методы автоматического распознавания речи / У. Ли. – М.: Мир, 1983. – 716 с.
16. Червенчук И.В. Моделирование информационных систем с помощью UML / И.В. Червенчук. - М.: Вильямс, 2006. – 48 с.
17. Иванов Д. Моделирование на UML. Учебно-методическое пособие / Д. Иванов, Ф. Новиков. – СПб.: СПбГУ ИТМО, 2010. – 200 с.
18. Буч Г. Язык UML. Руководство пользователя / Г. Буч, Д. Рамбо, И. Якобсон. – М.: Вильямс, 2006. – 496 с.
19. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк. – Киев: Наукова думка, 1987. – 262 с.
20. Айфичер Э. Цифровая обработка сигналов. Практический подход. 2-е издание / Э. Айфичер, Б.Джервис. - М.: Вильямс, 2004. – 992 с.
21. Oliver Rioul. Wavelets and signal processing // IEEE SP Magazine. - №10. – 1991. – Р. 14-34.
22. Потапова Р.К. Речь: коммуникация, информатика, кибернетика / Р.К. Потапова. – М.: Радио и связь – 1997. – 236 с.
23. Демида Б.Р. Обробка мовної інформації на IBM PC // Тези доповідей Третьої міжнар. конф. "Проблеми українізації комп'ютерів". Львів, 1993. - С. 43-44.
24. Ковтун В.В. Розробка критерію оцінки ефективності систем ідентифікації мовця. // Вісник Хмельницького національного університету. – 2005. – №4. – Ч.1. – Т.2(68). – С. 117-122.
25. Ковтун В.В. Вибір інформативних ознак в задачі ідентифікації мовця // В.В. Ковтун, М.М. Биков / Збірка матеріалів Міжнародної конференції з індуктивного моделювання “МКІМ–2002”.– Том 1.– Львів: НУЛП.– 2002.– С. 280–286.
26. Фланаган Дж. Анализ, синтез и восприятие речи / Дж. Фланаган. – М.: Связь, 1968. – 392 с.

27. Сорокин В.Н. Теория речеобразования / В.Н.Сорокин. – М.: Радио и связь, 1985. – 312 с.
28. Кортаев Г.А. Анализ и синтез речевого сигнала методом линейного предсказания // Зарубежная радиоэлектроника. – 1990. – №3. – С. 31–50.
29. Кельманов А.В. Решения проблем выделения спектрально–корреляционных признаков речевого сигнала / А.В. Кельманов. – М: Мир, 1991.– 170 с.
30. Kohonen F. Self organization and associative memory / F. Kohonen. – Berlin, 1984. – 416 p.
31. Бонгард. М. М. Проблема впізнавання / М.М. Бонгард. - М.: Наука, 1967. - 320 с.
32. Системи розпізнавання образів [Електронний ресурс]/ Режим доступу: [www.iskint.ru/?xid=books/sotnik/-part3](http://www.iskint.ru/?xid=books/sotnik/-part3)
33. Измаилов А. Ф. Численные методы оптимизации: Учеб. пособие. / А.Ф. Измаилов. - М.: ФИЗМАТЛИТ, 2003. - 304 с.
34. Васюра А.С. Техніка передавання дискретної інформації / А.С. Васюра, С.Г. Кривогубченко, А.Я. Кулик, М.М. Компанець. - Вінниця: ВДТУ, 1998.
35. Fant G. Voice Source Parameters in Continious Speech. // ICSLP–94. – 1994. – P. 1451–1454.
36. Bung E. Automatic speaker recognition by computers / E. Bung. – Proc. Carnahan conf. on crime contermeasures. – Lexington. – 1975. – P.37-48.
37. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети / А.П. Ротштейн. – Винница: "УНИВЕРСУМ-Вінниця", 1999. – 320 с.
38. Beigi Homayoon. Fundamentals of Speaker Recognition / Homayoon Beigi. - New York: Springer, 2011. - 146 p.
39. Дьяконов В. MATLAB. Обработка сигналов и изображений / В.Дьяконов, И. Абраменкова. – СПб.: Питер, 2002. – 680 с.
40. Курбатова Е.А. MATLAB 7. Самоучитель / Е.А. Курбатова. – М.: Вильямс, 2005г. - 256 с.

