

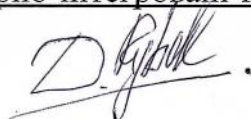
Вінницький національний технічний університет  
Факультет інтелектуальних інформаційних технологій та автоматизації  
Кафедра комп'ютерних систем управління

## МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:

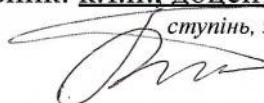
«Автоматизована система ідентифікації особи за її голосом»

Виконав: студент 2 курсу, групи 2АКІТ-22м  
спеціальності 151 – Автоматизація та  
комп'ютерно-інтегровані технології



Дмитро РИБАК  
Ім'я ПРІЗВИЩЕ

Керівник: к.т.н., доцент, проф. кафедри КСУ  
*ступінь, звання, посада*



Микола БИКОВ  
Ім'я ПРІЗВИЩЕ

«01» грудня 2023 р.

Опонент: к.т.н., доцент, доцент каф. АІТ  
*ступінь, звання, посада*

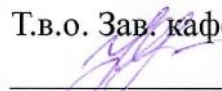


Юрій ІВАНОВ  
Ім'я ПРІЗВИЩЕ

«05» грудня 2023 р.

Допущено до захисту

Т.в.о. Зав. кафедри КСУ

 Марія ЮХИМЧУК

«07» грудня 2023

Вінницький національний технічний університет  
 Факультет інтелектуальних інформаційних технологій та автоматизації  
 Кафедра комп'ютерних систем управління  
 Рівень вищої освіти другий (магістерський)  
 Галузь знань – 15 Автоматика та приладобудування  
 Спеціальність - 151 Автоматизація та комп'ютерно-інтегровані технології  
 Освітньо-професійна програма - Інтелектуальні комп'ютерні системи

**ЗАТВЕРДЖУЮ**

**Т.в.о. Зав. кафедри КСУ**

**Марія ЮХИМЧУК**

**«09» жовтня 2023 року**

## **ЗАВДАННЯ**

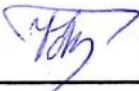
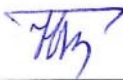
### **НА МАГІСТЕРСЬКУ КВАЛІФІКАЦІЙНУ РОБОТУ**

**студенту** Рибаку Дмитру Валерійовичу

(прізвище, ім'я, по батькові)

1. Тема роботи. «Автоматизована система ідентифікації особи за її голосом»  
керівник роботи Биков Микола Максимович, к.т.н., доцент  
затверджені наказом ВНТУ від “18” вересня 2023 року № 247
2. Термін подання студентом роботи “1” грудня 2023 року
3. Вихідні дані до роботи:  
Кількість мовців – <127; вибірка парольних фраз мовців у вигляді wav-файлів;  
досліджувані типи класифікаторів – апаратно-програмні;
4. Зміст текстової частини:  
Вступ; аналіз стану проблеми розпізнавання голосу мовця, дослідження  
методів і моделей опису і аналізу мовного сигналу мовців; розробка  
алгоритмів  
і програмного забезпечення ідентифікації особи за її голосом; машинний  
експеримент; економічна частина.
5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень):  
Математичне обґрунтування способу розпізнавання мовця – плакат; схема  
структурна підсистеми ідентифікації особи за голосом – 1 аркуш; 2 схеми  
програм обробки мовного сигналу – 3 аркуші; схема функціональна  
підсистеми – 1 аркуш; схема алгоритму ідентифікації – 1 аркуш; результати  
експерименту – плакат;


## 1. Консультанти розділів роботи

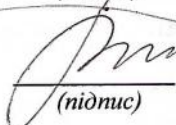
Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Виконання прийняв
4	Буреннікова Н.В. д.е.н., професор кафедри ЕПВМ		

## 2. Дата видачі завдання «09» жовтня 2023 року

## КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва та зміст етапу	Термін виконання етапів роботи	Примітка
1	Варіантний аналіз об'єкта автоматизації	02.10.2023	
2	Постановка задачі і розробка технічного завдання	04.10.2023	
3	Математичне обґрунтування вибраного методу ознакового опису голосу мовця	18.10.2023	
4	Розробка алгоритмічного та програмного забезпечення системи	5.11.2023	
5	Машинний експеримент	12.11.2023	
6	Оформлення пояснювальної записки, графічного матеріалу і презентації	20.11.2023	
7	Апробація результатів дослідження	22.11.2023	
8	Публікації	22.11.2023	
9	Графічні матеріали: - Плакати демонстраційні - Схеми алгоритмів і програм - Схеми системи - UML діаграми - Результати машинного експерименту	23.11.2023 23.11.2023 23.11.2023 26.11.2023 30.11.2023	
10	Захист МКР	20.12.2023	

Студент  Дмитро РИБАК  
(підпис) Ім'я ПРИЗВИЩЕ

Керівник роботи  Микола БИКОВ  
(підпис) Ім'я ПРИЗВИЩЕ

## АНОТАЦІЯ

УДК 681.3.019

Рибак Д. В. Автоматизована система ідентифікації особи за її голосом. Магістерська кваліфікаційна робота зі спеціальності 151 – Автоматизація та комп'ютерно-інтегровані технології, освітня програма – Інтелектуальні комп'ютерні системи. Вінниця: ВНТУ, 2023. 126 с.

Українською мовою. Бібліогр.: 53 назв; рис.: 13 ; табл. 16 .

Магістерська кваліфікаційна робота присвячена розробці методів, алгоритмів і програмного забезпечення автоматичної ідентифікації особи мовця за її голосом в системах санкціонованого доступу до інформаційних ресурсів і безпекових системах.

Проведено огляд математичних основ; розроблено алгоритми і програми аналізу сигналів мовлення; розроблено алгоритми і програми опису диктора такими ознаками, як частота основного тону та інтегральний спектр; розроблено нейромережевий алгоритм класифікації диктора за його ознаковим описом; проведені експериментальні дослідження з визначення якості розпізнавання диктора за допомогою розробленого програмного забезпечення.

Розробка дозволить покращити практичне застосування систем голосового доступу в комп'ютерних мережах.

В економічній частині проведено техніко-економічне обґрунтування доцільності та ефективності впровадження результатів розробки в реальні проекти.

Ілюстративна частина роботи містить 12 плакатів і креслень з ілюстрацією результатів роботи.

Ключові слова: мовлення, сигнал мови, темп мовлення, автоматизація контролю, комунікаційний центр, критична система.

## ANNOTATION

UDC 681.3.019

Dmytro Rybak. Automated system of identification of a person by his voice. Master's thesis on specialty 151 – “Automation and computer-integrated technologies”, educational program – “Intelligent computer systems”. Vinnytsia: VNTU, 2023. 126 p.

In Ukrainian. Bibliography: 53 titles; Fig.: 13; table 16.

The master's thesis is devoted to the development of methods, algorithms, and software for the automatic identification of the speaker's identity by their voice in systems of authorized access to information resources and security systems.

Mathematical foundations were examined, and algorithms and programs for speech signal analysis were developed. Algorithms and programs for describing the announcer using features such as the fundamental tone frequency and integral spectrum were developed, and a neural network algorithm for classifying the announcer based on its feature description was developed. Experimental studies were conducted to determine the quality of speaker recognition using the developed software.

The development will improve the practical application of voice access systems in computer networks.

In the economic section, a technical and economic justification of the feasibility and effectiveness of implementing the development results into real projects was carried out.

The illustrative part of the work contains 12 posters and drawings illustrating the results of the research.

Key words: speech, speech signal, speech rate, control automation, communication center, critical system.

## ЗМІСТ

<b>ВСТУП.....</b>	<b>8</b>
<b>1 АНАЛІЗ СТАНУ ПРОБЛЕМИ ІДЕНТИФІКАЦІЇ ОСОБИ ЗА ЇЇ ГОЛОСОМ.....</b>	<b>11</b>
1.1 Суть технічної проблеми ідентифікації голосу особи.....	11
1.2 Аналіз існуючих систем ідентифікації особи за голосом.....	14
1.3 Варіантний аналіз відомих систем ідентифікації мовця.....	20
1.4 Інтелектуальні системи ідентифікації мовця за його голосом.....	25
1.5 Принципи побудови сучасних систем ідентифікації особи.....	30
1.6 Обґрунтування методу ідентифікації особи і постановка задачі.....	31
<b>2 РОЗРОБКА МЕТОДІВ ОПИСУ ТА ІДЕНТИФІКАЦІЇ ОСОБИ ЗА ЇЇ ГОЛОСОМ.....</b>	<b>33</b>
2.1 Характеристика методів параметризації сигналу мовця.....	33
2.2 Аналіз способів відокремлення інформативних ознак з сигналу мовлення.....	35
2.3 Обчислення частоти основного тону за кількістю нульових перетинів.....	46
2.4 Висновки до розділу.....	48
<b>3 РОЗРОБКА АЛГОРИТМІВ І ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ СИСТЕМИ ІДЕНТИФІКАЦІЇ ОСОБИ.....</b>	<b>49</b>
3.1 Розробка схеми взаємодії програм для ідентифікації особи за її голосом.....	49
3.2 Розробка алгоритму спектрального аналізу сигналу мовця.....	50
3.3 Розробка алгоритму виділення частоти основного тону.....	53
3.4 Розробка алгоритмів навчання класифікатора для розпізнавання диктора.....	61
3.5 Машинний експеримент.....	70

3.6 Висновки до розділу.....	73
<b>4 ЕКОНОМІЧНА ЧАСТИНА.....</b>	<b>75</b>
4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки.....	76
4.2 Розрахунок узагальненого коефіцієнта якості розробки.....	79
4.3 Розрахунок витрат на проведення науково-дослідної роботи.....	81
4.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором.....	93
4.5 Висновки до розділу.....	98
<b>ВИСНОВКИ.....</b>	<b>99</b>
<b>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....</b>	<b>100</b>
<b>ДОДАТКИ.....</b>	<b>106</b>
<b>Додаток А – (обов’язковий) Протокол перевірки кваліфікаційної роботи на наявність текстових запозичень.....</b>	<b>107</b>
<b>Додаток Б – (обов’язковий) Технічне завдання.....</b>	<b>108</b>
<b>Додаток В – (довідниковий) Лістинги програм.....</b>	<b>111</b>
<b>Додаток Г – (обов’язковий) Ілюстративна частина.....</b>	<b>116</b>

## ВСТУП

**Актуальність теми.** Інтенсивна інформатизація сучасного суспільства призвела до виникнення великої кількості інформаційних об'єктів і сховищ, наприклад, хмарних, які потребують захисту від несанкціонованого доступу сторонніх осіб чи зловмисників. Велика кількість інформації з'являється і наростає з часом на сайтах мережі Інтернет, які обслуговують потреби функціонування комп'ютерно-інтегрованих систем виробництва. Ця інформація призначена для задоволення внутрішніх інформаційних потреб користувачів та визначених підрядників. З метою забезпечення стабільності їх роботи і запобігання несанкціонованому розголошенню технічної та комерційної інформації, необхідно встановити систему обмеження доступу до неї. В зв'язку з цим все більш актуальними стають проблеми використання надійних і зручних механізмів санкціонування доступу окремих осіб до даних інформаційних ресурсів, одним з яких на сьогодні вважається використання мовних технологій.

Значна частина сучасних засобів захисту інформації, що базуються на використанні деякого фізичного ключа, коду чи пароля, не є достатньо надійними, оскільки їх можна загубити, підібрати, чи під силовим примусом передати зловмиснику. Одним із поширених на сьогодні методів запобігання таким випадкам є використання в якості ключа біометричних характеристик індивідуальності людини: відбитків пальців, особливостей роговиці ока і, що досить зручно, індивідуальних ознак голосу особи-мовця. Ця зручність полягає в тому, що сигнал мовлення особи, що розпізнається, можна передавати на великі відстані. Проте, сьогодні системи ідентифікації диктора використовуються лише в певних сферах суспільного життя, так і не набувши широкого поширення. Це пояснюється тим, що існуючі системи розпізнавання диктора не володіють достатньою надійністю, легкістю в навчанні, зручністю в користуванні або доступністю. Зазвичай вони



використовуються як додаткові методи ідентифікації особи там, де потрібно високий рівень впевненості в ідентифікації. Таким чином, на сьогодні пріоритетом є дослідження, спрямовані на удосконалення методів та алгоритмів обробки мовних сигналів з метою створення ефективних механізмів автоматичної ідентифікації особи за голосом. Таким чином, завдання розробки автоматизованої системи ідентифікації особи за її голосом, яке вирішується у цій магістерській роботі, залишається високо актуальним.

**Метою роботи** є підвищення ефективності використання інформаційних ресурсів на сучасних автоматизованих виробництвах за рахунок розробки автоматизованих системи ідентифікацій особи за її голосом.

Для досягнення визначеної мети у цій роботі вирішуються наступні завдання:

- Розглянути та проаналізувати сучасні засоби санкціонування доступу персоналу до інформаційних ресурсів і об'єктів обмеженого використання;
- проаналізувати сучасні методи обробки сигналів мовлення з метою вибору оптимальних для задачі автоматичної ідентифікації особи за її голосом;
- обґрунтувати вибір оптимального класифікатора для використаних ознакових описів голосу особи;
- розробити алгоритми обробки сигналу мовлення і програмне забезпечення і програмного забезпечення автоматизованої системи ідентифікації особи.

**Об'єктом дослідження** є процеси обробки і аналізу сигналу мовлення особи, які характеризують її індивідуальність.

**Предметом дослідження** є методи і алгоритми автоматичної ідентифікації особи за її мовленнєвим сигналом.

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

Робота виконувалась у межах науково-дослідної тематики на кафедрі КСУ.

**Методи дослідження.** У процесі дослідження застосовувалися: теорія цифрової обробки сигналів мовлення; теорія ймовірностей; методи дискретний аналіз і теорія автоматів; теорія алгоритмів; системний аналіз; комп'ютерне моделювання для проведення машинних експериментів.

**Наукова новизна.** Представлено метод виділення частоти основного тону мовця шляхом підрахунку кількості нульових перетинів кліпованого сигналу мовлення.

**Практична цінність.** Розроблені алгоритми і програмні засоби дозволяють підвищити надійність ідентифікації особи за її голосом, що дає змогу підвищити інформаційну безпеку на комп'ютерно-інтегрованих автоматизованих виробництвах та інших інформаційних ресурсах.

**Особистий внесок здобувача.** Магістрантом розроблено алгоритми обробки сигналу мовлення і структура програмного забезпечення автоматизованої системи ідентифікації особи.

**Впровадження.** Отримані в магістерській роботі результати застосовуються в процесі викладання предметів, пов'язаних з інтелектуальними технологіями, і також в наукових дослідженнях на кафедрі Комп'ютерних Систем Управління.

**Апробація результатів роботи.** Отримані результати та висновки з проведених у роботі досліджень були представлені на Міжнародній науково-практичній Інтернет-конференції для студентів, аспірантів та молодих науковців підрозділів Вінницького національного технічного університету «МН» (2023).

**Публікації.** За результатами конференції опубліковані тези: Д.В. Рибак, М.М. Биков. Автоматизована система ідентифікації особи за її голосом. – [Електронний ресурс]. Режим доступу: <https://conferences.vntu.edu.ua/index.php/mn/mn2024/author/submission/19786>

# 1 АНАЛІЗ СТАНУ ПРОБЛЕМИ ІДЕНТИФІКАЦІЇ ОСОБИ ЗА ЇЇ ГОЛОСОМ

## 1.1 Суть технічної проблеми ідентифікації голосу особи

Сучасне функціонування різних сфер людського суспільства характеризується широким впровадженням методів штучного інтелекту (ШІ) в таких галузях, як військова справа, медицина, транспорт наземний, повітряний і водний, робототехніка, системи управління і багато інших [1]. Важче назвати сфери діяльності людини, де ШІ сьогодні не використовується, ніж ті, де він застосовується. Одним із важливих напрямків розвитку методів ШІ є мовні технології, пов'язані з задачами обробки природної мови (NLP – Natural Language Processing). Зусилля мовних технологій спрямовані на вирішення трьох основних задач – розпізнавання мови за сигналом мовлення, розпізнавання особи за її голосом та синтезу мови з тексту.

Розглядання аспектів ідентифікації особи за голосовими параметрами є значущим напрямком у сучасному світі, зокрема в контексті кібербезпеки, аутентифікації та приватності. Актуальність цієї проблеми викликана необхідністю впровадження безпекових заходів для охорони інформаційних ресурсів, які забезпечує надійне та ефективне функціонування багатьох організаційних і промислових систем [2]. Одним із підходів до розв'язання вказаної проблеми є автоматизація розмежування доступу різним особам до різних масивів конфіденційної інформації і секретних об'єктів. Це розмежування полягає в тому, що доступ дозволяється тільки тим особам, які мають допуск до відповідних об'єктів. Використання в цьому випадку такої біометричної характеристики індивідуальності людини, як її голос, особливо важливе і зручне в сучасний період, який характерний онлайновим принципом роботи на Інтернет ресурсах, в хмарних інформаційних середовищах, в організаціях, які займаються розробкою програмних

продуктів, а також для допуску до управління різними кіберфізичними об'єктами. Цей метод санкціонування допуску може бути вирішений тільки за рахунок розробки і впровадження автоматизованих систем розпізнавання мовця за його голосом.

Методи розпізнавання особи за її голосом поділяються на два класи:

1. ідентифікація особи;
2. верифікація особи.

Зазвичай більше уваги приділяється рішенням задач першого класу, оскільки задача верифікації (перевірки і підтвердження особистості мовця може бути вирішена тільки тоді, коли попередньо були проведені роботи з розпізнавання її голосу і формування його еталону в пам'яті системи розпізнавання).

Голос людини представляє собою одну з широко поширених біометричних характеристик, яка включає у себе унікальний набір особливостей, що можуть бути відносно легко виміряні (наприклад, частотний спектр голосового сигналу). Серед переваг голосової ідентифікації слід відзначити також зручність використання та застосування, а також доступність пристроїв, використовуваних для проведення ідентифікації, наприклад, мікрофонів. За допомогою цього методу можливо ефективно визначати особистість за голосовими характеристиками, що є необхідним в аспектах безпеки та ідентифікації осіб.

Можливості ідентифікації особи за голосовими даними охоплюють широкий спектр завдань, що робить їх відмінними серед інших біометричних систем. Важливо зауважити, що голосова ідентифікація давно використовується в різних системах обмеженого доступу до фізичних об'єктів та інформаційних ресурсів [3-6]. Цей метод також входить у склад теорії мовоутворення [7,8].

Додатково, голосова ідентифікація стає перспективною у нових застосуваннях, зокрема в системах, що базуються на телекомунікаційних каналах зв'язку. Наприклад, у сфері мобільного зв'язку голос може використовуватися для управління послугами, а використання голосової ідентифікації сприяє підвищенню рівня захисту від шахрайства [9,10].

Незважаючи на широку застосовуваність і перераховані вище переваги, методи ідентифікації особи за голосовою інформацією мають низку серйозних недоліків. До них відносяться, в першу чергу, невисока роздільна здатність методів та значний відсоток помилок як першого роду (хибно відкинуті особи, які мають право на допуск), і найбільш небезпечного другого роду (хибно допускаються до конфіденційної інформації особи, які права на допуск до неї не мають). Особливо ускладнює ситуацію проведення ідентифікації у реальних умовах, що супроводжуються набором несприятливих зовнішніх акторів.

Ідентифікація особи за голосом, що проводиться в реальних умовах, зустрічається з такими серйозними труднощами. По-перше, при даному способі ідентифікації виникають всілякі апаратні спотворення та перешкоди [11], обумовлені особливостями апаратури та пристроїв для запису, обробки та зберігання інформації. По-друге, на голосовий сигнал накладаються зовнішні акустичні шуми, які можуть суттєво спотворювати індивідуальні інформативні характеристики. В силу цього системи ідентифікації, які демонструють високу ефективність у лабораторних умовах, під час аналізу сигналу мовлення з зовнішніми шумами можуть показати надійність, значно нижчу [12]. Нарешті, у низці завдань доводиться проводити ідентифікацію в дуже складних умовах накладання голосів кількох дикторів, у тому числі і з близькими акустичними характеристиками. Потрібно відзначити, що дослідження можливостей голосової ідентифікації для цього найбільш складного випадку практично не проводилося.

Проведення голосової ідентифікації включає комплекс технічних, алгоритмічних та математичних методів, що охоплюють усі етапи, починаючи з запису голосу та закінчуючи класифікацією голосових даних.