41. Поршнеv С.В. MATLAB 7. Основы работы и программирования / С.В. Поршнеv. – М.:Бином. Лаборатория знаний, 2006г. - 320 с.
42. Бондарев В.Н., Трёстер Г., Чернега В.С. Цифровая обработка сигналов: методы и средства: Учеб. пособие для вузов. — Севастополь: Изд-во СевГТУ, 2002.-398с.
43. Jawerth В., Sweldens W. An overview of wavelet based multiresolution analyses//SIAM Rev.-1994. — 36(3).
44. Zurera M.R., Ferreras F.L., Amores M.P., Bascon S.M., Reyes N.R. New algorithm for translating psycho-acoustic information to the wavelet domain // Signal Processing 81(2001).
45. Строганов В.А. Применение wavelet-преобразования к задаче сжатия звуковых сигналов / В.А. Строганов, В.Н. Бондарев // Системы автоматки и автоматическое управление. Материалы студенч. науч.-техн. конф. г. Севастополь, 14-15мая 2002г. — Севастополь: Изд-во СевНТУ,2002.—С.13-16
46. Mohamed K. Nezami. RF Architectures & Digital Signal Processing Aspects of Digital Wireless Transeivers: 2003.
47. Оппенгейм А., Шафер Р. Цифровая обработка сигналов: Пер. с англ. - М.: Техносфера, 2006. - 856 с.
48. Сергиенко А.Б. Цифровая обработка сигналов: - СПб.: Питер, 2002. - 608 с.
49. Vaidyanathan P. P. Multirate systems and filters banks: -New Jersey: Prentice hall, 1993
50. Блаттер К. Вейвлет-анализ. Основы теории. — М.: Техносфера, 2006. — 280 с.

## **ДОДАТКИ**



Додаток А  
(обов'язковий)

ВНТУ

ТЕХНІЧНЕ ЗАВДАННЯ

на виконання магістерської кваліфікаційної роботи

**Нейромережева система автентифікації суб'єкта за візуальним  
представленням його голосу**

Студент групи 2АКІТ-18м Воловодівська О.О.

“ \_\_\_ ” \_\_\_\_\_ 2019 р.

Керівник к.т.н., доцент Ковтун В.В.

“ \_\_\_ ” \_\_\_\_\_ 2019 р.

Вінниця 2019

## 1. Назва та галузь застосування

1.1. Назва – Нейромережева система автентифікації суб'єкта за візуальним представленням його голосу.

1.2. Галузь застосування – Комп'ютерні системи біометричної інформаційної безпеки.

2. Підстава для проведення розробки.

Тема магістерської кваліфікаційної роботи затверджена наказом по ВНТУ № 254 від “ 02 “ \_\_\_\_\_ 10 \_\_\_\_\_ 20 19 р.

## 3. Мета та призначення розробки.

Метою магістерської кваліфікаційної роботи є розробка методу візуалізації мовних записів, оптимізованого для задачі автентифікації особи суб'єкта за голосом.

## 4. Вихідні дані для проведення розробки.

Магістерська кваліфікаційна робота виконується вперше. В ході проведення розробки повинні використовуватись такі документи:

1. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. – М.: Радио и связь, 1981. – 224 с. .;

2. Бондарев В.Н., Трёстер Г., Чернега В.С. Цифровая обработка сигналов: методы и средства: Учеб. пособие для вузов. — Севастополь: Изд-во СевГТУ, 2002.-398с.;

3. Дьяконов В. П. MATLAB 6.5 SP1/7.0 + Simulink 5/6/ Обработка сигналов и проектирование фильтров. — М.: СОЛОН-Пресс, 2005. — 676 с.

## 5. Вимоги до розробки.

5.1. Перелік головних функцій:

Програма повинна здійснювати приведення мовного сигналу до заданої тривалості звучання та здійснювати візуалізацію отриманого результату.

5.2. Основні технічні вимоги до розробки.

5.2.1. Вимоги до програмної платформи:

- ОС WINDOWS;
- ПЗ MATLAB.

5.2.2. Умови експлуатації системи:

- робота на стандартних ПЕОМ в приміщеннях зі стандартними умовами;
- можливість цілодобового функціонування системи;
- текст програмного забезпечення системи є цілком закритим.

## 6. Економічні показники

До економічних показників входять:

- термін окупності не більше \_\_\_\_\_ років;
- розмір капіталовкладень не більше \_\_\_\_\_ грн.;
- економічний ефект не менше \_\_\_\_\_ грн.;
- рентабельність не менше \_\_\_\_\_ %;
- інші економічні переваги у порівнянні з аналогами.

## 7. Стадії та етапи розробки.

## 7.1 Пояснювальна записка:

- |   |   |              |
|---|---|--------------|
| 1 | Аналіз методів оцінювання ризику прийняття керівних рішень при управлінні розгалужено-циклічними технологічними процесами. Постановка задач дослідження | 12.09.2019р. |
| 2 | Удосконалення технології прийняття рішень при управлінні розгалужено-циклічними технологічними процесами  | 22.09.2019р. |
| 3 | Практична реалізація та аналіз отриманих результатів  | 3.10.2019р.  |
| 4 | Підготовка економічної частини  | 12.11.2019р. |
| 5 | Апробація результатів дослідження   | 22.11.2019р. |
| 6 | Публікації  |              |
| 7 | Оформлення пояснювальної записки, графічного матеріалу і презентації  | 30.11.2019р. |
| 8 | Захист МКР  | 17.12.2019р. |

## 7.2 Графічні матеріали:

- алгоритм прийняття рішення при управлінні РЦТП: «01 » 12. 2019 р.
  - UML діаграма варіантів використання: «01 » 12. 2019 р.
  - UML-діаграма діяльності: «03 » 12. 2019 р..
  - UML-діаграма класів: «03 » 12. 2019 р.
  - вигляд екрану «Головне меню ІТ «TP Modeling»»: «06 » 12. 2019 р.
  - вигляд екрану «Total Risk»: «06 » 12. 2019 р.
  - 8. Порядок контролю і приймання.
- 8.1. Хід виконання магістерської кваліфікаційної роботи контролюється керівником роботи, консультантами з економічної частини. Рубіжний контроль провести до «7» грудня 2019 р.
- 8.2. Атестація проекту здійснюється на попередньому захисті. Попередній захист магістерської кваліфікаційної роботи провести до «10» грудня 2019 р.
- 8.3. Підсумкове рішення щодо оцінки якості виконання магістерської кваліфікаційної роботи приймається на засіданні ДЕК. Захист магістерської кваліфікаційної роботи провести «17» грудня 2019 р.

Додаток Б  
(обов'язковий)  
Лістинг програми

```

% Попередня обробка звукового сигналу
[signal,fs] = audioread('s1.wav') ;
d=length(signal);
tim=1; i=1;
while i<d-408 y=signal(i:i+408);
x(1)=0.0;
for j=2:409 x(j)=y(j)-y(j-1); end;
for j=1:409 z(j)=(0.54-0.46*cos(2*pi*(j-1)/408))*x(j); end;
%Вікно Хеммінга
C=fft(z,512); C=abs(C); % Перетворення Фур'є
S=C(1:256); % амплітуди

f=[0; 74.24; 156.4; 247.2; 347.6; 458.7; 581.6; 717.5;
867.9; 1034; 1218; 1422; 1647; 1895; 2171; 2475; 2812; 3184;
3596; 4052; 4556; 5113; 5730; 6412; 7166; 8000];

krok=16000/512;
a(1:26)=0; j=2; k=1; n(1:26)=0;
h=krok*(j-1);

while k<26
while and(f(k)<h,h<f(k+1)) alfa=(h-f(k))/(f(k+1)-f(k));
a(k+1)=a(k+1)+S(j)*alfa;
n(k+1)=n(k+1)+1;
a(k)=a(k)+S(j)*(1-alfa);
n(k)=n(k)+1;
j=j+1; h=krok*(j-1);
end;
a(k)=a(k)/n(k); k=k+1;
end;

O(tim,1:24)=a(2:25);
norma(tim)=norm(O(tim,1:24));
i=i+160; tim=tim+1;
end;
time=time-1;

normamax=max(norma(1:time));
O(1:time,1:24)= O(1:time,1:24)/normamax; % нормалізація