Розглянуті труднощі та недоліки якраз і складають суть технічної проблеми в побудові систем ідентифікації особи за її голосом. Аналіз цієї проблеми призводять до висновку, що подальший розвиток систем голосової ідентифікації вимагає розробки нових підходів, спрямованих на обробку великих масивів експериментальних акустичних сигналів, їх ефективний аналіз та надійну класифікацію. Це свідчить про актуальність досліджень зі створення покращених методів обробки, аналізу та класифікації голосових даних, що забезпечують надійність та достовірність ідентифікації особи.

В даній магістерській роботі пропонується вирішити дану проблему шляхом більш надійного способу виділення частоти основного тону з сигналу мовця і використанням класифікаторів на нейронній мережі.

## 1.2 Аналіз існуючих систем ідентифікації особи за голосом

Задача ідентифікації особи людини за її голосом була сформульована і поставлена для вирішення на порядок денний більше 55 років тому. На сьогоднішній день ідентифікація мовця за його голосом є активною областю досліджень, і багато робіт було проведено в цьому напрямку. Нижче наведено перелік тільки кількох важливих досліджень та робіт, результати яких на сьогодні є відомими і широко застосовуються.

Голосові помічники. Дослідження та розвиток у сфері голосових технологій, представлених, наприклад, у голосових асистентах, таких як Siri від Apple, Alexa від Amazon та Google Assistant. Основний акцент робиться на технологіях ідентифікації мовця для розрізнення голосових команд від різних користувачів [9].

Біометрична аутентифікація в фінансовому секторі: Використання ідентифікації за голосом для аутентифікації користувачів під час фінансових операцій. Особливий фокус досліджень в цьому напрямку – забезпечення безпеки використання таких систем.

Медичні застосування: Використання ідентифікації за голосом для діагностики медичних станів, таких як захворювання дихальних шляхів або психічні розлади.

Системи безпеки та контролю доступу: Голосова ідентифікація використовується для забезпечення контролю доступу до будівель, комп'ютерних систем та інфраструктури.

Розпізнавання голосу в масштабі великих даних: Дослідження над розпізнаванням голосу у великих наборах даних, використовуючи машинне навчання та нейронні мережі для підвищення точності.

Антифрод системи: Роботи у напрямку виявлення шахраїв та уникнення підробки голосу в системах ідентифікації за голосом.

Мовні технології: Використання технологій, що включають не лише розпізнавання голосу, але й аналіз мови та вимови для ідентифікації мовця.

Особиста приватність. Особлива увага приділяється етичним аспектам збору та зберігання голосових даних користувачів.

До перелічених вище можна також віднести ще низку важливих напрямків застосування ідентифікації особи за її голосом, як криміналістика і судова експертиза; електронна комерція; телекомунікації; Інтернет-сервіси.

Це лише кілька напрямків досліджень у сфері ідентифікації мовця за голосом. З інноваціями в галузях, пов'язаних з штучним інтелектом та обробкою природної мови, ця область може продовжувати розвиватися і знаходити нові застосування. Дослідження в даному напрямку продовжуються з постійною увагою, і на сьогоднішній час [3,8,10] спостерігається покращення ефективності систем розпізнавання голосової інформації. Проте важлива проблема автоматичної ідентифікації особи за її

голосом в повній мірі залишається невирішеною. Таким чином, пошук інноваційних рішень у цьому конкретному напрямку є актуальним завданням на сучасному етапі. Кожна людина має унікальні голосові характеристики, що обумовлені індивідуальною будовою голосового апарату. Під час слухання бесіди людина може підсвідомо розпізнавати голоси інших осіб, але розробка автоматичного класифікатора мовців стикається зі значними труднощами [3].

Метою завдання розпізнавання особи за голосом є вибір та класифікація людської мови з вхідного аудіо потоку, часто розв'язуючи підзадачі ідентифікації або верифікації мовця. Вибір методу для обчислення ступеня подібності між вхідними сигналами та опорними зразками є ключовим етапом вирішення цих завдань. Використання різних відстаней або ймовірнісних критеріїв дозволяє обчислити ступінь подібності між еталоном і тестовим зразком [13]. Алгоритм ідентифікації мовця може бути описаний як залежний чи незалежний від тексту. У випадку, коли алгоритм залежить від тексту, він може використовувати фіксовані або згенеровані системою фрази для ідентифікації. Незалежні від тексту системи призначені для обробки мов будь-якої природи [14].

Розв'язання завдання розпізнавання образів полягає в присвоєнні об'єктів, представлених у числовому чи символічному вигляді, одній з відомих груп відповідно до заданої міри схожості. Взаємозв'язок еквівалентності визначає належність оцінюваних об'єктів до конкретних категорій, розглядуваних як незалежні семантичні одиниці.

Під час створення алгоритмів розпізнавання, дослідник може визначати класи еквівалентності, використовуючи власні оригінальні концепції або залучаючи зовнішню додаткову інформацію щодо подібності та відмінності об'єктів. У випадках, коли система розпізнавання здійснює класифікацію за участю маркованої навчальної вибірки, це називається "контрольованим розпізнаванням або розпізнаванням з учителем" [8,9]. Якщо система виконує класифікацію без використання маркованої навчальної вибірки, то це відомо



як "неконтрольоване розпізнавання з навчанням без учителя". Алгоритми розпізнавання образів без учителя вимагають значних обчислювальних ресурсів або використання передових технологій штучного інтелекту.

Розрізняють два класи задач автоматичного розпізнавання диктора за його голосу – верифікація та ідентифікація, що показано на рис. 1.1.

У процесі перевірки диктора головною метою системи розпізнавання є визначити, чи особа, яка вимовляє парольну фразу, є тією самою особою, яку система навчилася розпізнавати. У цьому контексті ключовою є оцінка рівня ризику, яка визначає поріг прийняття рішення. Величина цього порогу визначає, чи розпізнана особа вважається собою чи ні [10,11].

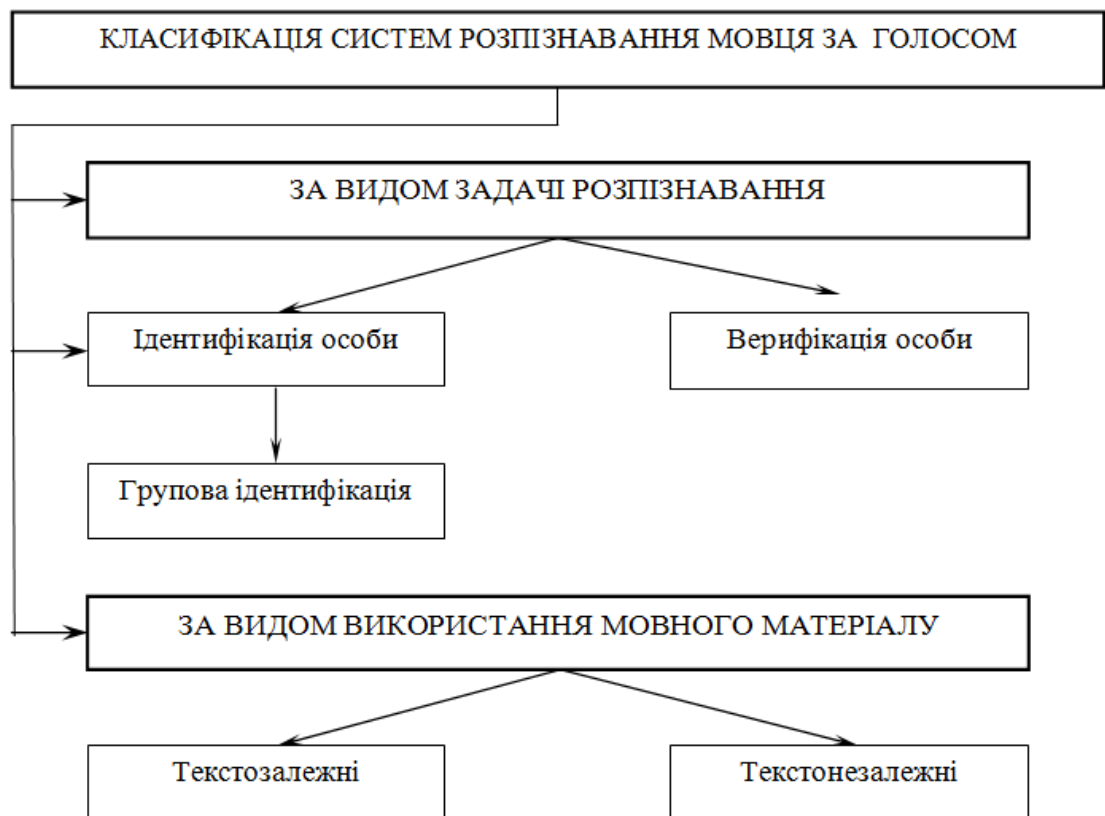


Рисунок 1.1 – Класифікація систем розпізнавання диктора

Схема системи верифікації диктора представлена на рис. 1.2.

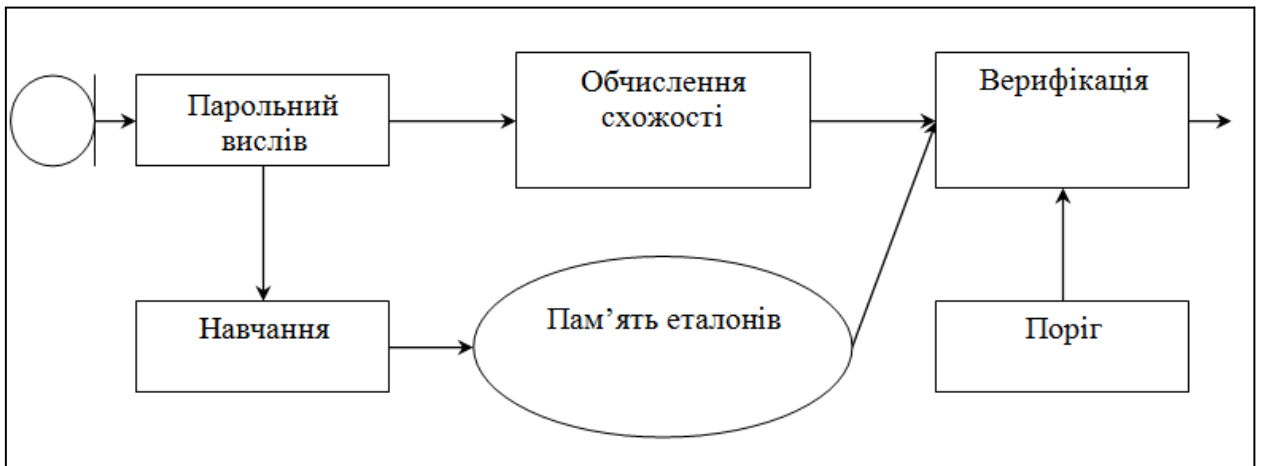


Рисунок 1.2 – Схема конфігурації системи автоматичної перевірки диктора.

Ефективність автоматичної системи верифікації особи мовця визначається двома ключовими параметрами: ймовірністю помилки 1-го роду (ситуація, коли зареєстрованого користувача помилково визнається зломисником) та ймовірністю помилки 2-го роду (ситуація, коли система помилково відмовляється визнавати зареєстрованого користувача).

При потребі під час експлуатації системи верифікації диктора можна внести зміни таким чином, щоб регулювати співвідношення помилок першого та другого роду. Це означає, що можна зменшувати кількість помилок одного типу за рахунок іншого, при цьому утримуючи всі інші важливі фактори, такі як тривалість мовного повідомлення, його характер, шуми і т.д. Визначення співвідношення помилок першого та другого роду встановлюється як прийнятий критерій якості в прийнятті рішень.

Розпізнавання особи мовця за допомогою його висловлювань визначається як ідентифікація диктора [12,13]. У деяких випадках цей процес здійснюється за допомогою конкретного парольного слова або фрази, а в інших використовується достатньо тривалий фрагмент мовлення будь-якого змісту. Структурна схема системи автоматичної ідентифікації диктора представлена на рис. 1.3.

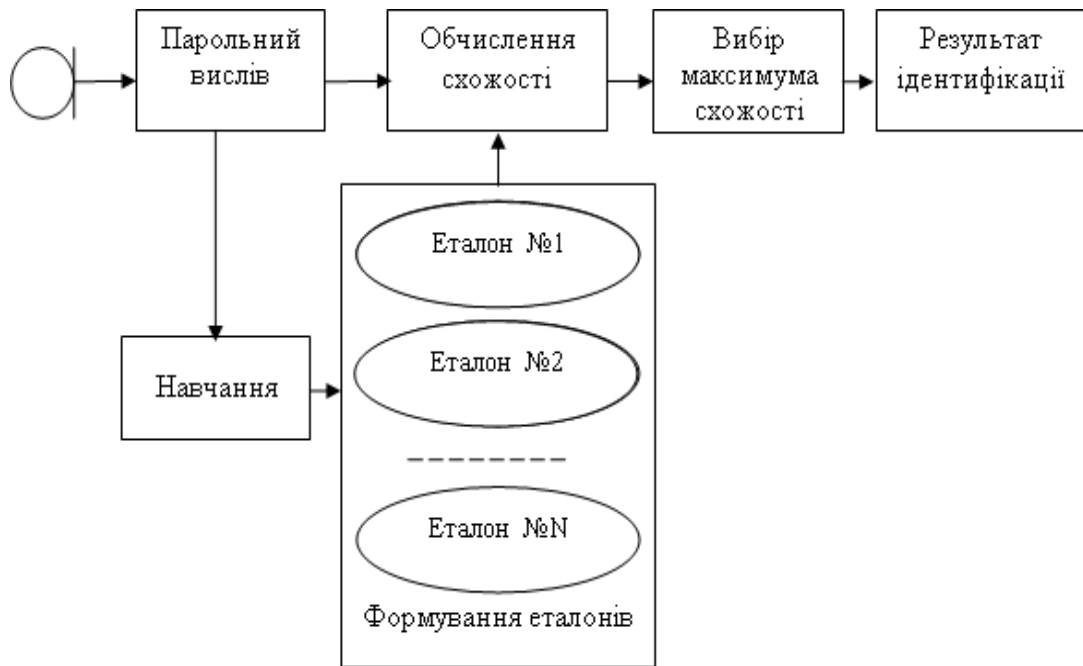


Рисунок 1.3 – Структурна схема системи ідентифікації мовця

Перед введенням системи ідентифікації мовця в експлуатаційний режим проводиться її навчання. На цьому етапі записи парольних фраз кожного мовця, який підлягає ідентифікації, подаються на вхід. З цих записів використовуються обрані алгоритми для виділення інформативних ознак, які формують еталонний опис мовця у вигляді вектора ознак в пам'яті еталонів. Щоб мінімізувати випадки помилок першого і другого роду, еталонні вектори ознак повинні враховувати всі можливі стани мовця. Тому кожен диктор, який бере участь у реєстраційній процедурі для ідентифікації, повинен вимовити парольну фразу декілька разів для чіткого визначення границь свого кластера в просторі ознак.

Під час експлуатації системи розпізнавання на вхід подається запис парольної фрази одного з зареєстрованих мовців для ідентифікації своєї особи. Інформативні ознаки виділяються з запису голосу, і створюється його образ-вектор в просторі ознак. Обчислюється відстань від цього вектора до кожного еталону в пам'яті системи, і згідно з прийнятим критерієм максимальної схожості приймається рішення щодо ідентифікації мовця з одним із дикторів, які були використані під час навчання системи.

Зазвичай точність функціонування системи визначається ймовірністю виникнення помилок 1-го та 2-го роду протягом заданого періоду експлуатації. Усі вищезгадані види завдань розпізнавання можна класифікувати на дві категорії в залежності від використовуваного мовного матеріалу: текстозалежні та текстонезалежні. Відмінність між цими двома категоріями полягає в тому, що в першому випадку система навчається на конкретних парольних виразах певної тривалості. У текстозалежних системах ідентифікації збіг лінгвістичної форми двох порівнюваних мовних записів сприяє полегшенню процесу розпізнавання, тому їх схемна та алгоритмічна структура вважається більш простою.

У сучасності великою популярністю користуються різні методи для опису особливостей мовця [14]. До ознакового опису мовця для текстозалежних систем відносять динамічне перетворення часу (Dynamic Time Warping; DTW) та приховані марковські моделі (Hidden Markov Model; HMM). Для текстонезалежних систем застосовуються такі методи, як векторне квантування (Vector Quantification; VQ), моделі гаусових змішаних моделей (GMM) і метод опорних векторів (SVM).

### 1.3 Варіантний аналіз відомих систем ідентифікації мовця

Для розгляду різних методів ідентифікації мовця розглянемо кілька конкретних систем, що були розроблені на сучасний момент. У всіх системах автоматичної ідентифікації особи за особливостями голосу відбувається порівняння записів голосу невідомого мовця з записами (моделями) голосу відомого мовця, які зареєстровані в системі. Це дозволяє визначити ступінь схожості голосу невідомого мовця з голосом користувача системи і визначити, чи належать записи мови різним людям чи тій самій людині.

### 1.3.1 Система автоматичної ідентифікації персоналу компанії Texas Instruments

Система, створена Доддінгтоном для автоматизації контрольного пункту у фірмі Texas Instruments, використовувала текстозалежний підхід, використовуючи записи парольних фраз із чотирьох слів як мовний сигнал. Обробка мовного сигналу відбувалася за наступною схемою (див. рис. 1.4). Процес розпізнавання включав постійне порівняння спектральних фрагментів вхідного мовного матеріалу, отриманого комп'ютером, із контрольним набором векторів, а також визначенням найбільш ймовірних опорних точок парольної фрази [15].

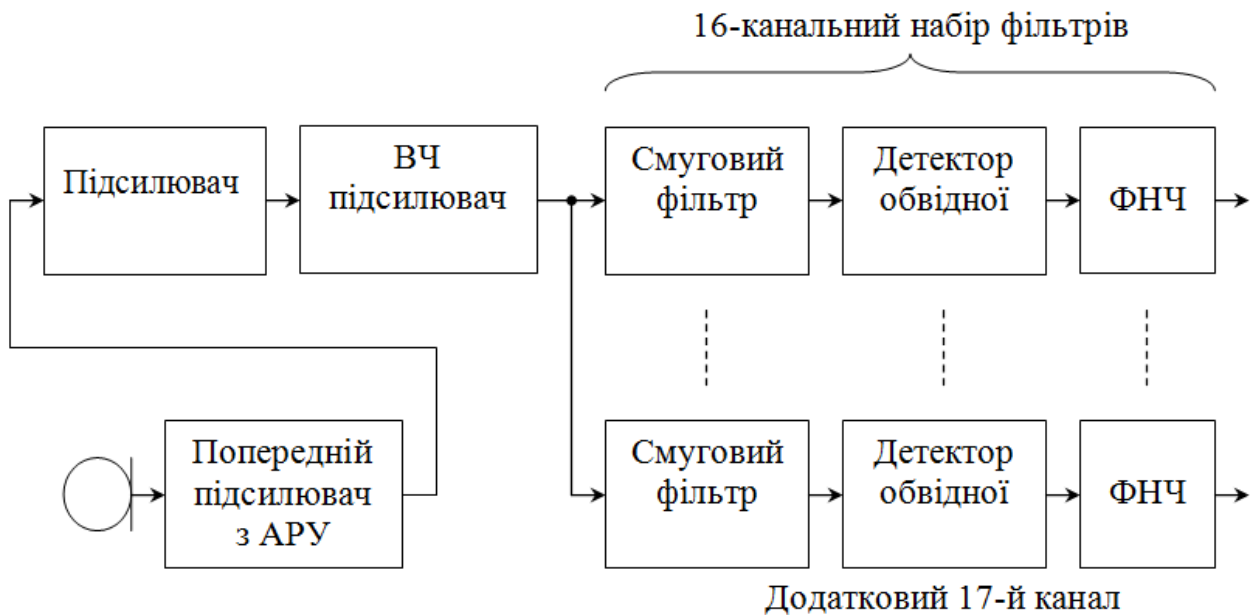


Рисунок 1.4 – Структурна схема системи ідентифікації особи за її голосом фірми Texas Instruments

Процес прийняття рішення включав у себе пошук мінімуму квадрату відхилення між спектральними частинами опорних точок парольної фрази та відповідними опорними векторами. Час, необхідний для розпізнавання, складав 5,8 секунд.