```

```

% Графіки
subplot(3,3,1); plot(signal);title('Звуковий сигнал');
subplot(3,3,2); plot(y);title('        Частина сигналу ');
subplot(3,3,3); plot(x);title('        Фільтр першого порядку
');
subplot(3,3,4); plot(z);title('        Хеммінг');
subplot(3,3,5); plot(C);title('        ШПФ ');
subplot(3,3,6); plot(O(time,1:24));title('Оброблена частина
сигналу');

% Обчислення МЧКК
    Tw = 20;                % тривалість кадру (мс)
    Ts = 10;                % тривалість рамки зсуву (мс)
    alpha = 0.97;          % коефіцієнт
    M = 20;                 % кількість каналів у банку
фільтрів
    C = 12;                 % кількість кепстральних
коефіцієнтів
    L = 22;                 % параметр кепстру
    LF = 300;               % межа низьких частот
    HF = 3700;              % межа високих частот
    wav_file = 's1.wav';    % вхідний звуковий сигнал
%Читання звукового сигналу з файлу
    [ speech, fs] = audioread( wav_file );
    [ MFCCs, FBEs, frames ] = ...
mfcc( speech, fs, Tw, Ts, alpha, @hamming, [LF HF], M,
C+1, L );

[ Nw, NF ] = size( frames );% довжина кадру та кількість
кадрів
time_frames = [0:NF-1]*Ts*0.001+0.5*Nw/fs;% вектор часу для
кадрів
time = [ 0:length(speech)-1 ]/fs;        % вектор часу для
сигналу
logFBEs = 20*log10( FBEs );                % обчислення логарифму
logFBEs_floor = max(logFBEs(:))-50;
logFBEs( logFBEs<logFBEs_floor ) = logFBEs_floor;
figure('Position', [30 30 800 600], 'PaperPositionMode',
'auto',...
'color', 'w', 'PaperOrientation', 'landscape', 'Visible',
'on' );
    subplot( 311 );
    plot( time, speech, 'k' );
    xlim( [ min(time_frames) max(time_frames) ] );

```

```

xlabel( 'Time (s)' );
ylabel( 'Amplitude' );
title( 'Speech waveform' );
subplot( 312 );
imagesc( time_frames, [1:M], logFBEs );
axis( 'xy' );
xlim( [ min(time_frames) max(time_frames) ] );
xlabel( 'Time (s)' );
ylabel( 'Channel index' );
title( 'Log (mel) filterbank energies' );
subplot( 313 );
imagesc( time_frames, [1:C], MFCCs(2:end,:) );
axis( 'xy' );
xlim( [ min(time_frames) max(time_frames) ] );
xlabel( 'Time (s)' );
ylabel( 'Cepstrum index' );
title( 'Mel frequency cepstrum' );

% Вейвлет як базис розкладу
wname = 'db10';
% Дискретний розклад із використанням періодизації
expand = 'ppd';
[sa, sd] = dwt (MFCCs, wname, 'mode', expand);
% Графіки
subplot (311)
plot (MFCCs), title ('Вихідний сигнал')
subplot (312)
plot (sa), title ('Коефіцієнти аппроксимації')
subplot (313)
plot (sd), title ('Коефіцієнти деталізації')

```

Додаток В – Перелік графічних матеріалів  
(Обов'язковий)

**ЗАТВЕРДЖУЮ**  
Завідувач кафедри КСУ  
д.т.н., проф. В.М. Дубовой

« \_\_\_\_\_ » \_\_\_\_\_ 2019 р.

**ПЕРЕЛІК  
ГРАФІЧНИХ МАТЕРІАЛІВ**

для захисту магістерської кваліфікаційної роботи  
на тему

**НЕЙРОМЕРЕЖЕВА СИСТЕМА АВТЕНТИФІКАЦІЇ СУБ'ЄКТА ЗА  
ВІЗУАЛЬНИМ ПРЕДСТАВЛЕННЯМ ЙОГО ГОЛОСУ**

1. Актуальність роботи
2. Об'єкт та предмет дослідження
3. Мета роботи
4. Технології розробки
5. UML-діаграма варіантів використання
6. UML-діаграма послідовності
7. UML-діаграма активності
8. Екранні форми та результати тестування створеного ПЗ
9. Результати автентифікації 20 мовців згортковою нейромережею на основі спектрограми або вейвлет-грами вхідного мовного сигналу.
10. Висновки

Розробив: Воловодівська О.О.

\_\_\_\_\_ (підпис) \_\_\_\_\_ (дата)

Перевірив: Ковтун В.В.

\_\_\_\_\_ (підпис) \_\_\_\_\_ (дата)

Рецензент: \_\_\_\_\_

\_\_\_\_\_ (підпис) \_\_\_\_\_ (дата)

Вінниця 2019

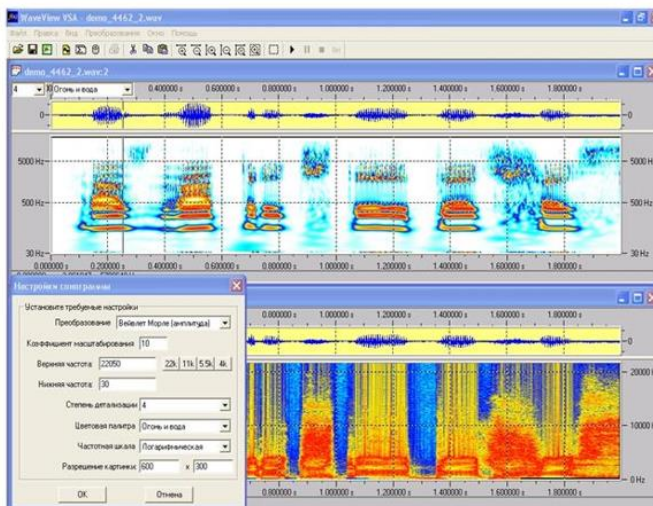
## Актуальність роботи

Можливість автентифікації суб'єкта за голосом засновано на аналізі унікальних характеристик мовного сигналу, обумовлених анатомічними властивостями (ротової порожнини, горла, голосових зв'язок тощо) та набутими мовленнєвими звичками (манера, гучність, швидкість мови).

Перевагою технології автентифікації суб'єкта за голосом забезпечується, насамперед, природністю мовного спілкування та невіддільністю мовної інформації від людини.

Складність такого підходу до автентифікації є те, що основним призначенням мовного спілкування є комунікація, отже, у мовному сигналі інформація про індивідуальні особливості голосу мовця накладаються на змістовну інформацію і для відокремлення цих складових мовний сигнал зазнає фільтрації та параметризації із представленням у вигляді спектрограми або вейвлет-діаграми.

## Актуальність роботи



Мовний  
сигнал

Вейвлет-  
діаграма

Мовний  
сигнал

Спектро-  
грама

## Об'єкт і предмет дослідження

**Об'єкт дослідження** - процес мовотворення людини.

**Предметом дослідження** - методи візуалізації мовних сигналів, математичні методи оброблювання цифрових рядів, методи цифрового оброблювання сигналів, методи теорії прийняття рішень і машинного навчання.



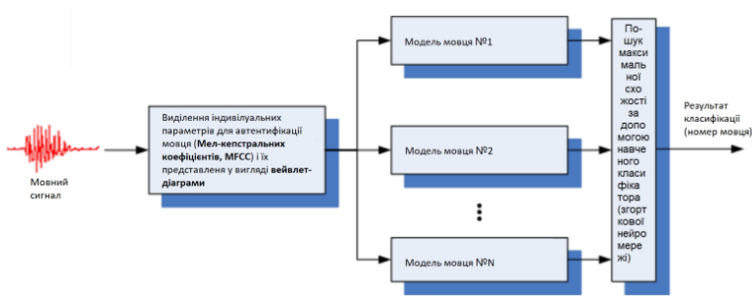
## Мета роботи

Метою є розроблення неймережевої системи автентифікації суб'єкта за візуальним представленням його голосу

### Задачі:

1. Здійснення аналітичного огляду і аналізу досягнень в галузі систем автентифікації суб'єкта за голосом;
2. Розроблення методу візуалізації мовних сигналів в задачі автентифікації суб'єкта за голосом;
3. Дослідження ефективності розробленого методу за допомогою відповідної програмної системи для автентифікації особи суб'єкта за голосом.

## Технології розробки



Створена на основі огляду аналогів структурна схема системи автентифікації Мовця за візуалізованим представленням його мовного сигналу

## Технології розробки

### Недоліки Фур'є-представлення (у вигляді спектрограми) мовного сигналу:

- втрата інформації в часі існування частотних компонентів сигналу;
- вихідний сигнал замінюється на періодичний, з періодом рівним тривалості досліджуваного зразка;
- ПФ не дає інформацію про динаміку зміни спектрального складу сигналу;
- ПФ погано працює при зміні параметрів процесу згодом (нестационарності), оскільки дає усереднені коефіцієнти для всього досліджуваного сигналу;
- ПФ вимагає знання сигналу не тільки в минулому, але й у майбутньому, що є неможливим для систем реального часу;
- погана пристосованість до сигналів з локальними особливостями.