Ця система була встановлена на блокпосту обчислювального центру компанії Texas Instruments. Згідно з висновками, представленими у [16], протягом одного року 180 користувачів звернулися до системи 150 000 разів для отримання допуску. Імовірність помилок 1-го виду становила 1%, 2-го виду – 0,3% (у випадку використання "тренажерів" – 0,6%). Однак недоліками цієї системи є обмежена надійність при відношенні сигнал/шум більше 30 дБ і значний вплив каналів зв'язку на точність її функціонування.

### 1.3.2 Система ідентифікації особи, запропонована компанію Bell Laboratories

Система ґрунтується на концепціях, визначених Доддінгтоном у попередній роботі та подальше удосконалення, яке внесли Ламміс, Сембур та Розенберг [17]. Варто відзначити, що у цій системі використовуються параметри мовного сигналу для ідентифікації мовця, що залишається стійким при високому відношенні сигнал/шум, і передача мовних сигналів до комп'ютерів реалізується через телефонні канали. В ролі параметрів мовних сигналів в цій системі обрано плавно змінюючіся траєкторії основного тону голосів мовців, їх інтенсивності, формантні частоти і коефіцієнти лінійного прогнозування.

В рамках цієї системи мовний сигнал піддається аналізу за допомогою методу швидкого перетворення Фур'є (ШПФ), що призводить до отримання трьох формантних частот, частоти основного тону та інтенсивності. Для підвищення продуктивності системи використовується метод паралельної обробки сигналу в просторі-часі [17]. Інтенсивність сигналу визначається через фільтр низьких частот із смугою пропускання до 900 Гц, випрямляча та плавного згладжування. Коефіцієнти лінійного прогнозування знаходяться в діапазоні частот від 300 до 4000 Гц за методом Атала та Ханауера [17]. Схема

послідовності виконання операцій при автоматичному розпізнаванні мовця наведена на рис. 1.5.



Рисунок 1.5 – Структурна схема системи фірми Bell Laboratories

Система має два входи, один з яких використовується для введення PIN-коду для виклику відповідного еталону з бази даних, а другий для введення паролльної фрази.

Після введення паролльної фрази проводиться її аналіз з виділенням параметрів, наведених вище, потім проходить нормалізація тривалості звучання паролльної фрази методом нелінійного стиснення та розтягнення [7] для кращого співставлення з еталоном.

Порівняння контурів інтенсивності дає можливість створити деформуєчу функцію, яка використовується для співставлення і інших контурів. Обчислена по всім контурам розбіжність порівнюється з пороговим значенням і використовується для прийняття рішень. Дослідження

ефективності використовуваних показників показало, що використання коефіцієнтів прогнозування (початково розглядали 12 коефіцієнтів, але найбільш інформативними виявилися  $a_2, a_7, a_{12}$ ) разом із даними про чотири формантні частоти та їх ширину вказує на більшу ефективність перших. Однак, навіть при зменшенні кількості показників виявлено, що час, потрібний для прийняття рішення системою, у 11 разів перевищує реальний.

Тестування системи виявило, що введення додаткових показників (коефіцієнтів прогнозування) та використання додаткової паролльної фрази впливає на зниження ймовірності виникнення помилок 1-го та 2-го роду під час експлуатації системи. Однак це також призводить до збільшення часу, необхідного для прийняття рішення системою, що для даної системи вже є досить великим.

### 1.3.3 Автоматизована система розпізнавання голосу САРГ

В основі відбору характеристик індивідуальності голосу в даній системі лежить припущення, що виразні особливості мовлення обумовлені як анатомічними особливостями мовного апарату людини, так і динамікою його функціонування [18]. Ця підстава вплинула на те, як сигнал мовця описано, включаючи подання мовного сигналу з використанням його "стаціонарних" характеристик та врахування його динамічних властивостей. Структурно, САРГ [18] є змішаною системою, яка складається з аналогової та цифрової частин. Аналогова частина включає блок для нормалізації рівня мовного сигналу, аналізатора та аналого-цифрового перетворювача (АЦП), який встановлює зв'язок між комп'ютерами. Спектрально-смуговий аналіз мови проводиться за допомогою 24 чверть-октавних фільтрів, що охоплюють діапазон від 100 до 6400 Гц. Сигнали, що надходять на виході аналізатора, піддаються корекції, згладжуються фільтром низьких частот і подаються разом із хвилюючим мовного сигналу та значенням основної частоти на



АЦП. Квантування на виході перетворювача здійснюється як за часом ( $\approx 14$  мс), так і за амплітудою на 16 рівнях.

У тестуванні системи взяли участь 64 чоловіки-диктори (11 – “Свої”, 53 – «Чужі», включаючи імітаторів). Було зібрано 622 записи парольних фраз, з яких 236 припадали на “Свої”. Рівень фону в приміщенні, де проводились експерименти, не перевищував 45 дБ. Внаслідок цього ймовірність помилок 1-го роду становила 0,8%, а 2-го роду – 1,9%. Отже, під час створення CAPT використовувалися ознаки, які оцінюють вплив анатомічних особливостей мовного апарату людини на індивідуальність мовлення. Важливо відзначити недосконалість процедури сегментації мовних сигналів, що використовувалася в системі, що значно вплинуло на достовірність аналізу інтонації та просодії вимови.

#### 1.4 Інтелектуальні системи ідентифікації мовця за його голосом

На сьогоднішній день ідентифікація мовця за його голосом є активною областю досліджень, і багато робіт було проведено в цьому напрямку. Нижче наведено перелік тільки кількох важливих досліджень та робіт, результати яких на сьогодні є відомими і широко застосовуються.

Голосові асистенти: Роботи і дослідження в галузі голосових асистентів, таких як Siri від Apple, Alexa від Amazon, Google Assistant, використовують технології ідентифікації мовця для розрізнення голосових команд від різних користувачів.

Банківська безпека. У фінансовому секторі ідентифікація користувачів за голосом використовується для біометричної аутентифікації при здійсненні банківських операцій. Роботи в цій галузі досліджують способи підвищення безпеки таких систем.

Медичні системи. Деякі дослідження використовують ідентифікацію мовця за голосом для діагностики медичних станів, таких як хвороби дихальних шляхів або психічні розлади.

Технології безпеки та контролю доступу. Голосова ідентифікація використовується для контролю доступу до будівель, комп'ютерних систем та інфраструктури.

Розпізнавання голосу в масштабі великих даних. Дослідники також працюють над розпізнаванням голосу у великих наборах даних, використовуючи машинне навчання та нейронні мережі для покращення точності.

Антифрод системи. Роботи в цьому напрямку спрямовані на виявлення шахраїв і підробку голосу в системах ідентифікації за голосом.

Мовні технології. Технології, які враховують не лише розпізнавання голосу, а й аналіз мови та вимови, щоб ідентифікувати мовця.

Етика й приватність. Особлива увага приділяється питанням етики і приватності при зборі та зберіганні голосових даних користувачів.

Це лише декілька аспектів, які вивчаються у галузі ідентифікації мовця за голосом. З новаторськими підходами у сфері штучного інтелекту та обробки природної мови, дана область може продовжити свій розвиток та знаходити нові сфери застосування.

Розглянемо принципи роботи і можливості таких голосових програм, як Siri, Alexa, Google Assistant [9].

Siri – це розроблений компанією Apple у 2011 році віртуальний голосовий помічник для пристроїв на платформах Mac OS, iPad OS, iOS, tvOS, watchOS та audio OS. Керуванням ним управляють голосовими командами, рухами рук, а також стандартними натисканнями кнопок. Робота цього голосового асистента базується на технології NLP, що дозволяє системі керувати роботою інтегрованих цифрових пристроїв, обробляти запити та відповідати на них, та надавати рекомендації.

Apple Siri відповідає на користувацькі запитання, використовуючи вбудований генератор мови та виводить необхідну інформацію на дисплей пристрою, на якому вона працює. Цим голосовим асистентом також можна користуватися для диктування електронної пошти, SMS різного характеру, опрацювання вхідних повідомлень та листів.

Функціонал Siri розглядається як комбінація трьох основних компонентів:

- 1) Розмовний інтерфейс.
- 2) Поінформованість про контекст.
- 3) Делегування завдань.

Перший елемент використовується для визначення голосових команд користувача та розуміння їх змісту. Цей компонент є важливим для забезпечення якісного функціонування таких сервісів, і йому приділяється особлива увага розробниками. Незаперечною перевагою Siri є її здатність розпізнавати мову та відповідати на запитання в 20 різних мовах. Достатньо промовити універсальну фразу «Hey Siri» біля будь-якого пристрою Apple, включаючи "розумні" домашні пристрої, для активації асистента.

Голосовий помічник Alexa представляє собою програмний продукт, розроблений компанією Amazon і введений на ринок у 2014 році. Порівнюючи його з Siri, важливо відзначити, що продукт від Amazon успішно утримує провідні позиції на американському ринку, і його частка становить більше 60% протягом не одного року. Мова йде про смарт-динамік Amazon Echo, який виступає центральним хабом для цього голосового асистента. Крім цього, хмарна платформа підтримує різні пристрої від сторонніх виробників. Програму можна встановити на різних пристроях, таких як смартфони, планшети, смарт-телевізори і інші. Якщо Alexa не є частиною вашого пристрою за замовчуванням, ви маєте можливість завантажити її самостійно з App Store або Google Play.

У більшості сценаріїв використання сервісу його можна активувати, вимовивши фразу "Alexa, open wake up", хоча наявні пристрої, які не підтримують цю функцію, і в таких випадках його необхідно запускати за допомогою натискання кнопки. Помічник розпізнає та взаємодіє з користувачем на восьми різних мовах. Однією з важливих переваг цього особистого асистента є його універсальність.

Amazon дозволяє стороннім розробникам інтегрувати віртуального помічника Alexa у свої пристрої та керувати ними за допомогою спеціальної функції Alexa Voice Service. Ця функція дозволяє виконувати переадресацію дзвінків та реалізовувати різні інші операції. Таким чином, розробники можуть як створювати нові пристрої з підтримкою, так і інтегрувати програму в існуючі пристрої.

Ще однією значущою перевагою цього віртуального асистента є його велика кількість навичок, яка на даний момент перевищує 25 тисяч і продовжує зростати. Крім того, Amazon дає стороннім розробникам можливість самостійно створювати нові функції для цього "розумного" асистента. Завдяки цим навичкам користувачі можуть гнучко керувати різними аспектами своїх пристроїв за допомогою голосових команд, а їх створення відбувається за допомогою Alexa Skills Kit.

Серед найпопулярніших можливостей цього віртуального помічника варто виділити такі:

- Здійснення голосових команд для керування пристроями та виконання різних завдань.
- Відтворення музики та аудіокниг з різних джерел, таких як Amazon Music, Spotify, аудіокниги Audible тощо.
- Надання актуальної інформації, погоди, новин та інших корисних даних.
- Створення списків завдань, календарних подій, а також нагадувань для користувача.

- Використання штучного інтелекту для кращого розуміння запитань та команд користувача.
- Надання порад щодо здоров'я та фітнесу, запуск тренувань, облік кроків тощо.
- Проведення різних розважальних ігор та вікторин, використання навчальних навичок.
- Відповіді на різноманітні питання, пошук інформації в Інтернеті.

Програмний продукт Google Assistant був розроблений компанією Google у травні 2016 року, пізніше від конкурентів Siri і Alexa. Це дозволило їй врахувати досвід попередників та використати їхні переваги. Ця програма доступна сьогодні для установки на смартфони і планшети з операційною системою Android, а також для пристроїв керування керує «розумним» будинком Google Home. Окрім того, цей голосовий асистент сумісний з рядом інших розумних пристроїв, таких як розумні дисплеї, телевізори, інтелектуальні автомобільні системи тощо. Даний застосунок доступний практично на будь-якому пристрої з ОС версії Android 5.0 і вище. Його також можуть встановити користувачі Apple. Assistant підтримує активацію голосом, для цього потрібно сказати пробуджувальну фразу «OK Google».

За допомогою команд на 40 мовах Google Assistant може надавати можливість виконувати такі дії:

- Здійснення голосових команд для керування смарт-пристроями, виконання завдань та запуску додатків.
- Надання відповідей на різні запитання, пошук інформації в Інтернеті та оновлення про новини.
- Надання вказівок для навігації, інформації про транспорт та умови дорожнього руху.
- Створення списків нагадувань, календарних подій та планування завдань.

- Відтворення музики, подкастів та аудіокниг, управління відео на підтримуваних платформах.
- Надання порад щодо здоров'я, фітнес-тренувань та стеження за активністю.
- Взаємодія та інтеграція з різними додатками, включаючи електронну пошту, календарі, картки лояльності та інші.

### 1.5 Принципи побудови сучасних систем ідентифікації особи

Принцип функціонування сучасних систем ідентифікації мовця ґрунтується на виділенні мови з фонограм та подальшому попарному порівнянні біометричних ознак, таких як індивідуальні та ідентифікаційно значущі характеристики особистості, що містяться у голосі.

На даний момент, розповсюдженим підходом до вирішення проблеми текстонезалежної ідентифікації є використання моделі гаусових сумішей (Gaussian mixture models, GMM) [19]. Зазвичай, у більшості систем ідентифікації в якості інформативних ознак використовують кепстральні мел-частотні коефіцієнти (MFCC) [20].

У випадках текстонезалежної ідентифікації мовця найвищу надійність продемонстрували GMM-системи, засновані на використанні спільного факторного аналізу (JFA), який був представлений та досліджений у роботах [21-23].

Рішення ідентифікації особи мовця в цих системах ґрунтується на використанні відношення правдоподібності. Еталони дикторів обчислюються або за допомогою класичної MAP-адаптації [24] GMM-моделі на UBM-моделі (universal background model), або з використанням більш потужних методів створення незалежних від каналів передачі еталонних образів диктора [22, 23].

Як перспективний метод для ідентифікації мовців розглядається метод опорних векторів (support vector machine, SVM) [25,26]. Він є дискримінантним (розподільним) методом, на відміну від породжувального методу GMM. Сучасні дослідження підтверджують, що гібридна система GMM-SVM, де SVM діє в модельному просторі супервекторів середніх GMM, виявляється найефективнішою для завдань ідентифікації мовця. Супервектор GMM середніх представляє собою відображення набору векторів MFCC голосового сигналу мовця в багатовимірний вектор.

На сьогоднішній день, зазначений метод MFCC-GMM-SVM є передовим в системах текстонезалежної ідентифікації диктора.

## 1.6 Обґрунтування методу ідентифікації особи і постановка задачі

Проведений в розділах 1.2 і 1.3 аналіз принципів побудови відомих систем ідентифікації мовця показав, що поряд з сучасними методами розпізнавання голосу особи використовуються і раніше розроблені інші ефективні методи. До них можна віднести спектрально-формантний (СФ) і метод ідентифікації мовця на базі статистик основного тону (СОТ) [13, 27]. Дані методи забезпечують точність, яка досягає точності методу MFCC-GMM-SVM тільки для випадку коротких і сильно зачумлених парольних фонограм. В таблиці 1.1 наведені порівняльні характеристики вказаних методів (число знаків «+» відображає ступінь залежності методу від параметрів сигналу).

Аналіз наведених в таблиці 1.1 основних характеристик проаналізованих методів ідентифікації особи, у кожного з них методів є свої недоліки і переваги. Тому проведення досліджень, спрямованих на покращення якості роботи існуючих систем автоматичної ідентифікації особи є доцільними.

Таблиця 1.1 – Основні характеристики методів ідентифікації особи

Метод	Параметри сигналу		
	Якість	Тривалість	Фізичний і емоціональний стан диктора
MFCC–GMM–SVM	+	++	++
СФ	++	+++	+
СОТ	+	++	++++

З ціллю оптимізації функціонування методів спектрально-формантного розпізнавання мовця, ця магістерська робота прагне підвищити точність визначення частоти основного тону, яка визначає тембр голосу мовця. Зокрема, робота спрямована на аналіз енергій короткочасного спектру мовленнєвого сигналу та використання нейронної мережі в ролі класифікатора.

З метою вирішення поставленої мети в роботі розв'язуються наступні задачі:

- Створити математичну модель виділення частоти основного тону шляхом підрахунку частоти переходів сигналу мови через нуль;
- обрати оптимальну для задачі розпізнавання голосу особи модель нейромережевого класифікатора;
- розробити структурну схему системи та схему взаємодії програм;
- розробити алгоритми і програмне забезпечення системи для ознакового опису голосу мовця;
- провести машинний експеримент для визначення адекватності розробленого програмного забезпечення.



## 2 РОЗРОБКА МЕТОДІВ ОПИСУ ТА ІДЕНТИФІКАЦІЇ ОСОБИ ЗА ЇЇ ГОЛОСОМ

В даній роботі об'єктом автоматизації є підсистема ідентифікації особи, що належить до користувачів даної системи розпізнавання її голосу. Система ідентифікації повинна автоматично обробити мовний сигнал, проголошений мовцем, виділити з нього ознаки, інформативні до індивідуальності даної особи, виконати її опис його у вигляді вектора ознак та з використанням цього опису прийняти рішення про входження даний мовець до зареєстрованого в системі кола користувачів. Одним із особливо важливих етапів при цьому є виділення інформативних ознак голосу мовця. Розглянемо основні види ознак та методи, що використовуються для їх виділення.

### 2.1 Характеристика методів параметризації сигналу мовця

Головною метою лінгвістичного взаємодії є передача інформації. Індивідуальні особливості мови взаємодіють з структурами, що визначають смисловий зміст, і існують, не заважаючи комунікативному процесу. З цього випливає, що окремі мовні особливості не виокремлюються як ізольовані параметри, а взаємодіють з параметрами, що визначають мовну структуру та зміст мовного сигналу. Індивідуальні особливості існують завдяки значній адаптивності мовної системи, яка дозволяє здійснювати комунікативний акт у разі великої мінливості одних і тих самих повідомлень (індивідуальних, емоційних, ситуативних тощо). Мова кожної особи має свої індивідуальні риси, оскільки відсутня абсолютна норма як на фонетичному рівні, так і на рівні просодичних характеристик (інтонація, ритм).

За допомогою акустичної моделі формування мовлення [28-30] можна попередньо передбачити місця, де можна очікувати виявлення конкретних мовних особливостей, які є інформативними для вирішення задачі

ідентифікації мовця. Серед таких особливостей визначаються характеристики джерела голосу (низький рівень) [28-30] та акустичного фільтра [17, 30] (артикуляційний тракт). Мовна структура мовних звуків формується акустичним фільтром, тому характеристики повинні бути менше залежними від індивідуальних особливостей артикуляційного тракту. Джерело голосу відповідає лише за створення звуку в артикуляційному тракті, тобто найбільша змінливість параметрів мови спостерігається в роботі джерела голосу, що відображається у таких характеристиках:

- Середнє та дисперсія частоти основного тону;
- частотна модуляція періодів пітча
- амплітудна модуляція періодів пітча;
- контур основного тону;
- розподіл періодів частоти основного тону (пітча);
- форма збуджуючої хвилі джерела голосу.

При аналізі індивідуальних рис мовного сигналу, викликаних дією акустичного фільтра, застосовують два підходи:

1. Моделювання роботи артикуляторного тракту, хоча ефективно, має свої обмеження. Процедури моделювання, хоча великі та точні відносно, все ж таки припускають осьову симетрію артикуляторного тракту, що може не відповідати індивідуальним особливостям. Також питання залишається щодо того, наскільки отримані апроксимації артикуляторного тракту враховують індивідуальні розходження, оскільки можуть бути згладжені активними артикуляторними органами.
2. Визначення впливу параметрів артикуляторного тракту на мовний сигнал за допомогою аналізу мовної хвилі. У цьому випадку прив'язка до форми артикуляторного тракту є умовною, оскільки дійсно аналізується форма мовної хвилі.

В сучасних системах автоматичної ідентифікації мовця для параметризації мовного сигналу використовуються такі методи опису:

- Коефіцієнти лінійного прогнозу [28];
- основні спектральні компоненти [29, 30];
- параметри кепстра [28].

Ці підходи базуються на формальному аналізі мовного сигналу і вимагають подальшого експериментування для визначення їхньої ефективності. Оскільки різноманітність мовного матеріалу включає різні умови шуму та перешкод, обрання оптимального методу первинної параметризації мовного сигналу залишається завданням експериментальної перевірки.

## 2.2 Аналіз способів відокремлення інформативних ознак з сигналу мовлення

Як зазначалося вище [29], основний тон (пітч) характеризує вплив джерела голосу (голосової низки) людини на мовленнєвий процес, а тому є індивідуальною характеристикою мовця. Це може варіюватися в залежності від емоційного відтінку мови, проте в межах відносно обмежених діапазонів, як видно з табл.2.1. Сигнал основного тону породжується голосовими низками, має імпульсний характер, що породжує в спектрі сигналу мовця гармоніки обертопу, які характеризують тембр його голосу.