## Технології розробки

**Вейвлет-перетворення** дозволяє судити не тільки про частотний спектр сигналу, але також про те, у який момент часу з'явилася та або інша гармоніка. З їхньою допомогою можна легко аналізувати переривчасті сигнали, або сигнали з гострими сплесками.

### **Переваги вейвлет-представлення мовного сигналу:**

локалізація в часовій і частотній області, що дозволяє робити ефективний частотно-часовий аналіз нестационарних сигналів, прикладом яких є мова;

- можливість масштабного перетворення й зсувів (стиск-розтягання досліджуваного сигналу за допомогою хвильових функцій різної періодичності);
- математичний апарат, який було розроблено для локалізації й класифікації особливих точок сигналу;
- відмінно відображають динаміку зміни сигналу уздовж „осі масштабів”, локалізацію різномасштабних деталей (спектр мови людини радикально змінюється з часом, а характер цих змін - дуже важлива інформація).

**У роботі для аналізу мовних сигналів викрписано вевлет Морле**

## Технології розробки

Після фільтрації шумів для виділення із мовного сигналу окремих елементів застосовується аналіз **енергії сигналу** на протязі кожних 10-20 мс. Обробка вхідного оцифрованого сигналу з метою зменшення обсягу вхідних даних полягає у застосуванні різних методів спектрального аналізу.

**Спектральний аналіз** даних реалізується або за допомогою віконного дискретного перетворення Фур'є або за допомогою дискретних вейвлет-перетворень. На сьогодні загальноприйнято проводити **стиснення спектру** за допомогою **процедури визначення мел-кепстральних коефіцієнтів**. Її результатом являється 16-24 коефіцієнтів, які достатньо повно характеризують весь діапазон звукових частот, які відчуває людина (від 16 Гц до 20000 Гц).

**Обчислення мел-частотних кепстральних коефіцієнтів** для обраного фрагменту мовного сигналу передбачає:

- пропускання фрагменту мовного сигналу крізь гребінку смугових фільтрів, центральні частоти яких розташовані за Мел-шкалою;
- обчислення мел-частотного кепстрального коефіцієнту всіх синалів, отриманих на виході гребінки фільтрів.

## Технології розробки

**Згорткова ж нейрона** – це різновид **глибоких нейромереж**, яка за рахунок застосування спеціальної операції – власне, згортки – дозволяє водночас зменшити кількість інформації, що зберігається в пам'яті, і виділити опорні ознаки на досліджуваному зображенні, такі як ребра, контури або грані. На наступному рівні обробки з цих ребер і граней можна розпізнати повторювані фрагменти текстур, які далі можуть скластися в фрагменти зображення. По суті кожен шар багатосарової згорткової нейромережі використовує власне перетворення. Якщо на перших шарах мережа оперує такими поняттями як "ребра", "грані" і т.п., то на осанніх шарах використовуються поняття "текстура", "частини об'єктів".

**В результаті такого опрацювання ми можемо виділити потрібний об'єкт на зображенні – індивідуальний рисунок енергії мел-частотних кепстральних коефіцієнтів, властивих конкретному мовцю.**

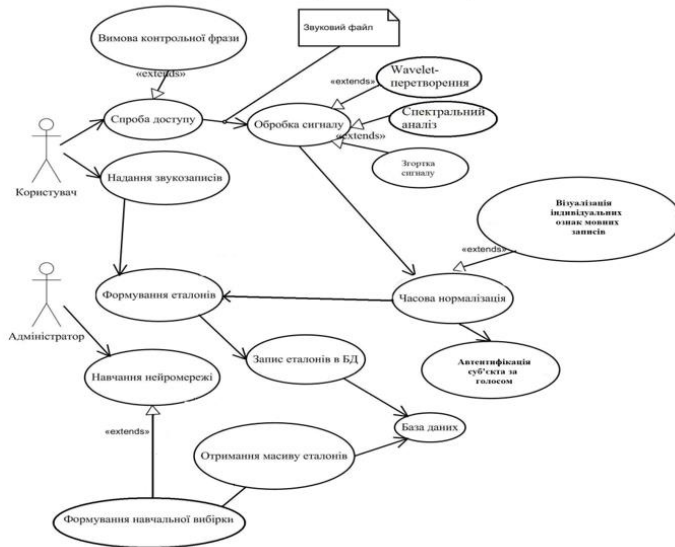
## Технології розробки

Для реалізації створеного ПЗ обрано програмне середовище **Matlab** виходячи із таких міркувань:

1. Система Matlab спеціально створена для проведення саме інженерних обчислювань: математичний апарат, що використовується в ній, гранично наближений до сучасного математичного апарата інженера, ученого й спирається на роботу з матрицями, векторами й комплексними числами; графічне подання функціональних залежностей тут організовано у формі як вимагає саме інженерна документація;
2. Мова програмування системи Matlab близька до актуальних мов програмування, але має свій синтаксис та особливості;
3. На відміну від більшості систем комп'ютерної математики, Matlab є відкритою системою; це означає, що практично всі процедури й функції Matlab доступні не тільки для використання, але й для модифікації; користувач може розширювати базу процедур та функцій створюючи власні підпрограми у тому числі за допомогою вбудованого графічного редактора користувача GUIDE; це робить Matlab незамінним засобом наукових обчислювальних робіт та імітаційного моделювання фізичних процесів;
- 4) останні версії Matlab дозволяють інтегрувати її з текстовим редактором Word, що робить зручним оформлення науково-технічних звітів, статей, дипломних та курсових проектів із включенням у них складних обчислень та виведенням графіків у текст.

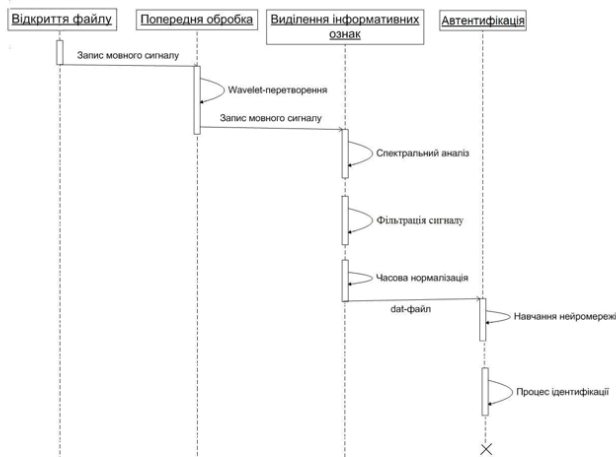
UML-діаграма варіантів використання

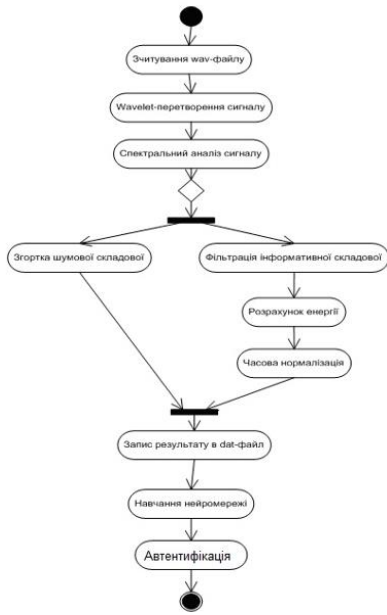
5



UML-діаграма послідовності

6



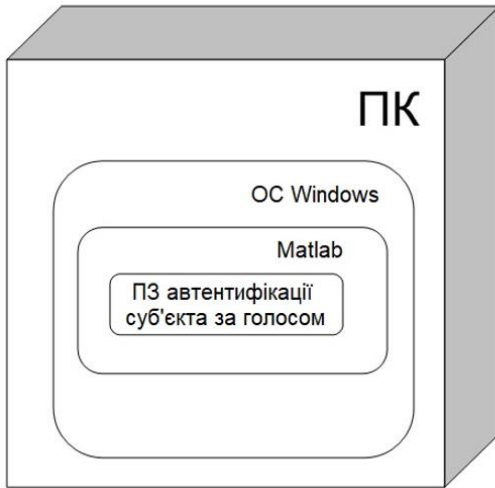


7

UML-діаграма активності

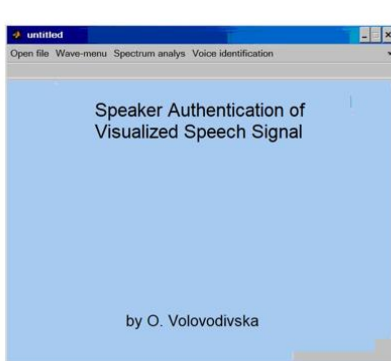
UML-діаграма розгортання

8

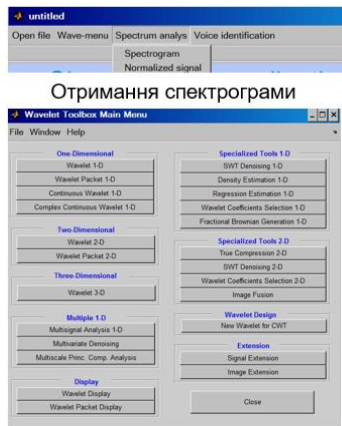


Екранні форми та результати тестування створеного ПЗ

9



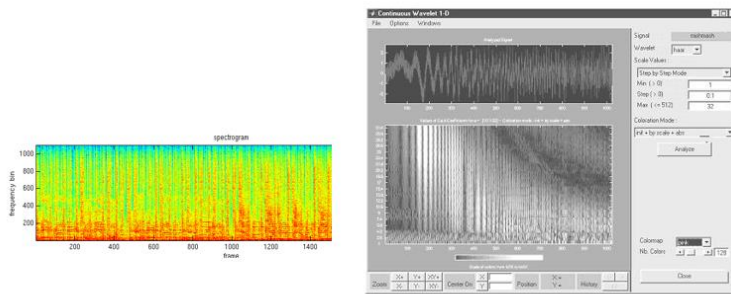
Стартове вікно ПЗ



Отримання спектрограми

Отримання вейвлет-перетворення

## Екранні форми та результати тестування створеного ПЗ



Приклад отриманих за допомогою створеного ПЗ:  
 -спектрограми (зліва),  
 -вейвлет-діаграми (справа),  
 для мовного сигналу, візуалізованого справа вверху.