Таблиця 2.1 – Діапазон частот пітча і довжини голосових низок особи від типу голосу та статі

Голоси	Тип голосу	Діапазон частоти основного тону під час співу, Гц	Діапазон частоти основного тону під час розмови, Гц	Довжина голосових низок, мм
Жіночі	Меццо-сопрано	220-880	160-340	18-19

	Сопрано	260-1020		14-17
	Контральто	170-680		18-21
Чоловічі	Бас	80-350	90-210	24-25
	Баритон	100-400		22-24
	Тенор	130-510		18-24

Під час параметричного кодування мови приймають, той факт, що частота основного тону у людини розташована у діапазоні від 80 до 400 Гц, можна класифікувати за методами визначення пітчу (основного тону) на наступні групи:

- Амплітудна селекція;
- кореляційні методи;
- вибір частоти.

Надалі розглядаються фундаментальні теоретичні принципи, що становлять основу для кожної з категорій методологій.

### 2.2.1 Метод амплітудної селекції

У областях стаціонарності мовного сигналу, де відбувається відтворення аудіозаписів при низькому рівні шуму, форма мовних коливань майже ідентично повторюється в кожному циклі основного тону. Розмір відстані між максимальними піками мовного сигналу можна приблизно вважати періодом основного тону. Алгоритми вибору амплітуди характеризуються проблемою ідентифікації та послаблення локальних помилкових піків мовного сигналу. Зазвичай цього досягається підвищенням порогу активації схеми пошуку максимуму, що, однак, може збільшити ймовірність пропуску реального піка. Для підвищення надійності визначення періоду основного тону можна, наприклад, включити другий канал вибору

амплітуди, який вказує положення мінімумів мовного сигналу. На рис. 2.1 наведено часові діаграми такого пристрою, де  $S(t)$  – мовний сигнал з періодом основного тону (пітча).

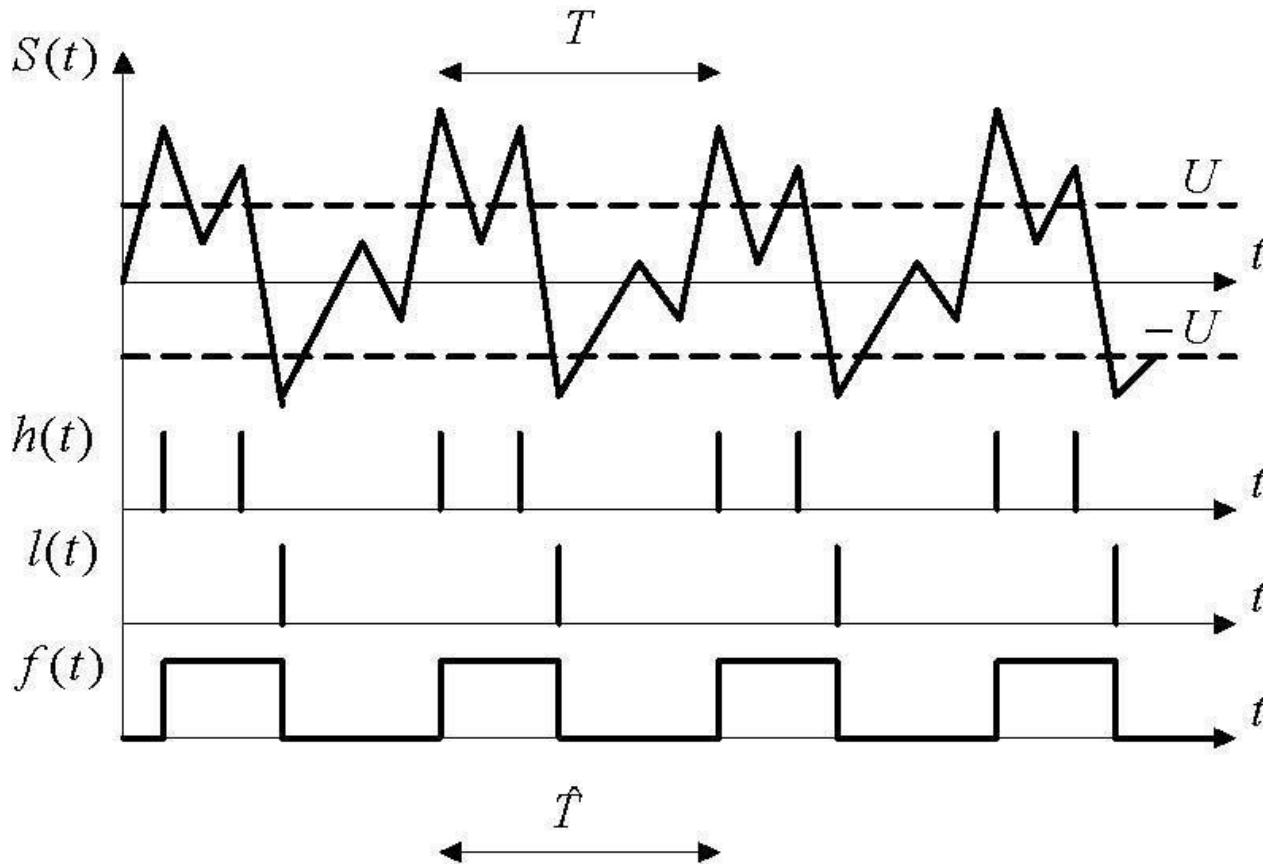


Рисунок 2.1 – Часові діаграми амплітудної селекції пітча

Сигнали  $h(t)$  та  $l(t)$  – представляють собою послідовності імпульсів, відповідних максимальним і мінімальним значенням мовного сигналу. Ці імпульси взаємодіють з тригером, формуючи на його виході сигнал  $f(t)$  з періодом  $\hat{T}$ , наближений до реального значення періоду основного тону  $T$ .

Основним плюсом амплітудних методів селекції є їх надзвичайна простота в реалізації, яка охоплює як пристрої, так і алгоритми. Головним недоліком є обмежена точність визначення частоти основного тону.

### 2.2.2 Кореляційні методи оцінки періоду пітчу

Представимо мовний сигнал у вигляді послідовності відліків  $S(t)$ ,  $t=0,1,2$ . Для вокалізованих звуків можна вважати, що

$$S(n) \approx S(n - T),$$

де  $T$  – період основного тону, виражений в кількості відліків. Для його оцінки в момент часу  $n$  виберемо значення  $k$  таке, що мінімізує функцію

$$L(n, k) = \sum_{i=0}^{N-1} (S(n + i) - S(n - k + i))^2 \quad (2.1)$$

Припустимо, що енергія мовного сигналу залишається незмінною в області стаціонарності. Таким чином, оцінка періоду основного тону має на меті максимізувати кореляційну функцію.

$$R(n, k) = \sum_{i=0}^{N-1} S(n + i) S(n - k + i)$$

Цей підхід гарантує значно вищу точність визначення періоду основного тону порівняно з методами амплітудної селекції. Важливо відзначити, що кореляційні методи визначення основного тону супроводжуються значною обчислювальною складністю. Також існують модифікації кореляційних підходів для визначення основного тону, які базуються на розрахунку взаємкореляційної функції.

$$R(n, k) = \sum_{i=0}^{N-1} S(n + 1) F(S(n - k + 1)) \quad (2.2)$$

Шляхом підбору функції  $F(x)$  можна спростити алгоритм, зробивши його придатним для апаратної реалізації: нехай функція  $F(x)$  кліппує мовний сигнал на три рівні  $\{-1, 0, 1\}$ . Тоді обчислювач взаємкореляційної функції (2.2) можна побудувати без помножувача

Розглянуті методи кореляції для оцінки періоду основного тону мають загальний недолік: їх нестійка ефективність виявляється у випадку, коли мовний сигнал зазнає амплітудної модуляції. Це стає особливо актуальним,

оскільки енергія реального мовлення, яке має емоційний відтінок, відмінно змінюється навіть на стаціонарних ділянках, що відповідають одній фонемі.

Змінимо функцію (2.1)

$$L(n, k) = \sum_{i=0}^{N-1} (S(n+i) - a_k S(n-k+i))^2 \quad (2.3)$$

де  $a_k$  – коефіцієнт підсилення. Для зсуву  $k$  оптимальне значення  $a_k$

обчислюється за формулою

$$a_k = \frac{\sum_{i=0}^{N-1} S(n+1) S(n-k+i)}{\sum_{i=0}^{N-1} S(n-k+i)^2} \quad (2.4)$$

Отже, як оцінку періоду основного тону у момент часу  $n$  слід взяти таке значення  $k$ , яке б максимізувало функцію

$$M(n, k) = \frac{\sum_{i=0}^{N-1} S(n+1) S(n-k+i)^2}{\sum_{i=0}^{N-1} S(n-k+i)^2} \quad (2.5)$$

Цей метод дозволяє отримувати точні оцінки основного тону, які плавно змінюються в часі з урахуванням змін у мовному сигналі. Тепер цей алгоритм активно використовується в стандарті G.723, який належно регулює стилізацію мовних сигналів для відеоконференцій.

### 2.2.3 Частотна селекція частоти основного тону

При вокалізованому збудженні мовного тракту у спектрі мовного сигналу можна виявити піки на частотах, що є кратними частотам пітчу. Якщо скласти дискретне перетворення Фур'є з досить малим кроком квантування по частоті, можна використовувати частоту, яка відповідає максимальному значенню енергетичного спектру, як оцінку частоти пітчу. Пошук максимуму рекомендується проводити в діапазоні 80-400 Гц. Однак

іноді в цьому діапазоні може знаходитися друга гармоніка основного тону, іноді навіть із більшою енергією. У таких випадках існує ризик помилкового визначення цієї гармоніки як оцінки основного тону. З метою уникнення цього, проводиться пошук максимуму не в спектрі, а в певній функції [45-46]

$$P(n, k) = \prod_{r=1}^R |X(n, kr)|^2$$

де введено індекс  $n$  для підкреслення того, що як спектр  $X(n, k)$ , так і функція  $P(n, k)$  обчислюються в конкретний момент часу  $n$ . Враховуючи той факт, що логарифм  $P(n, k)$  є монотонно зростаючою функцією в межах допустимих значень, цільову функцію можна виразити у вигляді:

$$\overline{P}(n, k) = \frac{1}{2} \ln(P(n, k)) = \sum_{r=1}^R \ln|X(n, k * r)| \quad (2.6)$$

Ця функція є сумою  $R$  стиснених по частоті в  $r$  разів логарифмів спектру потужності мовного сигналу. Основна ідея полягає в тому, що для фактичної частоти основного тону друга гармоніка другого доданку формується за рахунок першої гармоніки першого доданку, що призводить до її посилення. Аналогічно цей принцип діє і для третього доданку та подібних випадків. Внаслідок виконання для аудіосигналу, пов'язаного із голосовими виразами, спостерігатиметься виділений пік функції  $P(n, k)$  на частоті основного тону, тоді як для неголосового сигналу акумуляція буде випадкового характеру. Рисунок 2.2 ілюструє функціонування описаного методу.



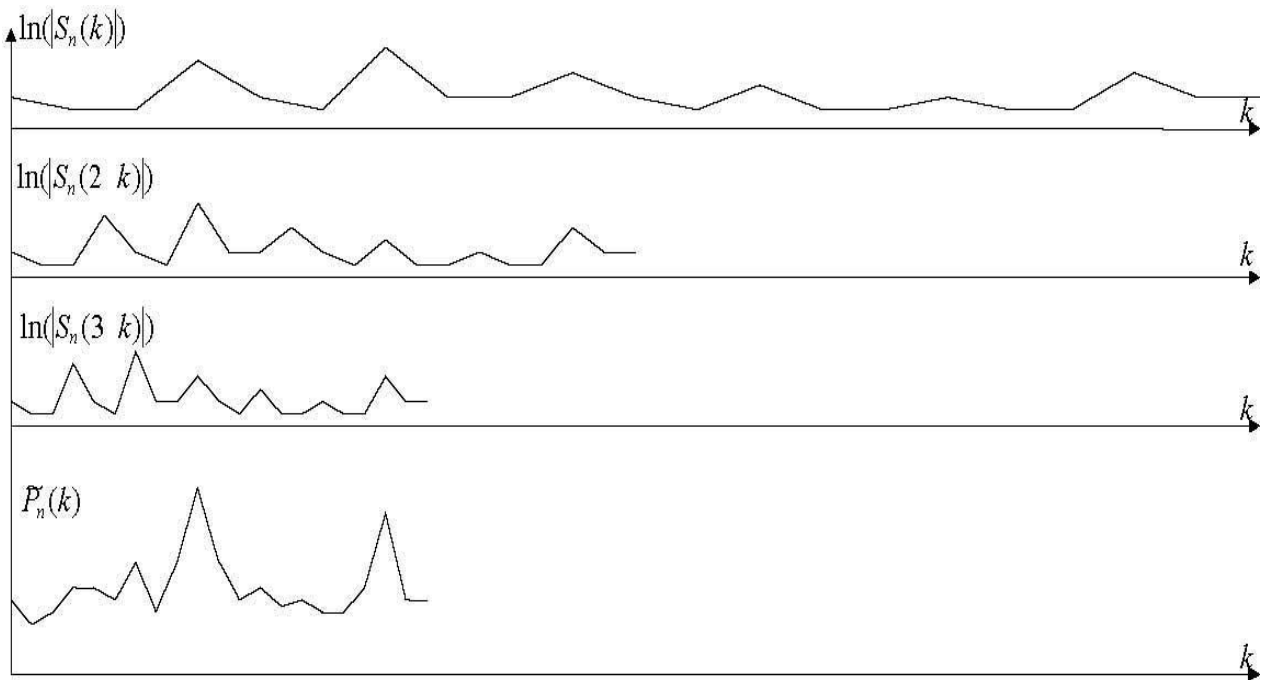


Рисунок 2.2 – Частотний метод визначення пітча

Між недоліками даного методу можна відзначити високу складність обчислень і обмеження, пов'язані із використанням базису Фур'є для отримання спектру мовного сигналу, який не завжди забезпечує необхідну роздільну здатність в просторі частота-час.

Один із найбільш поширених методів аналізу мовних сигналів - це короткочасний спектральний аналіз [18, 28-30], заснований на припущенні, що на відносно короткому часовому інтервалі мову можна розглядати як стаціонарний процес. Це припущення дозволяє вивчати суттєві часові зміни у мовному сигналі. Математично короткочасний спектр сигналу  $f(t)$  визначається через кругову частоту та час згідно з таким виразом:

$$F(\omega, t) = \int_{-\infty}^t f(\tau)h(t - \tau)e^{-j\omega\tau} d\tau \quad (2.7)$$

де  $h(t)$  – вагова віконна функція, яка виділяє лише частину сигналу, яка примикає до аналізованого моменту часу. На практиці, ефективна довжина функції  $h(t)$  становить 20-30 мс. Часто поряд з короткочасним спектром розглядають амплітудний спектр  $F(\omega, t)$  або енергетичний спектр  $F(\omega, t)^2$ .

Для впровадження короточасного спектрального аналізу використовуються різноманітні підходи, серед яких найбільше поширений – це проведення аналізу за допомогою фільтрації за гребінцею смугових фільтрів.

Представимо спектр  $F(\omega, t)$  в наступному вигляді:

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

$$= a(\omega, t) - jb(\omega, t) \quad (2.8)$$

Енергетичний спектр визначається як

$$|F(\omega, t)|^2 = a^2(\omega, t) + b^2(\omega, t) \quad (2.9)$$

Для реалізації енергетичного спектру сигнал  $f(t)$  потрібно помножити на  $\cos \omega t$  і  $\sin \omega t$ , відфільтрувати гребінкою фільтрів з імпульсними характеристиками  $h(t)$ , піднести до квадрату та отримати їх суму. Цей процес зображено на рис. 2.3 [28].

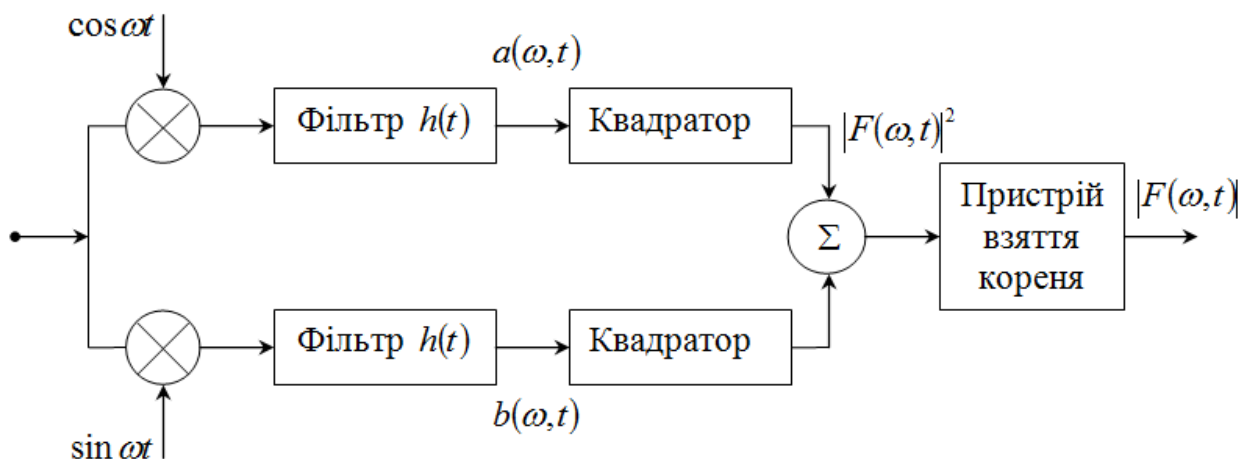


Рисунок 2.3 – Схема виділення енергетичного спектру

Для отримання амплітудного спектру до зазначеної процедури вводиться виділення кореня енергетичного спектру  $|F(\omega, t)|^2$ . Проте такий

метод виділення спектру є складним. Зазвичай на практиці для спрощення процедури визначення функції використовують часову обвідну  $a(\omega, t)$  або  $b(\omega, t)$ . Фізично цю обвідну апроксимують формою обвідної на виході смугового фільтру, використовуючи детектування сигналів  $a(\omega, t)$  або  $b(\omega, t)$  і подальше пропускання результату через фільтр низьких частот [28].

З аналогічним підходом можна здобути енергетичний спектр на короткий період часу. Вимірювання енергії на виході кожного смугового фільтру надає надійну оцінку спектру на центральній частоті, що обумовлює вибір емпірично діапазонів, кількості смугових фільтрів та їх вагових функцій відповідно до поставленої задачі.

Викладені сценарії аналізу в часі та частоті можна використовувати для конвертації сигналу  $f(t)$  у цифрову форму. Існує різноманіття технік цифрового представлення мовних сигналів, проте найпоширенішим є використання методів швидкого перетворення Фур'є для короткочасного часово-частотного аналізу мовних сигналів [17, 29, 30].

Нехай мовний сигнал  $f(t)$  представлено послідовністю із  $N$  відліків  $f(nT)$ ,  $0 \leq n \leq N - 1$ , де  $T$  – проміжок дискретизації сигналу у часовій площині. Тоді спектр дискретизованого сигналу можна представити як:

$$F(\omega) = \sum_{n=0}^{N-1} f(nT)e^{-j\omega nT} \quad (2.10)$$

Згідно теореми відліків в частотній області для обмеженої в часі функції на інтервалі  $NT$  спектральна щільність однозначно визначається визначенням своїх дискретних значень в точках  $\omega = \frac{2\pi}{NT}k$ , де  $k = 0, 1, 2$ . Ввівши позначення  $\Omega = \frac{2\pi}{NT}$ , одержимо наступний вигляд для дискретного перетворення Фур'є:

$$F(k\Omega) = \sum_{n=0}^{N-1} f(nT)e^{-jk\Omega nT} \quad (2.11)$$

Процес відтворення вихідної послідовності мовного сигналу з використанням оберненого дискретного перетворення Фур'є може бути описаний наступним чином:

$$f(lT) = \frac{1}{N} \sum_{k=0}^{N-1} F(k\Omega) e^{-j l T k \Omega} \quad (2.12)$$

Позначивши  $e^{-j\Omega T} = W$  та  $f(nT) = f_n$ , пряме та обернене ДПФ запишемо як:

$$F_k = \sum_{n=0}^{N-1} f_n W^{nk}, \quad f_n = \frac{1}{N} \sum_{k=0}^{N-1} F_k W^{-lk} \quad (2.13)$$

З метою зменшення обчислювальних операцій, потрібних для отримання ДПФ мовного сигналу, використовують алгоритми швидкого перетворення Фур'є (ШПФ). Розглянемо конкретний підхід, відомий як метод часового розрідження. Якщо кількість відліків  $N$  ділиться на 2, то з однієї послідовності значень  $f_n$  можливо скласти дві коротші послідовності, утворені із парних та непарних відліків. Перша послідовність відліків утворює функцію  $g_d = f_{2d}$ , друга – функцію  $h_d = f_{2d+1}$ ,  $d = 0, 1, 2, N/2 - 1$ .

Обчислюючи для кожної з цих функцій ДПФ, одержимо:

$$G_k = \sum_{d=0}^{N/2-1} g_d W^{2dk} \quad \text{та} \quad H_k = \sum_{d=0}^{N/2-1} h_d W^{2dk}$$

Дискретний перетворений часовий ряд можна виразити, відокремивши парні та непарні точки під символом суми.

$$F_k = \sum_{d=0}^{N/2-1} (g_d W^{2dk} + h_d W^{(2d+1)k}), \quad k = 0, 1, 2, N - 1 \quad (2.14)$$

або

$$F_k = \sum_{d=0}^{N/2-1} g_d W^{2dk} + W^k \sum_{d=0}^{N/2-1} h_d W^{2dk} = G^k + W^k H_k \quad (2.15)$$

Для описаного методу кількість обчислень склала  $N^r/2 + N$ , а для ДПФ –  $N^r$ . Якщо  $N/2$  кратне 2, то кількість операцій можна ще зменшити.

Інша форма ШПФ – частотне прорідження. Припустимо, що функція  $f_d$  з парним числом точок  $N$  розділена на дві послідовності по  $N/2$  точок кожна, наприклад,  $g_d$  та  $h_d$ , де  $g_d = f_d$  та  $h_d = f_d + N/2$ , для  $d = 0, 1, 2, N/2 - 1$ .

ДПФ для  $N$  точок, позначене як  $F_k$ , може бути записано через  $g_d$  та  $h_d$

$$F_k = \sum_{d=0}^{N/2-1} (g_d W^{dk} + h_d W^{(d+N/2)k}) = \sum_{d=0}^{N/2-1} (g_d W^{dk} + h_d W^{dk} W^{(N/2)k}) \quad (2.16)$$

Підставляючи  $W^{(N/2)k} = e^{-jNTk\Omega/2} = e^{-j\pi k}$ , одержимо:

$$F_k = \sum_{d=0}^{N/2-1} (g_d + h_d) W^{2dk} \quad (2.17)$$

Розглянемо парні та непарні відліки  $F_k$ . Замінивши  $k$  на  $2k$ , одержимо рівняння:

$$F_{2k} = \sum_{d=0}^{N/2-1} (g_d + h_d) W^{2dk} \quad (2.18)$$

яке являє собою ДПФ  $N/2$  точок функції  $g_d + h_d$ .

Аналогічно, для спектру непарних точок одержимо такий вираз:

$$F_{2k+1} = \sum_{d=0}^{N/2-1} |(g_d - h_d) W_d| W^{2dk} \quad (2.19)$$

який є ДПФ  $N/2$  точок функції  $(g_d + h_d)W^d$ .

Подібно до сценарію часового розрідження, кількість обчислень для отримання ДПФ мовного сигналу відповідно до цього підходу дорівнює  $N * \log_2 N$ .

Серед недоліків використання спектрального аналізу голосових сигналів у межах розпізнавання мовців можна відзначити випадки, коли тимчасові зміни у мовному сигналі ускладнюють завдання, і застосування базису Фур'є не забезпечує достатньої роздільної здатності для ефективної реалізації в реальному часі.

### 2.3 Обчислення частоти основного тону за кількістю нульових перетинів

Для знаходження частоти основного тону диктора по кількості нульових перетинів можна використати наступну формулу [31]:

$$\omega(i) = \frac{W(i)}{2(M-1)} f_d \quad (2.20)$$

$$W_{K_D}(i) = \left\{ \frac{W_{K_D}(i)+1, \operatorname{sgn}[z_{K_D}(n)] \neq \operatorname{sgn}[z_{K_D}(n-1)]}{W_{K_D}(i), \operatorname{sgn}[z_{K_D}(n)] \neq \operatorname{sgn}[z_{K_D}(n-1)]} \right\},$$

$$W_{K_S}(j) = \left\{ \frac{W_{K_S}(j)+1, \operatorname{sgn}[z_{K_S}(n)] \neq \operatorname{sgn}[z_{K_S}(n-1)]}{W_{K_S}(j), \operatorname{sgn}[z_{K_S}(n)] \neq \operatorname{sgn}[z_{K_S}(n-1)]} \right\},$$

якщо  $W_{K_D}(i) = W_{K_S}(j)$ , то  $W(i) = W_{K_D}(i) = W_{K_S}(j)$ , де

$\omega(i)$  – домінантна частота,

$i$  – порядковий номер  $K_D$  в діапазоні  $0 \leq K_D \leq K_{D \max}$ ,

$j$  – порядковий номер  $K_S$  у діапазоні  $0 \leq K_S \leq K_{S \max}$ ,

$M$  – кількість відліків,

$f_d$  – частота дискретизації;

$$W_{K_s}(j) = \left\{ \frac{W_{K_s}(j)+1, \operatorname{sgn}[z_{K_s}(n)] \neq \operatorname{sgn}[z_{K_s}(n-1)]}{W_{K_s}(j), \operatorname{sgn}[z_{K_s}(n)] \neq \operatorname{sgn}[z_{K_s}(n-1)]} \right\} - \text{кількість нульових}$$

перетинів повторно-різницевої вибірки;

$$W_{K_s}(j) = \left\{ \frac{W_{K_s}(j)+1, \operatorname{sgn}[z_{K_s}(n)] \neq \operatorname{sgn}[z_{K_s}(n-1)]}{W_{K_s}(j), \operatorname{sgn}[z_{K_s}(n)] \neq \operatorname{sgn}[z_{K_s}(n-1)]} \right\} - \text{кількість точок, де}$$

вибірка повторного підсумовування перетинає нульовий рівень;

$$z_{K_D}(n) = \nabla^{K_D}(z_s(n)) = \nabla(\nabla^{K_D-1}(z_s)(n)), n = 1, \dots, M_D, M_D = M_S - K_D,$$

$K_D = 0, 2, K_{max}$  – сформована повторно-різницева вибірка,

$$z_s(n) = \Delta^{K_s}s(n) = \Delta(\Delta^{K_s-1}s(n)), n = 1, \dots, M_s, M_s = M - K_s -$$

сформована вибірка повторного підсумовування значень  $s(n)$ .

Отже, розроблено метод визначення частоти основного тону, який базується на кореляції функцій, що визначають кількість перетинів амплітудою сигналу нульового рівня. Використано повторно-різницевої та повторно-підсумовуючі вибірки, що базуються на модифікованій математичній моделі слухової системи людини. Цей підхід сприяє підвищенню роздільності між класами дикторів у просторі ознак та дозволяє виділяти частоту основного тону при наявності шумів навколишнього середовища в мовному сигналі. Програма для виділення частоти основного тону, яка використовується як одна з ознак для характеристики диктора, реалізована на мові програмування MATLAB і надається у додатку В.

## 2.4 Висновки до розділу

В цьому розділі магістерської кваліфікаційної роботи було проведено аналіз моделей індивідуальності особи в його голосовому сигналі, вибрано інформативні ознаки для опису диктора за його голосом, розглянуто математичні основи методів їх виділення з мовленнєвого сигналу і запропоновано метод і алгоритм виділення частоти основного тону шляхом підрахунку кількості нульових перетинів сигналу через нуль, запропоновано тип класифікатора для розпізнавання голосу мовця.



### 3 РОЗРОБКА АЛГОРИТМІВ І ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ СИСТЕМИ ІДЕНТИФІКАЦІЇ ОСОБИ

#### 3.1 Розробка схеми взаємодії програм для ідентифікації особи за її голосом

Враховуючи аналіз, проведений в першому розділі магістерської роботи та сучасний досвід проектування систем розпізнавання образів, розроблено структуру програмного забезпечення для автоматизованої ідентифікації особи в системах доступу до інформаційних ресурсів, яку показано на рисунку 3.1.

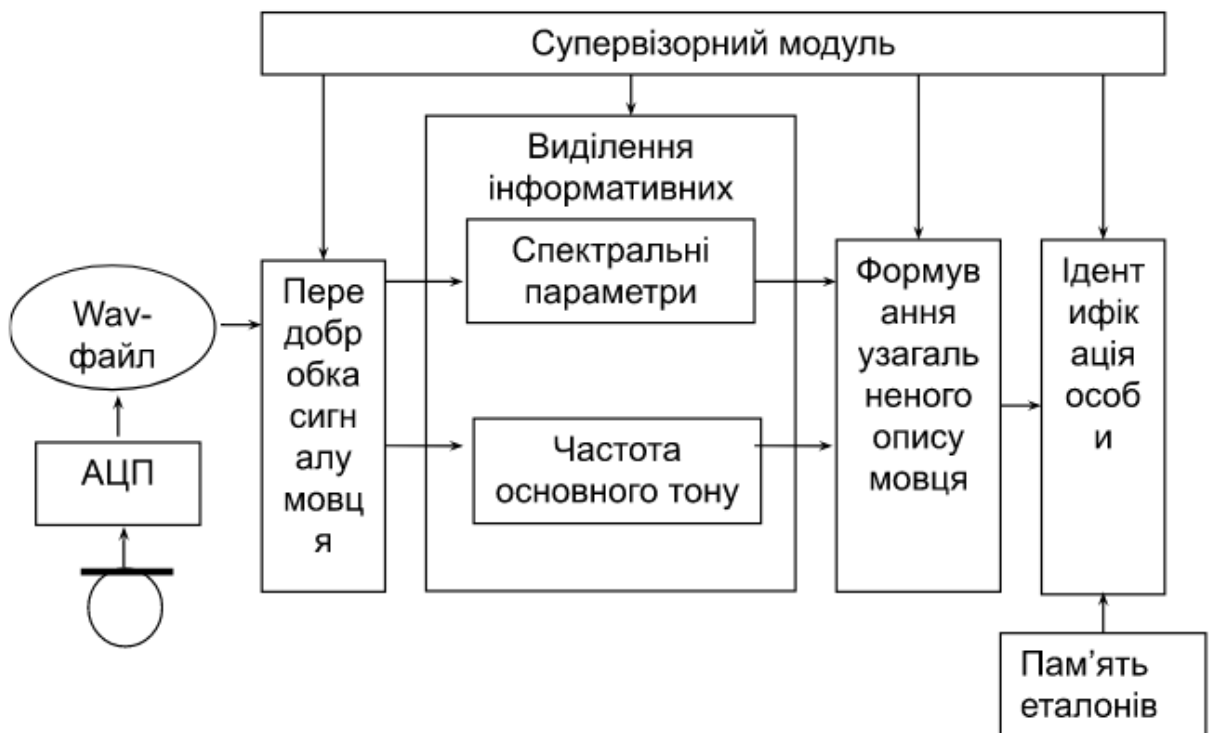


Рисунок 3.1 – Схема взаємодії програм ідентифікації особи за її голосом

Як видно з рис. 3.1, програмне забезпечення складається з таких основних модулів:

- Модуль керування, який координує роботу інших компонентів у програмному забезпеченні для ідентифікації диктора за його голосом;
- модуль попередньої обробки цифрового мовного сигналу, який застосовує фільтрацію, масштабування та сегментацію вхідного мовного сигналу за допомогою розроблених алгоритмів;
- модуль виділення інформативних ознак, який обробляє цифровий мовний сигнал за допомогою спеціальних алгоритмів для виділення характеристик, таких як спектральні параметри та кореляційні функції частоти основного тону;
- модуль формування узагальненого образу диктора, що базується на векторах інформативних ознак, і створює загальний образ диктора у просторі, об'єднуючи всі інформаційні характеристики
- модуль прийняття остаточного рішення, де використовуючи навчений класифікатор, проводиться порівняння узагальненого образу диктора із зразками образів дикторів, на яких тренувався класифікатор, і здійснюється рішення щодо ідентифікації особи диктора.

Враховуючи спосіб, яким люди сприймають мовні сигнали, та конструктивні особливості апаратної частини системи розпізнавання голосу диктора, програмне забезпечення отримує безпосередньо цифровий мовний сигнал у форматі wav, з частотою дискретизації 10 кГц та розширеністю 16 біт. Це дозволяє зберегти усі унікальні особливості голосу диктора, як вказано у [14]. У випадку передачі парольного файлу через мережу, сигнал мовця спочатку відтворюється з архіву.

### 3.2 Розробка алгоритму спектрального аналізу сигналу мовця

У контексті створення програмного забезпечення для розпізнавання диктора за його голосом, були розроблені спеціалізовані алгоритми виділення

ключових ознак з мовного сигналу, враховуючи унікальні особливості його формування диктором. Ці алгоритми включають:

- FLT\_OGIB: алгоритми фільтрації мовного сигналу, спрямовані на проведення спектрального аналізу та визначення енергії сигналу в конкретних частотних діапазонах;
- OBR\_SYGN: алгоритм обробки сигналу, що визначає параметри цифрових фільтрів і запускає процес фільтрації мовного сигналу у визначених смугах частот.

При використанні спектрального аналізу мовного сигналу за допомогою смугових фільтрів, які мають центральні частоти розташовані за логарифмічним законом вздовж частотної осі, в окремих частотних піддіапазонах  $\Delta F_1$ ,  $\Delta F_2$ ,  $\Delta F_3$  знаходяться ті, які визначені в таблиці 3.1.

Таблиця 3.1 – Розташування частотних піддіапазонів

Частотний діапазон	Смуга частот фільтра	Частотний піддіапазон $\Delta f_{ki}$	Середня частота, Гц.
$\Delta F_1$	252 / 317	$\Delta f_{11}=252/400$	317
	317 / 400		504
	400 / 504	$\Delta f_{12}=400/635$	
	504 / 635		800
	635 / 800	$\Delta f_{13}=635/1008$	
	800 / 1008		
$\Delta F_2$	800 / 1008	$\Delta f_{21}=800/1270$	1008
	1008 / 1270		1425
	1270 / 1600	$\Delta f_{22}=1270/1600$	2016
	1600 / 2016		
	2016 / 2540	$\Delta f_{23}=1600/2540$	

$\Delta F_3$	1600 / 2016	$\Delta f_{31}=1600/2540$	2016
	2016 / 2540		2850
	2540 / 3200	$\Delta f_{32}=2540/3200$	3592
	3200 / 4032	$\Delta f_{33}=3200/4032$	

Алгоритми реалізовані у вигляді відповідних підпрограм мовою Matlab.

Підпрограма FLT\_OGIB призначена для обробки вхідного сигналу за допомогою цифрового смугового фільтра. Під час виконання цієї підпрограми генерується вектор енергії звукового сигналу на інтервалі 20 мсек і обчислюється інтегральна енергія на виході кожного фільтра за весь період виконання парольної фрази. Для цього використовуються параметри цифрового смугового фільтра, верхня і нижня межі частоти пропускання, а також номер фільтра. Результатом є рядок матриці, що представляє вектор енергії звукового сигналу, відфільтрований конкретним смуговим фільтром. Номер смугового фільтра відповідає номеру рядка вихідної матриці. Схема алгоритму зображена на рисунку 3.2, а лістинг програми наведений в додатку В.

Друга зовнішня підпрограма – функція OBR\_SIGN – використовується в основній програмі і відіграє роль завантаження звукового файлу та активації кожного з п'ятнадцяти фільтрів функції FLT\_OGIB. Вона працює з вектором вхідного сигналу, визначає верхні та нижні межі частоти пропускання смугових фільтрів та їхні номери. Результатом є матриця, яка містить значення енергій звукового сигналу, виділених п'ятнадцятьма смуговими фільтрами, з періодом дискретизації 20 мсек. Схема алгоритму показана на рисунку 3.3, а лістинг програми можна знайти в додатку В.

### 3.3 Розробка алгоритму виділення частоти основного тону

Загалом вважається, що на відрізках вокалізованого звуку артикуляторний тракт людини стає об'єктом періодичних коливань низької частоти. Ці періодичні коливання, відомі як період основного тону, вважаються індивідуальною характеристикою для ідентифікації диктора за його голосом [1,2,17]. У розробці систем для ідентифікації диктора за голосом, часто використовується характеристика, обернена до періоду основного тону, а саме – частота основного тону. Ця характеристика може зазнавати змін в залежності від емоційного забарвлення мови, але ці зміни зафіксовані у відносно стислих межах. У параметричному кодуванні мови припускається, що частота основного тону розташовується в певних діапазонах (див. підрозділ 1.3).

У даній роботі розроблено метод виділення частоти основного тону, який комбінує амплітудні та кореляційні підходи [1,2,17], використовуючи метод нульових перетинів [1,2,17] та аналіз короткочасної енергії мовного сигналу [1,2,17]. Математично обгрунтований підхід до розробленого методу подано у подальшому викладі.

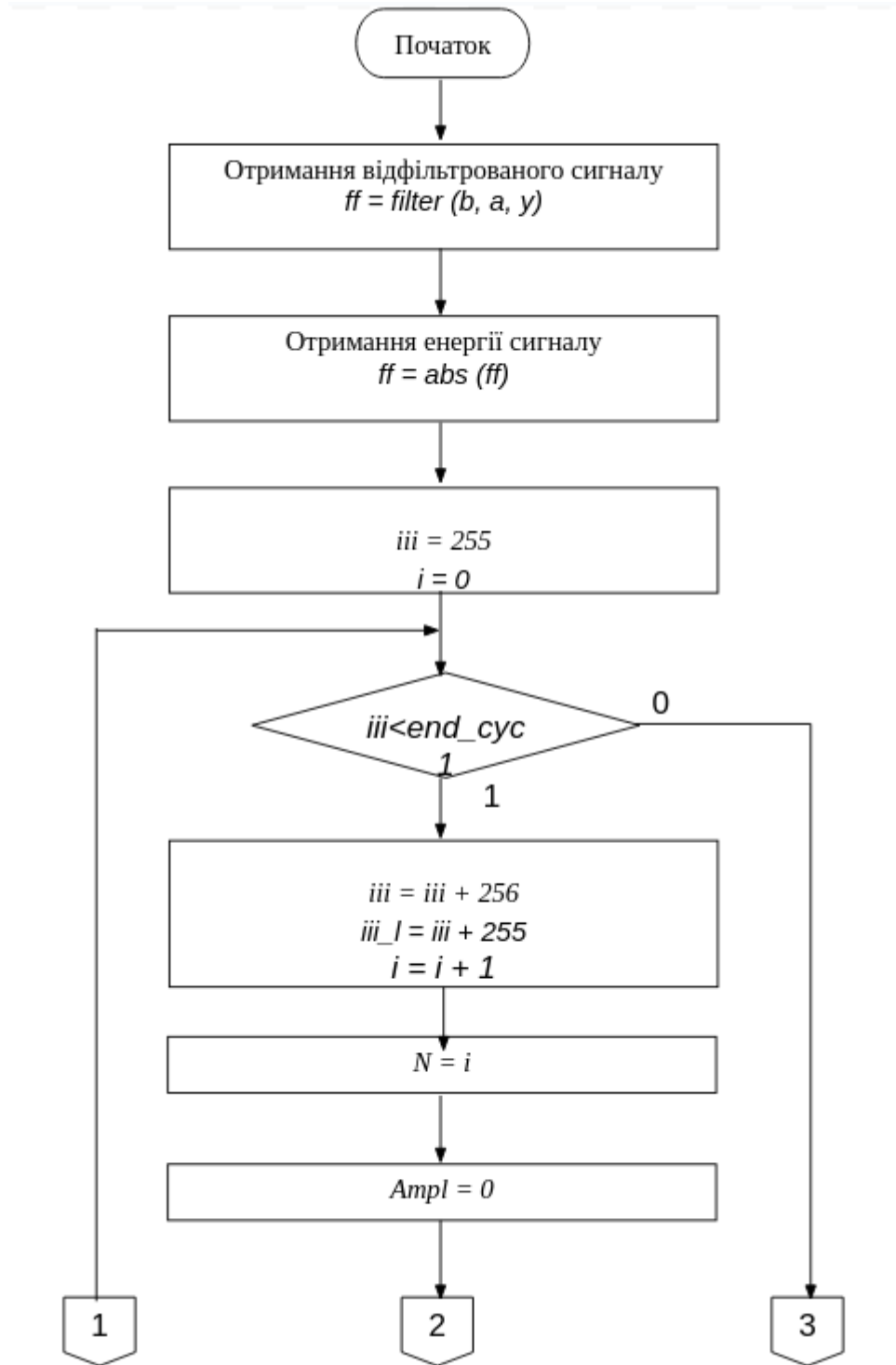


Рисунок 3.2 – Схема підпрограми FLT\_OGIB (початок)