### Отримання вейвлет-перетворення

Належність мовного сигналу	Результат автентифікації	
	Спектрограма	Вейвлет-грама
speaker1	Вірно	Вірно
speaker2	Вірно	Вірно
speaker3	Не вірно	Не вірно
speaker4	Вірно	Вірно
speaker5	Вірно	Вірно
speaker6	Вірно	Вірно
speaker7	Вірно	Вірно
speaker8	Не вірно	Вірно
speaker9	Вірно	Вірно
speaker10	Вірно	Вірно
speaker11	Вірно	Вірно
speaker12	Вірно	Вірно
speaker13	Вірно	Вірно
speaker14	Не вірно	Вірно
speaker15	Вірно	Вірно
speaker16	Вірно	Вірно
speaker17	Вірно	Вірно
speaker18	Вірно	Вірно
speaker19	Вірно	Вірно
speaker20	Вірно	Вірно

Результати автентифікації 20 мовців згортковою нейромережею на основі спектрограми або вейвлет-грами вхідного мовного сигналу.

Для навчання використовувалися 1500 зображень (30 с мовного матеріалу).  
 Для тестування – 150 зображень (3 парольних фрази привалістю по 1 с)

## ВИСНОВКИ

1. Розглянуто і проаналізовано за обраними критеріями основні методи аналізу та оброблення мовних сигналів
2. Здійснено порівняння і аналіз актуальності існуючих систем автентифікації суб'єктів за голосом.
3. Викладена суть проблеми автентифікації суб'єктів за голосом та отримано узагальнену постановку задачі синтезу нейромережевої системи автентифікації суб'єкта за візуальним представленням його голосу.
4. Розроблено структурну схему та UML-діаграми нейромережевої системи автентифікації суб'єкта за візуальним представленням його голосу.
5. Викладено математику процедур вейвлет-перетворення мовних сигналів, їх фільтрації та часової нормалізації тривалості їх звування, в т.ч. алгоритму нелінійної часової нормалізації.
6. У програмному середовищі MATLAB розроблено програмне забезпечення для автентифікації суб'єкта за візуальним представленням його голосу.
7. Проведено дослідження доцільності використання часової нормалізації мовних сигналів у задачі автентифікації суб'єктів за голосом та встановлено, що використання процедури нормалізації є доцільним.