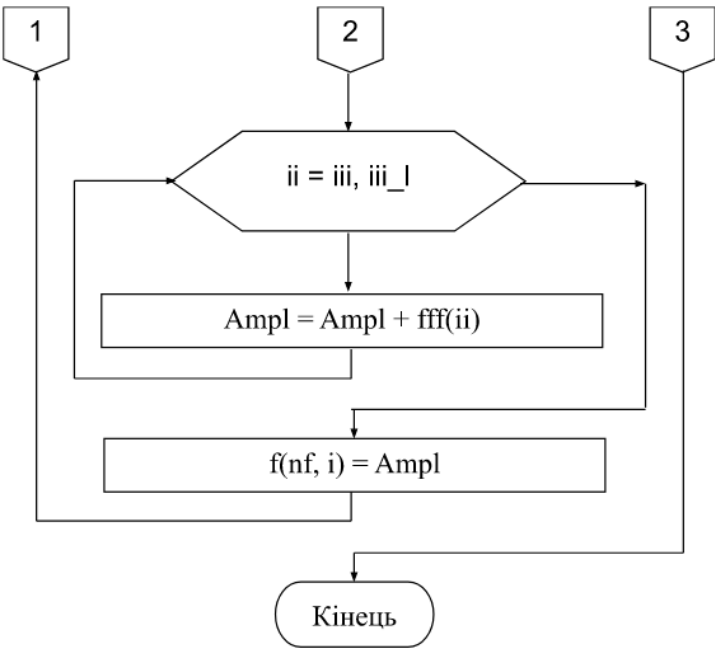


Рисунок 3.2 – Схема підпрограми FLT\_OGIB (закінчення)

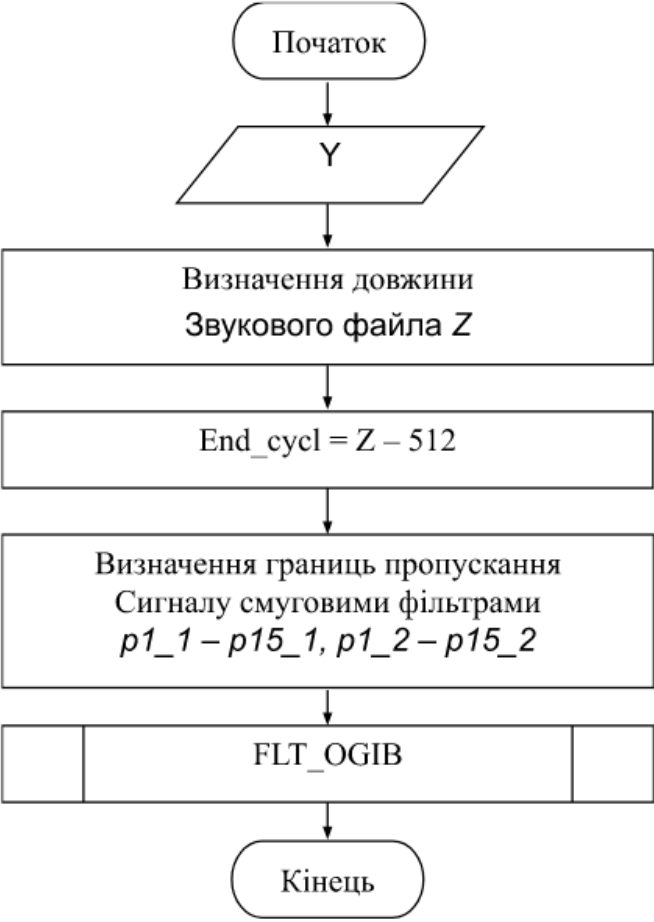


Рисунок 3.3 – Схема підпрограми OBR\_SIGN.

Нехай  $s(n) = s(0), s(1), \dots, s(N - 1)$  – часовий ряд, що є стаціонарним із нульовим середнім значенням. Згідно з теоремою Уолда, коваріаційна функція такого процесу може бути виражена у вигляді:

$$\gamma(k) = \int_{-\pi}^{\pi} \cos(k\omega) dG(\omega) \quad (3.1)$$

де  $G(\omega)$  – спектральна щільність потужності сигналу,

$k = 0, \pm 1, \pm 2, \dots$  – часове зміщення.

Автокореляційна функція – нормована величина  $R(k) = \gamma(k)/\gamma(0)$  з урахуванням (3.1) може бути виражена таким рівнянням

$$R(k) = \frac{\int_{-\pi}^{\pi} \cos(k\omega) dG(\omega)}{\int_{-\pi}^{\pi} dG(\omega)}, k = 0, \pm 1, \pm 2, \dots \quad (3.2)$$

Визначимо двосторонньо-обмежений двійковий ряд  $y(n) = \begin{cases} 1, & s(n) \geq 0 \\ 0, & s(n) < 0 \end{cases}$ ,  $0 \leq n \leq N - 1$ , і введемо індикаторну функцію для моменту часу  $F_n = (y(n) - y(n - 1))^2$

Якщо  $F_n = 1$ , то коли амплітудне значення сигналу нульового рівня перетинається в момент часу  $t$ , а в протилежному випадку,  $F_n = 0$ . Кількість перетинів амплітудним значенням сигналу нульового рівня для ряду  $s(n)$ ,  $n = 0, 1, 2, \dots, N - 1$ , визначається наступним співвідношенням:

$$W = \sum_{i=0}^{N-1} F_i$$

Для опису вказаного процесу, що використовує розподіл Гауса із нульовим середнім значенням, кількість перетинів амплітудних значень сигналу нульового рівня можна виразити таким чином:



$$E\{W_1\} = (N - 1)\left(\frac{1}{2} - \frac{1}{\pi} \arcsin r(1)\right), \text{ де}$$

$E\{ \}$  – математичне сподівання,

$r(1)$  знаходимо за формулою (3.2) при  $k = 1$ .

$$\text{Тоді } r(1) = \cos\left(\frac{\pi E\{W_1\}}{N-1}\right) = \frac{\int_{-\pi}^{\pi} \cos(k\omega) dG(\omega)}{\int_{-\pi}^{\pi} dG(\omega)} \quad (3.3)$$

Для гармонійного сигналу з частотою  $\omega_0$  спектральна щільність потужності при  $\omega \in [0, \pi]$  повинна задовольняти умові [78]

$$G(\omega) = \frac{A^2}{4} [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)] \quad (3.4)$$

де  $A$  – амплітуда гармоніки,

$\delta(\omega)$  – дельта-функція.

Підставляючи (3.3) в (3.4), переходимо до наступної рівності:

$$\frac{\pi E\{W_1\}}{N-1} = \omega_0 \quad (3.5)$$

Іншими термінами, частота гармоніки збігається із нормалізованим значенням математичного сподівання кількості перетинів амплітудних значень сигналу на нульовому рівні. Для дискретного сигналу і лінійної частоти вираз (3.5) може бути виражений таким чином:

$$\frac{E\{W_1\}}{N-1} * \frac{f_d}{2} = f_0$$

де  $f_d$  – частота дискретизації,

$f_0$  – частота гармоніки.

У більш загальному контексті, коли в сигналі присутня певна домінуюча частота на тлі завад, рівність (3.5) не виконується. Проте можна припустити, що нормалізоване значення кількості перетинів амплітудних значень сигналу на нульовому рівні буде прямувати до основної частоти. Без врахування операції математичного сподівання можна стверджувати, що у випадку, коли певна частота стає домінуючою, величина буде мати значення, що відповідає або близьке цій частоті.

Отже, якщо в спектрі сигналу присутня домінуюча частота, її можна визначити за допомогою оцінки кількості перетинів амплітудним значенням сигналу нульового рівня. Цей метод ефективний, коли в сигналі присутня лише одна гармоніка і відношення сигнал/шум велике. У випадках, коли є кілька гармонік або амплітуда гармонійної компоненти значно менша за шум, варто застосовувати фільтри для виділення періодичних компонентів сигналу та оцінки домінуючої частоти в різних частотних смугах.

Щоб обмежити частотний діапазон, де спостерігається частота основного тону, рекомендується використовувати полосовий фільтр. Потім виділений частотний проміжок розбивається на частотні полоси, до яких по чергово застосовуються процедури низькочастотної і високочастотної фільтрації. Це здійснюється за допомогою нерекурсивних цифрових фільтрів КІХ із симетричною імпульсною характеристикою. Різницеве рівняння для цих фільтрів виражено в наступному виразі:

$$y(n) = \sum_{i=0}^N h(i)s(n - i)$$

де  $h(i)$  – коефіцієнти фільтру,  $N$  - порядок фільтру.

Для виділення сигналу в області низьких частот можна використовувати операцію повторного підсумовування (фільтр низьких частот), при цьому перша сума буде записана у наступному вигляді:

$$\Delta s(n) = s(n) + s(n - 1)$$

А  $k$ -та кінцева сума запишеться як:

$$\Delta^k s(n) = \sum_{j=0}^k C_k^j s(n - j)$$

Позначимо  $W_k$  за кількість перетинів амплітудою сигналу нульового рівня для ряду  $\Delta^{k-1} s(n)$ . Оскільки оператор повторного сумування є фільтром нижніх частот, то ряд  $\Delta^k s(n)$  буде менш осцилюючим в порівнянні з рядом  $\Delta^{k-1} s(n)$ , а отже, і  $W_{k+1} < W_k$ .

В ролі фільтру низьких частот використовується кінцево-різницевий оператор  $\nabla$ , його перша різниця представлена таким чином:

$$\nabla s(n) = s(n) - s(n - 1),$$

а  $k$ -та кінцева різниця запишеться як:

$$\Delta^k s(n) = \sum_{j=0}^k C_k^j (-1)^j s(n - j), C_k^j = \frac{k!}{(k-j)! j!}$$

Позначимо  $W_D$  за кількість перетинів амплітудою сигналу нульового рівня для ряду  $\nabla^{k-1} s(n)$ . З метою отримання результатів, що володіють вищою гнучкістю та точністю при оцінці властивостей сигналу, рекомендується одночасно використовувати параметри  $W_D$  і  $k$ , послідовно застосовуючи до сигналу процедури фільтрації нижніх частот і смугової фільтрації  $\nabla^{k-1} \Delta^{j-1} s(n)$ , підраховуючи число нулів. Тоді кількість нулів ряду  $W$  буде відображати різні смуги частот сигналу.

Отже, для визначення частоти основного тону диктора за кількістю нульових перетинів можна використовувати наступний вираз [37]:

$$w(i) = \frac{W(i)}{2^{(M-1)}} f_d \quad (3.6)$$

$$W_{K_D}(i) = \left\{ \frac{W_{K_D}(i)+1, \operatorname{sgn}[z_{K_D}(n)] \neq \operatorname{sgn}[z_{K_D}(n-1)]}{W_{K_D}(i), \operatorname{sgn}[z_{K_D}(n)] \neq \operatorname{sgn}[z_{K_D}(n-1)]} \right\},$$

$$W_{K_S}(j) = \left\{ \frac{W_{K_S}(j)+1, \operatorname{sgn}[z_{K_S}(n)] \neq \operatorname{sgn}[z_{K_S}(n-1)]}{W_{K_S}(j), \operatorname{sgn}[z_{K_S}(n)] \neq \operatorname{sgn}[z_{K_S}(n-1)]} \right\},$$

якщо  $W_{K_D}(i) = W_{K_S}(j)$ , то  $W(i) = W_{K_D}(i) = W_{K_S}(j)$ , де

$\omega(i)$  – доміантна частота,

$i$  – порядковий номер  $K_D$  в діапазоні  $0 \leq K_D \leq K_{D \max}$ ,

$j$  – порядковий номер  $K_S$  у діапазоні  $0 \leq K_S \leq K_{S \max}$ ,

$M$  – кількість відліків,

$f_d$  – частота дискретизації;

$$W_{K_S}(j) = \left\{ \frac{W_{K_S}(j)+1, \operatorname{sgn}[z_{K_S}(n)] \neq \operatorname{sgn}[z_{K_S}(n-1)]}{W_{K_S}(j), \operatorname{sgn}[z_{K_S}(n)] \neq \operatorname{sgn}[z_{K_S}(n-1)]} \right\} - \text{кількість нульових}$$

перетинів повторно-різницевої вибірки;

$$W_{K_S}(j) = \left\{ \frac{W_{K_S}(j)+1, \operatorname{sgn}[z_{K_S}(n)] \neq \operatorname{sgn}[z_{K_S}(n-1)]}{W_{K_S}(j), \operatorname{sgn}[z_{K_S}(n)] \neq \operatorname{sgn}[z_{K_S}(n-1)]} \right\} - \text{кількість точок, де}$$

вибірка повторного підсумовування перетинає нульовий рівень;

$$z_{K_D}(n) = \nabla^{K_D}(z_s(n)) = \nabla(\nabla^{K_D-1}(z_s)(n)), n = 1, \dots, M_D, M_D = M_S - K_D,$$

$K_D = 0, 2, K_{\max}$  – сформована повторно-різницева вибірка,

$$z_s(n) = \Delta^{K_S} s(n) = \Delta(\Delta^{K_S-1} s(n)), n = 1, \dots, M_S, M_S = M - K_S -$$

сформована вибірка повторного підсумовування значень  $s(n)$ .

Отже, розроблено метод визначення частоти основного тону, що ґрунтується на кореляції функцій кількості перетинів амплітудою сигналу нульового рівня. Цей метод використовує повторно-різницеві та повторно-підсумовуючі вибірки, базуючись на модифікованій математичній моделі слухової системи людини. Його застосування сприяє покращенню роздільності класів дикторів у просторі ознак та виокремленню частоти основного тону при наявності шумів у мовному сигналі. Програма для виділення частоти основного тону, яка є однією з ознак для характеристики диктора, була розроблена на мові Matlab та подана в додатку В.

#### 3.4 Розробка алгоритмів навчання класифікатора для розпізнавання диктора

Автоматичний класифікатор – це комплекс програм та обладнання, який має здатність автономно вирішувати конкретні завдання класифікації. Теоретичною основою для таких класифікаторів може слугувати формула умовних ймовірностей Байєса, метод фазового інтервалу, концепція нечітких множин, логічне програмування або нейронні мережі .

Використання математичних методів класифікації можливе лише в тому випадку, коли підтверджується так звана гіпотеза компактності, тобто зображення об'єктів одного і того ж класу займає в просторі ознак деяку компактну область. Найпростішим методом класифікації є знаходження аналітичних виразів деяких функцій, за допомогою яких можна розділити ці області в просторі ознак. Такі функції називаються вирішувальними.

В розділі 1 магістерської роботи зазначалось, що сьогодні здобувають все більшого поширення перцептронні системи розпізнавання, в основі яких використовується природна модель розпізнавання образів нейронною мережею людського мозку.

Модель нейронної мережі з трьома шарами подана на рис.3.4.

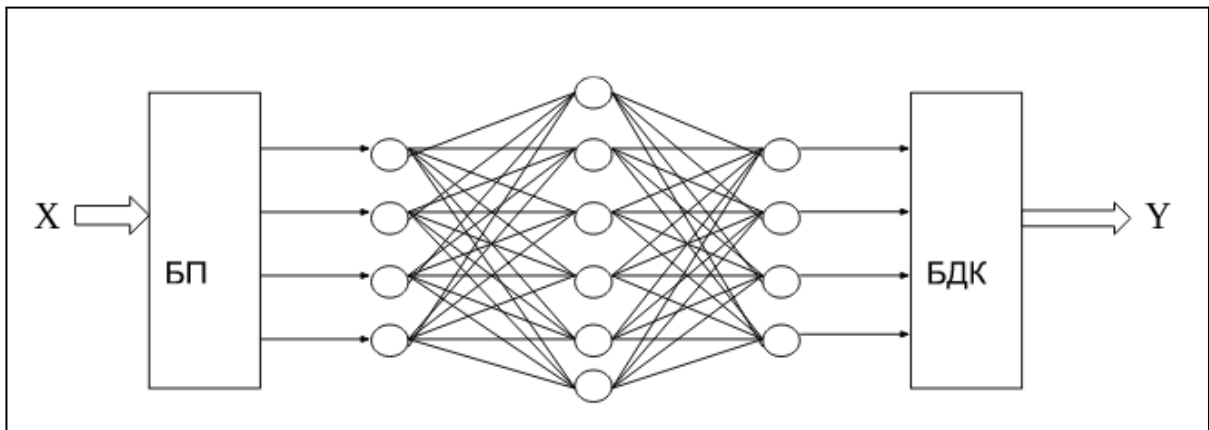


Рисунок 3.4 – Багатошаровий персептрон для розпізнавання мовних образів

Для упорядкування взаємодії між нейронною мережею та навколишнім оточенням використовуються блоки параметризації (БП) та декодування (БДК), що служать для передачі інформації про звуковий об'єкт (диктор). З метою опису обчислювальної моделі нейронної мережі можна використовувати матриці ваг міжнейронних зв'язків.

Ваги міжнейронних зв'язків представляються як матриці наступного виду [34]:

$$W^{\mu} = \{W_{ij}^{\mu}\}, \mu = \overline{1, K}, i = \overline{1, N_{\mu}}, j = \overline{1, N_{\mu-1}}, \quad (3.7)$$

де  $\mu$  – номер матриці міжнейронних зв'язків;

$W_{ij}^{\mu}$  – вага зв'язку між  $j$ -тим нейроном вихідного шару і  $i$ -того нейрона

вихідного шару для  $\mu$ - тої матриці зв'язків;

$N_{\mu}$  – число нейронів  $\mu$ -го шару;

$N_{\mu-1}$  – число нейронів  $\mu-1$ -го шару.

Задача ідентифікації звукових образів нейронними мережами може бути подана в такому вигляді.

Дано:  $(X^P, \hat{y}^P)$ ,  $p = \overline{1, P}$  – навчальна вибірка у вигляді  $P$  пар типу input-output, де

$$X^P = \{x_1^P, x_2^P, \dots, x_n^P\} - \text{вхідний вектор};$$

$$\hat{y}^P - \text{вихідний параметр/}$$

Необхідно: визначити такі параметри нейронної мережі, при яких забезпечується мінімальне відхилення моделі від об'єкта:

$$\sum_{p=1}^P (F(W^1, W^1, \dots, W^K, x_1^P, x_2^P, \dots, x_n^P) - \hat{y}^P)^2 \rightarrow \min \quad (3.8)$$

Вирішення задачі оптимізації (3.8) відбувається за допомогою процедури навчання нейронної мережі. На сьогодні здобув великого поширення метод навчання нейронних мереж зі зворотнім поширенням помилки.

Основна відмінність та перевага нейромереж порівняно з традиційними методами прогнозування та класифікації полягають у їхній здатності до навчання. Процес навчання передбачає обчислення синаптичних коефіцієнтів шляхом розв'язання задач нейромережею, де відповідь формується не на основі правил, а за допомогою прикладів, об'єднаних у навчальні множини. Таким чином, нейромережа під час навчання виконує функцію експерта у підготовці даних для побудови експертної системи.

Для тренування нейронної мережі необхідні відповідні навчальні дані, які можуть включати в себе характеристики представлення та елементи випадковості або послідовності. Характер цих даних залежить від конкретної класифікації завдання, що вирішується. Навчальні дані представляють собою набір прикладів, де для кожного з них надається значення вихідного параметра, яке ми бажаємо отримати. Цей процес можна охарактеризувати як контрольоване навчання, де "вчитель" надає мережі вхідний вектор даних і

вказує очікуване значення результату обчислень на виході.

Процес контрольованого навчання нейромережі можна розглядати як вирішення оптимізаційної задачі, спрямованої на мінімізацію функції помилки  $E$  для даної множини прикладів через вибір оптимальних значень ваг  $W$ . Досягнення мінімуму цієї функції відоме як збіжність процесу навчання. Розенблат розглядав цю можливість, адже через нелінійний характер залежності помилки від ваг, аналітичне отримання рішення стає неможливим. Отже, глобальний мінімум функції шукається за допомогою ітераційного процесу, відомого як навчальний алгоритм. Існує багато різних навчальних алгоритмів, які відрізняються стратегією оптимізації та критерієм помилки. Зазвичай використовується середньоквадратична помилка (СКП) як міра похибки.

$$E = \sqrt{\frac{\sum_{i=1}^M (d_i - y_i)^2}{M}} \quad (3.9)$$

де  $M$  - кількість прикладів у навчальній множині.

Мінімізація значення  $E$  виконується шляхом використання градієнтних методів. Зміни у вагах проводяться в напрямку, протилежному напрямку найбільшого нахилу функції:

$$W(t + 1) = W(t) + \varepsilon * \frac{\partial E}{\partial W} \quad (3.10)$$

де  $\varepsilon$  – коефіцієнт навчання.

Правило Хебба для навчання мережі передбачає збільшення ваги активного входу нейрона у випадку, коли його вихід має бути активним, і зменшення ваги, якщо вихід повинен бути неактивним. Це правило відзначається вищим пріоритетом над іншими методами навчання, включаючи широко вживаний на сьогодні алгоритм зворотного поширення



помилки.

Один з найпоширеніших методів навчання нейромереж прямого поширення включає в себе алгоритм зворотного поширення помилки (Error Back Propagation – ЕВР). Цей алгоритм, відкритий і популяризований у 1986 році Румельхартом та МакЛеландом з групи паралельно розподілених процесів у Массачусетському технологічному інституті, є методом градієнтного спуску, спрямованим на мінімізацію загальної квадратичної помилки.

Згідно з принципом методу найменших квадратів [40-43], функціонал помилки (НМ), який прагне знизити своє значення визначається так:

$$E(\omega) = \frac{1}{2} \sum_{j,p} (y_{j,p}^{(N)} - d_{j,p})^2 \quad (3.11)$$

де  $y_{j,p}^{(N)}$  – вихідний сигнал нейрона  $j$  вихідного шару  $N$  нейромережі при подачі на її вхід  $p$ -го образу;  $d_{j,p}$  – очікуваний (бажаний) вихідний сигнал для цього нейрона.

Здійснюється сумування по всіх нейронах вихідного шару та оброблюваних мережею образах. Процес мінімізації здійснюється за допомогою методу градієнтного спуску, що включає в себе коригування вагових коефіцієнтів:

$$\Delta \omega_{ij}^{(n)} = - \eta * \frac{\partial E}{\partial W_{ij}} \quad (3.12)$$

У даному контексті  $w_{ij}$  – представляє ваговий коефіцієнт синаптичного зв'язку, між  $i$ -тим нейроном шару  $n-1$  із  $j$ -им нейроном шару  $n$ , де  $\eta$  – є коефіцієнтом швидкості навчання і набуває значень в межах від 0 до 1.

Як наведено вище,

$$\frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial y_j} * \frac{\partial y_j}{\partial s_{ij}} * \frac{\partial s_{ij}}{\partial W_{ij}} \quad (3.13)$$

У цьому контексті  $y_j$ , вказує на вихід нейрона  $j$ , а  $s_j$  – зважена сума його вхідних сигналів, іншими словами, сума, вагована відповідно до активаційної функції. Оскільки  $dy_j/ds_j$  є похідною цієї функції за її аргументом, це означає, що похідна активаційної функції має бути визначена на всьому діапазоні значень аргументу. З цієї причини функції з неоднорідностями, такі як функція одиничного стрибка та інші, не є придатними для розглядуваних нейромереж. У цьому випадку використовують гладкі функції, такі як гіперболічний тангенс або класичний сигмоїд з експонентою. Якщо взяти, наприклад, гіперболічний тангенс [44]

$$\frac{dy}{ds} = 1 - s^2 \quad (3.14)$$

Третій множник  $\partial s_j / \partial w_{ij}$ , мабуть, дорівнює виходові нейрона з попереднього шару  $y_i^{(n-1)}$ .

Що стосується першого множника в (3.13), він легко розкладається в такий спосіб:

$$\frac{\partial E}{\partial y_{ij}} = \sum_k \frac{\partial E}{\partial y_k} * \frac{\partial y_k}{\partial s_k} * \frac{\partial s_k}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} * \frac{\partial y_k}{\partial s_k} * w_{jk}^{(m+1)} \quad (3.15)$$

В цьому місці сумовування по  $k$  виконується серед нейронів з шару  $n+1$ .

Додавши нову змінну: ми отримуємо формулу, яка використовується для ітеративних обчислень величин  $\delta_j^{(n)}$  шару  $n$  із величин  $\delta_k^{(n+1)}$  старшого шару  $n+1$ .

$$\delta_j^{(n)} = \frac{\partial E}{\partial y_j} * \frac{dy_j}{ds_j} \quad (3.16)$$

$$\delta_j^{(n)} = \left[ \sum_k \delta_k^{(n+1)} * w_{jk}^{(n+1)} \right] * \frac{dy_j}{ds_j} \quad (3.17)$$

Для вихідного шару:

$$\delta_1^{(N)} = (y_1^{(N)} - d_1) * \frac{dy_1}{ds_1} \quad (3.18)$$

Тепер ми можемо записати (3.8) у розкритому вигляді:

$$\Delta\omega_{ij}^{(n)} = -\eta * \delta_j^{(n)} * y_i^{(n-1)} \quad (3.19)$$

Часом, для усунення різких коливань під час корекції ваг, пов'язаних із переміщенням по поверхні цільової функції, (3.19) доповнюється значенням зміни ваги на попередньому кроці ітерації.

$$\Delta\omega_{ij}^{(n)} = -\eta(\mu * \Delta\omega_{ij}^{(n)}(t-1) + (1-\mu) * \delta_j^{(n)} * y_i^{(n-1)}) \quad (3.20)$$

де  $\mu$  – коефіцієнт інерційності,  $t$  - номер поточної ітерації.

Отже, вся процедура навчання нейромережі за допомогою алгоритму зворотного поширення помилки конструюється наступним чином:

1. Подавши на входи мережі один із можливих сценаріїв, а система працюючи в нормальному режимі, коли сигнали передаються від входів до виходів, обчислити значення останніх. Нагадаємо, що

$$S_j^{(n)} = \sum_{i=0}^M y_i^{(n-1)} * w_{ij}^{(n)} \quad (3.21)$$

де  $M$  – число нейронів із шару  $n-1$ , враховуючи нейрон з постійним вихідним станом  $+1$ , що задає зсув;  $y_i^{(n-1)} = x_{ij}^{(n)}$  -  $i$ -ий вхід нейрона  $j$  шару  $n$ .

$$y_j^{(n)} = f(s_j^{(n)}), \quad (3.22)$$

$$y_q^{(0)} = I_q, \quad (3.23)$$

де  $f()$  – сигмоїд;

$I_q$  -  $q$ -а компонента вектора вхідного образу.

2. Обрахувати  $\delta^{(N)}$  для вихідного шару по формулі (3.18).
3. Розрахувати по формулі (3.19) або (3.20) зміни ваг  $\Delta w^{(N)}$  шару  $N$ .
4. Розрахувати по формулах (3.17) і (3.19) (або (3.16) і (3.20))  $\delta^{(N)}$  і  $\Delta w^{(n)}$  для всіх інших шарів,  $n = N-1, \dots, 1$ .
5. Скорегувати усі ваги в НМ

$$\omega_{ij}^{(n)}(t) = \omega_{ij}^{(n)}(t-1) + \Delta \omega_{ij}^{(n)}(t). \quad (3.24)$$

6. У разі значущої помилки мережі перехід на перший крок є необхідним. У випадку, коли помилок незначно або ж їх немає, завершити процес.

На першому етапі, мережі представляють усі тренувальні приклади у випадковому порядку, створюючи умови, при яких мережа, метафорично кажучи, зберігає попередні враження, навіть коли засвоює нову інформацію.

З виразу (3.19) видно, що ефективність навчання помітно знижується, коли значення  $u_i^{(n-1)}$ , отримане на виході, наближається до нуля. У випадку двійкових вхідних векторів половина вагових коефіцієнтів, у середньому, не піддається коригуванню. Тому для оптимізації області можливих значень виходів нейронів з  $[0,1]$  до  $[-0.5, +0.5]$  рекомендується використовувати модифікації логістичних функцій. Наприклад, можна змінити сигмоїд з експонентою, щоб отримати:

$$f(x) = -0.5 + \frac{1}{1+e^{-a*x}} \quad (3.25)$$

Схема алгоритму наведена на рис. 3.5.

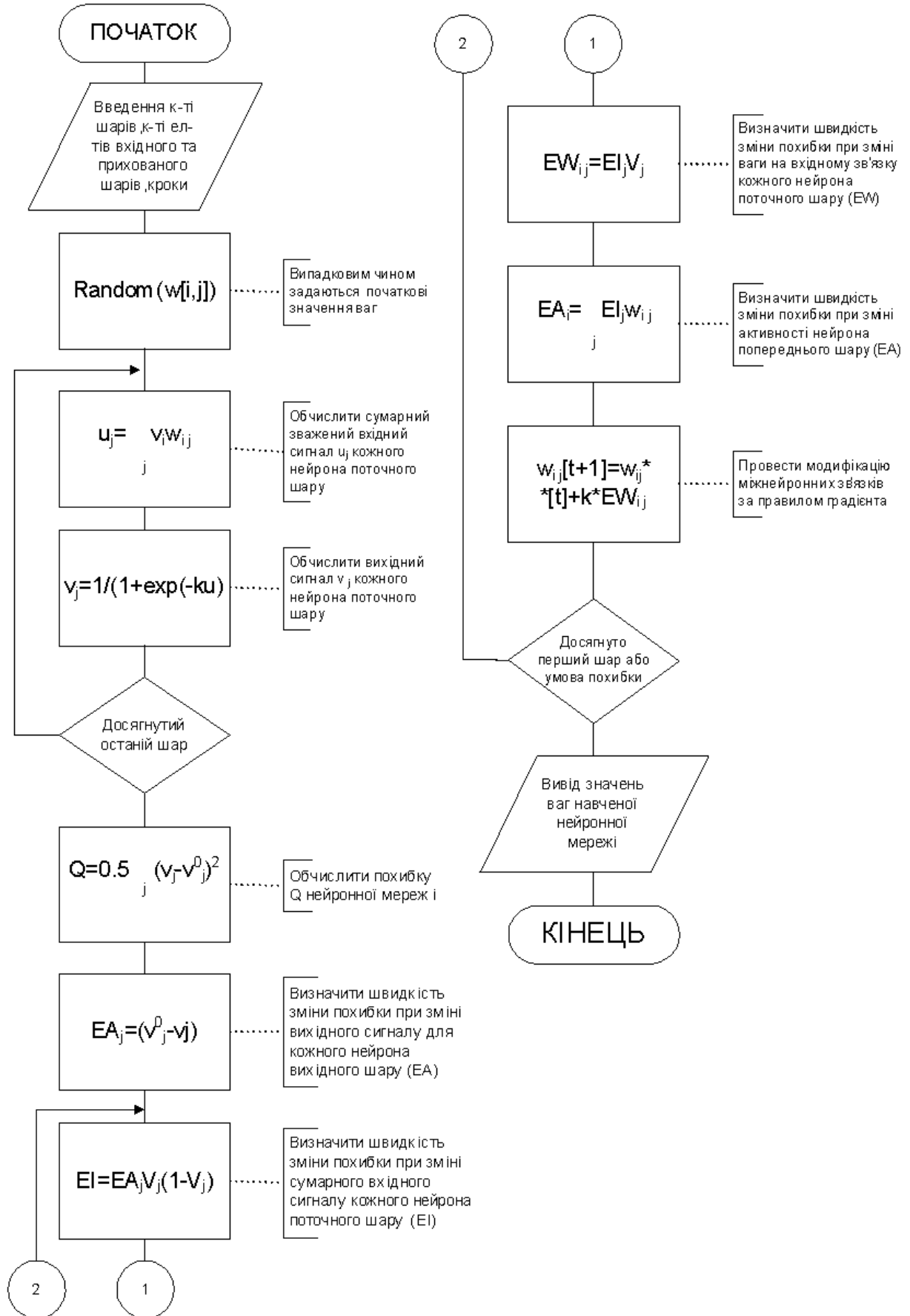


Рисунок 3.5 – Схема алгоритму навчання нейромережевого класифікатора

### 3.5 Машинний експеримент

Як було вказано в п.3.4 роботи для реалізації класифікатора використовуємо тришарову нейромережу.

У розробленій системі автоматичної ідентифікації диктора, мовний матеріал (парольна фраза, вимовлена диктором) піддається аналізу на підставі таких характеристик:

- інтегральний спектр;
- частота основного тону.

Отже, має сенс розробити індивідуальні нейромережі для вирішення завдання ідентифікації особи диктора на основі кожної з наведених характеристик.

Аналізуючи запис парольної фрази за допомогою алгоритмів виділення ознак, розроблених в ході роботи, одержуємо вектор параметрів, що відповідає запису парольної фрази:

$$v_i = \{IS(1, 2, n_{IS}), AKS(1, 2, n_{AKS}), VF(1, 2, n_{VF}), KL(1, 2, n_{KL}), LP(1, 2, n_{LP})\}$$

де  $i$  – номер диктора,  $IS(1, 2, n_{IS})$  – вектор параметрів, що одержуємо аналізуючи запис парольної фрази за інтегральним спектром,  $n_{IS}$  – довжина цього вектора;  $AKS(1, 2, n_{AKS})$  – вектор параметрів, що одержуємо аналізуючи запис парольної фрази за амплітудними значеннями короточасних спектрів на інтервалах стаціонарності сигналу,  $n_{AKS}$  – довжина цього вектора і т.д.

На вхід кожної нейромережі буде подаватися відповідна частина вектора параметрів  $v_i$ .

Далі, для продемонстрування ефективності системи, здійснено процес навчання та тестування нейромереж для розпізнавання 10 дикторів за

парольним словом "пароль". Записи вимови даного слова були виконані кожним диктором по 30 разів для формування навчальної вибірки, а також окремо по 5 разів для отримання тестової вибірки. Для кожної нейромережі проведено аналіз залежності точності ідентифікації  $T$  та часу навчання  $t$  від кількості нейронів в прихованому шарі нейромережі  $N_2$  та метода навчання.

Точність ідентифікації  $T$  визначається за формулою

$$T = (1 - \frac{k}{50}) * 100, \quad (3.26)$$

де  $k$  – кількість помилок навченої нейромережі.

Для експерименту використано комп'ютер з такими технічними характеристиками: CPU Intel I3-380M, 4GB DDR3, 500GB HDD, операційна система – MS Windows 10.

Результати проведеного експерименту і їх аналіз наведено нижче.

Проводиться аналіз амплітудних характеристик короткочасних спектрів у 12 смугах полосового фільтра, який рівномірно охоплює частотний проміжок  $f_{min} / f_{max}$  Гц без перекриття вікон (властивості смуг фільтра можна змінювати). Частоти, що визначають межі кожної смуги, розраховуються за таким способом:

$$f_{\text{початкове}} = (n - 1) * K + f_{min}, f = n * K + f_{max} \quad (3.27)$$

де  $K = \frac{f_{min} - f_{max}}{n}$ ,  $n$  – це номер смуги фільтра.

Кількість нейронів на вхідному шарі нейромережі  $N_1$  відповідає кількості фільтрів, тобто  $N_1 = 12$ .

Початково кількість нейронів на прихованому шарі взято  $N_2 = 20$ .

На вихідному шарі нейромережі кількість нейронів  $N_3$  відповідає кількості дикторів для ідентифікації яких нейромережа навчається (відповідно до умови експерименту  $N_3 = 10$ ).

Нейромережу реалізовано у середовищі програмування Matlab. Згідно умовам експерименту, навчальний набір створений із 30 стандартів для кожного з 10 говорителів. Ці стандарти виникли через виділення амплітудних значень короткочасних спектрів записів парольних фраз на 12 смугах фільтра. Розміри навчального набору 12x300.

Для тестування навченої нейромережі сформовано матрицю розмірами 12x50, одержану таким же чином.

Результати тестування наведено в табл. 3.2.

Таблиця 3.2 – Результати експерименту з навчання і тестування мережі для ідентифікації особи за короткочасними спектрами

$N_2$	Метод навчання	t, c	T, %
10	Gradient descent backpropagation	124	82
	Gradient descent with momentum backpropagation	114	82
	Levenberg-Marquardt backpropagation	96	63
	RPROP backpropagation	103	83
20	Gradient descent backpropagation	142	84
	Gradient descent with momentum backpropagation	128	83
	Levenberg-Marquardt backpropagation	107	64
	RPROP backpropagation	126	86
30	Gradient descent backpropagation	149	84
	Gradient descent with momentum backpropagation	132	84
	Levenberg-Marquardt backpropagation	112	63
	RPROP backpropagation	126	86

Аналізуючи дані таблиці 3.2 робимо висновок, що найкраще відношення показників час навчання/точність ідентифікації має метод RPROP backpropagation. Його середня точність ідентифікації – 86%.

Нейромережу реалізовано в програмному середовищі Matlab.

В таблиці 3.3 наведено результати навчання та тестування нейромережі з використанням додаткового в якості ознаки частоти основного тону.



Таблиця 3.3 – Результати експерименту з навчання та тестування нейромережі для ідентифікації особи за частотою основного тону

$N_2$	Метод навчання	t, с	T, %
10	Levenberg-Marquardt backpropagation	68	62
	Gradient descent backpropagation	112	89
	Gradient descent with momentum backpropagation	108	89
	RPROP backpropagation	105	90
20	Levenberg-Marquardt backpropagation	67	68
	Gradient descent backpropagation	115	92
	Gradient descent with momentum backpropagation	108	93
	RPROP backpropagation	116	95
30	Levenberg-Marquardt backpropagation	69	65
	Gradient descent backpropagation	121	94
	Gradient descent with momentum backpropagation	117	94
	RPROP backpropagation	119	96

Як видно з таблиці 3.3 найкращі результати за точністю ідентифікації отримані навчанням нейромережі з використанням методу RPROP backpropagation (96%).

Таким чином, аналіз експериментальних даних, наведених в таблицях 3.2 і 3.3 можна зробити висновок, що найбільш інформативними ознаками для опису голосу мовця в задачі ідентифікації є короткочасний спектр і частота основного тону.

### 3.6 Висновки до розділу

У цьому розділі висвітлено математичні основи методів виділення спектральних ознак та частоти основного тону, а також методу навчання нейромережевого класифікатора. З метою реалізації цих методів були створені відповідні алгоритми, призначені для їх ефективної реалізації. В рамках даного дослідження було розроблено докладну структуру програмного забезпечення автоматизованої системи ідентифікації особи за допомогою голосу.

В склад програмного забезпечення автоматизованої системи ідентифікації особи за голосом включені різноманітні програми, серед яких можна виокремити програми виділення ознак частоти основного тону, програму виділення спектральних характеристик сигналу мовця, програму навчання нейромережевого класифікатора та програму розпізнавання голосу мовця з використанням нейромережевого класифікатора. Розроблені програмні модулі були втілені в середовищі Matlab, а тексти програм зазначені у додатку В, який додається до даного документу.

## 4 ЕКОНОМІЧНА ЧАСТИНА

Науково-технічна розробка має право на існування та впровадження, якщо вона відповідає вимогам часу, як в напрямку науково-технічного прогресу та і в плані економіки. Тому для науково-дослідної роботи необхідно оцінювати економічну ефективність результатів виконаної роботи.

Магістерська кваліфікаційна робота з розробки та дослідження «Автоматизована система ідентифікації особи за її голосом» відноситься до науково-технічних робіт, які орієнтовані на виведення на ринок (або рішення про виведення науково-технічної розробки на ринок може бути прийнято у процесі проведення самої роботи), тобто коли відбувається так звана комерціалізація науково-технічної розробки. Цей напрямок є пріоритетним, оскільки результатами розробки можуть користуватися інші споживачі, отримуючи при цьому певний економічний ефект. Але для цього потрібно знайти потенційного інвестора, який би взявся за реалізацію цього проекту і переконати його в економічній доцільності такого кроку.

Для наведеного випадку нами мають бути виконані такі етапи робіт:

- 1) проведено комерційний аудит науково-технічної розробки, тобто встановлення її науково-технічного рівня та комерційного потенціалу;
- 2) розраховано витрати на здійснення науково-технічної розробки;
- 3) розрахована економічна ефективність науково-технічної розробки у випадку її впровадження і комерціалізації потенційним інвестором і проведено обґрунтування економічної доцільності комерціалізації потенційним інвестором.

#### 4.1 Проведення комерційного та технологічного аудиту науково-технічної розробки

Метою проведення комерційного і технологічного аудиту дослідження за темою «Автоматизована система ідентифікації особи за її голосом» є оцінювання науково-технічного рівня та рівня комерційного потенціалу розробки, створеної в результаті науково-технічної діяльності.

Оцінювання науково-технічного рівня розробки та її комерційного потенціалу рекомендується здійснювати із застосуванням 5-ти бальної системи оцінювання за 12-ма критеріями, наведеними в табл. 4.1 [52].

Таблиця 4.1 – Рекомендовані критерії оцінювання науково-технічного рівня і комерційного потенціалу розробки та бальна оцінка

Бали (за 5-ти бальною шкалою)					
	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 працездатність 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

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
Ринкові перспективи					
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	Конкуренти в немає
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
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
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

Результати оцінювання науково-технічного рівня та комерційного потенціалу науково-технічної розробки потрібно звести до таблиці 4.2.

За результатами розрахунків, наведених в таблиці 4.2, зробимо висновок щодо науково-технічного рівня і рівня комерційного потенціалу розробки.

Таблиця 4.2 – Результати оцінювання науково-технічного рівня і комерційного потенціалу розробки експертами

Критерії	Експерт (ПІБ, посада)		
	1	2	3
	Бали:		
1. Технічна здійсненність концепції	4	3	4
2. Ринкові переваги (наявність аналогів)	3	3	2
3. Ринкові переваги (ціна продукту)	3	2	3
4. Ринкові переваги (технічні властивості)	2	3	3
5. Ринкові переваги (експлуатаційні витрати)	2	2	2
6. Ринкові перспективи (розмір ринку)	2	2	2
7. Ринкові перспективи (конкуренція)	2	2	2
8. Практична здійсненність (наявність фахівців)	4	4	4
9. Практична здійсненність (наявність фінансів)	2	3	2
10. Практична здійсненність (необхідність нових матеріалів)	2	2	2
11. Практична здійсненність (термін реалізації)	3	4	4
12. Практична здійсненність (розробка документів)	4	4	4
Сума балів	33	34	34
Середньоарифметична сума балів $СБ_c$	33,7		

При цьому використаємо рекомендації, наведені в табл. 4.3 [52].

Таблиця 4.3 – Науково-технічні рівні та комерційні потенціали розробки

Середньоарифметична сума балів СБ , розрахована на основі висновків експертів	Науково-технічний рівень та комерційний потенціал розробки
41...48	Високий
31...40	Вище середнього
21...30	Середній
11...20	Нижче середнього
0...10	Низький

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Автоматизована система ідентифікації особи за її голосом» становить 33,7 бала, що, відповідно до таблиці 4.3, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки вище середнього).

#### 4.2 Розрахунок узагальненого коефіцієнта якості розробки

Окрім комерційного аудиту розробки доцільно також розглянути технічний рівень якості розробки, розглянувши її основні технічні показники. Ці показники по-різному впливають на загальну якість проектної розробки.

Узагальнений коефіцієнт якості ( $B_H$ ) для нового технічного рішення розраховуємо за формулою [53]:

$$B_H = \sum_{i=1}^k a_i * \beta_i \quad (4.1)$$

де  $k$  – кількість найбільш важливих технічних показників, які впливають на якість нового технічного рішення;

$\alpha_i$  – коефіцієнт, який враховує питому вагу  $i$ -го технічного показника в загальній якості розробки. Коефіцієнт  $\alpha_i$  визначається експертним

шляхом і при цьому має виконуватись умова  $\sum_{i=1}^k \alpha_i = 1$ ;

$\beta_i$  – відносне значення  $i$ -го технічного показника якості нової розробки.

Відносні значення  $\beta_i$  для різних випадків розраховуємо за такими формулами:

- для показників, зростання яких вказує на підвищення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{ni}}{I_{ai}} \quad (4.2)$$

де  $I_{ni}$  та  $I_{ai}$  – чисельні значення конкретного  $i$ -го технічного показника якості відповідно для нової розробки та аналога;

- для показників, зростання яких вказує на погіршення в лінійній залежності якості нової розробки:

$$\beta_i = \frac{I_{ai}}{I_{ni}} \quad (4.3)$$

Використовуючи наведені залежності можемо проаналізувати та порівняти техніко-економічні характеристики аналогу та розробки на основі отриманих наявних та проектних показників, а результати порівняння зведемо до таблиці 4.4.

Таблиця 4.4 – Порівняння основних параметрів розробки та аналога.

Показники (параметри)	Одиниця вимірювання	Аналог	Проектований пристрій	Відношення параметрів нової розробки до аналога	Питома вага показника
Час ідентифікації	с	2,5	1,8	1,39	0,25



Точність ідентифікації	%	95	98	1,03	0,3
Стійкість до шуму	бал	7	7	1	0,15
Споживана потужність	Вт	100	70	1,43	0,2
Робоча відстань	М	1...3	1...3	1	0,1

Узагальнений коефіцієнт якості ( $B_H$ ) для нового технічного рішення складе:

$$B_H = \sum_{i=1}^k a_i * \beta_i = 1,39 \cdot 0,25 + 1,03 \cdot 0,3 + 1 \cdot 0,15 + 1,43 \cdot 0,2 + 1 \cdot 0,1 = 1,19.$$

Отже за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,19 рази.

#### 4.3 Розрахунок витрат на проведення науково-дослідної роботи

Витрати, пов'язані з проведенням науково-дослідної роботи на тему «Автоматизована система ідентифікації особи за її голосом», під час планування, обліку і калькулювання собівартості науково-дослідної роботи групуємо за відповідними статтями.

##### 4.3.1 Витрати на оплату праці

До статті «Витрати на оплату праці» належать витрати на виплату основно та додаткової заробітної плати керівникам відділів, лабораторій, секторів і груп, науковим, інженерно-технічним працівникам, конструкторам, технологам, креслярам, копіювальникам, лаборантам, робітникам, студентам, аспірантам та іншим працівникам, безпосередньо зайнятим виконанням

конкретної теми, обчисленої за посадовими окладами, відрядними розцінками, тарифними ставками згідно з чинними в організаціях системами оплати праці.

Основна заробітна плата дослідників

Витрати на основну заробітну плату дослідників ( $Z_o$ ) розраховуємо у відповідності до посадових окладів працівників, за формулою [52]:

$$Z_o = \sum_{i=1}^k \frac{M_{ni} * t_i}{T_p} \quad (4.4)$$

де  $k$  – кількість посад дослідників залучених до процесу досліджень;

$M_{ni}$  – місячний посадовий оклад конкретного дослідника, грн;

$t_i$  – число днів роботи конкретного дослідника, дн.;

$T_p$  – середнє число робочих днів в місяці,  $T_p=21$  дні.

$$Z_o = 17210,00 \cdot 26 / 21 = 21307,62 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.5 – Витрати на заробітну плату дослідників

Найменування посади	Місячний посадовий оклад, грн	Оплата за робочий день, грн	Число днів роботи	Витрати на заробітну плату, грн
Керівник проекту	17210,00	819,52	26	21307,62
Інженер-розробник програмного забезпечення	16670,00	793,81	26	20639,05
Провідний фахівець дослідження біометричних характеристик людини	16050,00	764,29	5	3821,43
Всього				45768,10

Основна заробітна плата робітників

Витрати на основну заробітну плату робітників ( $Z_p$ ) за відповідними найменуваннями робіт НДР на тему «Автоматизована система ідентифікації особи за її голосом» розраховуємо за формулою:

$$Z_p = \sum_{i=1}^n C_i * t_i, \quad (4.5)$$

де  $C_i$  – погодинна тарифна ставка робітника відповідного розряду, за виконану відповідну роботу, грн/год;

$t_i$  – час роботи робітника при виконанні визначеної роботи, год.

Погодинну тарифну ставку робітника відповідного розряду  $C_i$  можна визначити за формулою:

$$C_i = \frac{M_M * K_i * K_c}{T_p * t_{зм}} \quad (4.6)$$

де  $M_M$  – розмір прожиткового мінімуму працездатної особи, або мінімальної місячної заробітної плати (в залежності від діючого законодавства), прийmemo  $M_M=6700,00$  грн;

$K_i$  – коефіцієнт міжкваліфікаційного співвідношення для встановлення тарифної ставки робітнику відповідного розряду (табл. Б.2, додаток Б) [52];

$K_c$  – мінімальний коефіцієнт співвідношень місячних тарифних ставок робітників першого розряду з нормальними умовами праці виробничих об'єднань і підприємств до законодавчо встановленого розміру мінімальної заробітної плати.

$T_p$  – середнє число робочих днів в місяці, приблизно  $T_p = 21$  дн;

$t_{зм}$  – тривалість зміни, год.

$$C_i = 6700,00 \cdot 1,10 \cdot 1,35 / (21 \cdot 8) = 59,22 \text{ грн.}$$

$$Z_{pl} = 59,22 \cdot 7,30 = 432,33 \text{ грн.}$$

Таблиця 4.6 – Величина витрат на основну заробітну плату робітників

Найменування робіт	Тривалість роботи, год	Розряд роботи	Тарифний коефіцієнт	Погодинна тарифна ставка, грн	Величина оплати на робітника грн
--------------------	------------------------	---------------	---------------------	-------------------------------	----------------------------------

Установка обладнання для розробки програмного забезпечення	7,30	2	1,10	59,22	432,33
Підготовка робочого місця розробника програмного забезпечення	6,10	3	1,35	72,68	443,37
Інсталяція програмного забезпечення для розпізнавання зображення	5,65	5	1,70	91,53	517,13
Компіляція програмних блоків	4,00	4	1,50	80,76	323,04
Налагодження програмних блоків	4,30	5	1,70	91,53	393,57
Формування бази даних результатів дослідження	12,00	4	1,50	80,76	969,11
Всього					3078,53

Додаткова заробітна плата дослідників та робітників

Додаткову заробітну плату розраховуємо як 10 ... 12% від суми основної заробітної плати дослідників та робітників за формулою:

$$Z_{\text{дод}} = (Z_o + Z_p) * \frac{H_{\text{дод}}}{100\%} \quad (4.7)$$

де  $H_{\text{дод}}$  – норма нарахування додаткової заробітної плати. Прийmemo 11%.

$$Z_{\text{дод}} = (45768,10 + 3078,53) \cdot 11 / 100\% = 5373,13 \text{ грн.}$$

### 4.3.2 Відрахування на соціальні заходи

Нарахування на заробітну плату дослідників та робітників розраховуємо як 22% від суми основної та додаткової заробітної плати дослідників і робітників за формулою:

$$З_{\text{н}} = (З_{\text{о}} + З_{\text{р}} + З_{\text{дод}}) * \frac{H_{\text{зп}}}{100\%} \quad (4.8)$$

де  $H_{\text{зп}}$  – норма нарахування на заробітну плату. Приймаємо 22%.

$$З_{\text{н}} = (45768,10 + 3078,53 + 5373,13) \cdot 22 / 100\% = 11928,35 \text{ грн.}$$

### 4.3.3 Сировина та матеріали

До статті «Сировина та матеріали» належать витрати на сировину, основні та допоміжні матеріали, інструменти, пристрої та інші засоби і предмети праці, які придбані у сторонніх підприємств, установ і організацій та витрачені на проведення досліджень за темою «Автоматизована система ідентифікації особи за її голосом».

Витрати на матеріали ( $M$ ), у вартісному вираженні розраховуються окремо по кожному виду матеріалів за формулою:

$$M = \sum_{j=1}^n H_{\beta j} \cdot Ц_j \cdot K_j - \sum_{j=1}^n B_j \cdot Ц \quad (4.9)$$

де  $H_j$  – норма витрат матеріалу  $j$ -го найменування, кг;

$n$  – кількість видів матеріалів;

$Ц_j$  – вартість матеріалу  $j$ -го найменування, грн/кг;

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

$B_j$  – маса відходів  $j$ -го найменування, кг;

$Ц_{\beta j}$  – вартість відходів  $j$ -го найменування, грн/кг.

$$M_1 = 4,0 \cdot 206,00 \cdot 1,01 - 0 \cdot 0 = 832,24 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.7 – Витрати на матеріали

Найменування матеріалу, марка, тип, сорт	Ціна за 1 кг, грн	Норма витрат, кг	Величина відходів, кг	Ціна відходів, грн/кг	Вартість витраченого матеріалу, грн
Папір канцелярський офісний (А4)	206,00	4,0	0	0	832,24
Папір для заміток (А5)	120,00	3,0	0	0	363,60
Начиння канцелярське	178,00	3,0	0	0	539,34
Органайзер офісний	199,00	4,0	0	0	803,96
Картридж для принтера	2018,00	1,0	0	0	2038,18
Всього					4577,32

#### 4.3.4 Розрахунок витрат на комплектуючі

Витрати на комплектуючі ( $K_6$ ), які використовують при проведенні НДР на тему «Автоматизована система ідентифікації особи за її голосом», розраховуємо, згідно з їхньою номенклатурою, за формулою:

$$K_6 = \sum_{j=1}^n H_j * C_j * K_j \quad (4.10)$$

де  $H_j$  – кількість комплектуючих  $j$ -го виду, шт.;

$C_j$  – покупна ціна комплектуючих  $j$ -го виду, грн;

$K_j$  – коефіцієнт транспортних витрат, ( $K_j = 1,1 \dots 1,15$ ).

$$K_6 = 1 \cdot 3079,00 \cdot 1,01 = 3109,79 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.8 – Витрати на комплектуючі

Найменування комплектуючих	Кількість, шт.	Ціна за штуку, грн	Сума, грн
Зовнішній жорсткий диск 2.5" 2TB Seagate (STGD2000200)	1	3079,00	3109,79
Мікрофон Fifine A6V AMPLIGAME	1	1799,00	1816,99
Підсилювач	1	2780,00	2807,80
Лінії затримки	1	1899,00	1917,99
Формувач сигналу	1	980,00	989,80
Дешифратор	1	1200,00	1212,00
Аналого-цифровий перетворювач	1	1545,00	1560,45
Всього			13414,82

#### 4.3.5 Спецустаткування для наукових (експериментальних) робіт

До статті «Спецустаткування для наукових (експериментальних) робіт» належать витрати на виготовлення та придбання спецустаткування необхідного для проведення досліджень, також витрати на їх проектування, виготовлення, транспортування, монтаж та встановлення.

Балансову вартість спецустаткування розраховуємо за формулою:

$$V_{\text{спец}} = \sum_{i=1}^k \Pi_i * C_{\text{пр.і}} * K_i \quad (4.11)$$

де  $\Pi_i$  – ціна придбання одиниці спецустаткування даного виду, марки, грн;

$C_{\text{пр.і}}$  – кількість одиниць устаткування відповідного найменування, які придбані для проведення досліджень, шт.;

$K_i$  – коефіцієнт, що враховує доставку, монтаж, налагодження устаткування тощо, ( $K_i = 1, 10 \dots 1, 12$ );

$k$  – кількість найменувань устаткування.

$$B_{\text{спец}} = 12400,00 \cdot 1 \cdot 1,01 = 12524,00 \text{ грн.}$$

Отримані результати зведемо до таблиці:

Таблиця 4.9 – Витрати на придбання спецустаткування по кожному виду

Найменування устаткування	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Осцилограф цифровий ОСЦ-2000	1	12400,00	12524,00
Аналізатор спектру	1	10250,00	10352,50
Всього			22876,50

#### 4.3.6 Програмне забезпечення для наукових робіт

До статті «Програмне забезпечення для наукових (експериментальних) робіт» належать витрати на розробку та придбання спеціальних програмних засобів і програмного забезпечення, (програм, алгоритмів, баз даних) необхідних для проведення досліджень, також витрати на їх проектування, формування та встановлення.

Балансову вартість програмного забезпечення розраховуємо за формулою:

$$B_{\text{прг}} = \sum_{i=1}^k C_{\text{іпрг}} * C_{\text{прг.і}} * K_i \quad (4.12)$$

де  $C_{\text{іпрг}}$  – ціна придбання одиниці програмного засобу даного виду, грн;

$C_{\text{прг.і}}$  – кількість одиниць програмного забезпечення відповідного найменування, які придбані для проведення досліджень, шт.;

$K_i$  – коефіцієнт, що враховує інсталяцію, налагодження програмного засобу тощо, ( $K_i = 1,10 \dots 1,12$ );

$k$  – кількість найменувань програмних засобів.

$$B_{\text{прг}} = 8410,00 \cdot 1 \cdot 1,01 = 8494,10 \text{ грн.}$$



Отримані результати зведемо до таблиці:

Таблиця 4.10 – Витрати на придбання програмних засобів по кожному виду

Найменування програмного засобу	Кількість, шт	Ціна за одиницю, грн	Вартість, грн
Програмний продукт Visual Studio 2017	1	8410,00	8494,10
Програмне забезпечення MATLAB 7.0	1	7770,00	7847,70
Середовище програмування мови C++	1	4320,00	4363,20
Всього			20705,00

#### 4.3.7 Амортизація обладнання, програмних засобів та приміщень

В спрощеному вигляді амортизаційні відрахування по кожному виду обладнання, приміщень та програмному забезпеченню тощо, розраховуємо з використанням прямолінійного методу амортизації за формулою:

$$A_{\text{обл}} = \frac{Ц_б}{T_в} * \frac{t_{\text{вик}}}{12} \quad (4.13)$$

де  $Ц_б$  – балансова вартість обладнання, програмних засобів, приміщень тощо, які використовувались для проведення досліджень, грн;

$t_{\text{вик}}$  – термін використання обладнання, програмних засобів, приміщень під час досліджень, місяців;

$T_в$  – строк корисного використання обладнання, програмних засобів, приміщень тощо, років.

$$A_{\text{обл}} = (27799,00 \cdot 2) / (2 \cdot 12) = 2316,58 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.11 – Амортизаційні відрахування по кожному виду обладнання

Найменування обладнання	Балансова вартість, грн	Строк корисного використання, років	Термін використання обладнання, місяців	Амортизаційні відрахування, грн
Персональний комп'ютер розробника програмного забезпечення	27799,00	2	2	2316,58
Робоче місце інженера-програміста	7250,00	3	2	402,78
Пристрої передачі даних	6510,00	5	2	217,00
Офісна оргтехніка	7960,00	5	2	265,33
Приміщення лабораторії розробки	405000,00	25	2	2700,00
Електронно-обчислювальний комплекс	48500,00	2	2	4041,67
Всього				9943,36

Витрати на силову електроенергію ( $B_e$ ) розраховуємо за формулою:

$$B_e = \sum_{i=1}^n \frac{W_{yi} * t_i * C_e * K_{vni}}{\eta_i} \quad (4.14)$$

де  $W_{yi}$  – встановлена потужність обладнання на визначеному етапі розробки, кВт;

$t_i$  – тривалість роботи обладнання на етапі дослідження, год;

$C_e$  – вартість 1 кВт-години електроенергії, грн; (вартість електроенергії визначається за даними енергопостачальної компанії), прийmemo  $C_e = 7,50$  грн;

$K_{vni}$  – коефіцієнт, що враховує використання потужності,  $K_{vni} < 1$ ;

$\eta_i$  – коефіцієнт корисної дії обладнання,  $\eta_i < 1$ .

$$B_e = 0,24 \cdot 200,0 \cdot 7,50 \cdot 0,95 / 0,97 = 360,00 \text{ грн.}$$

Проведені розрахунки зведемо до таблиці.

Таблиця 4.12 – Витрати на електроенергію

Найменування обладнання	Встановлена потужність, кВт	Тривалість роботи, год	Сума, грн
Персональний комп'ютер розробника програмного забезпечення	0,24	200,0	360,00
Робоче місце інженера-програміста	0,10	200,0	150,00
Пристрої передачі даних	0,03	200,0	45,00
Офісна оргтехніка	0,65	1,5	7,07
Електронно-обчислювальний комплекс	0,40	200,0	600,00
Осцилограф цифровий ОСЦ-2000	0,20	35,0	52,50
Аналізатор спектру	0,12	35,0	31,50
Всього			1246,07

#### 4.3.8 Витрати на роботи, які виконують сторонні підприємства, установи і організації

Витрати за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації» розраховуємо як 30...45% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{cn} = (Z_o + Z_p) * \frac{H_{cn}}{100\%} \quad (4.15)$$

де  $H_{cn}$  – норма нарахування за статтею «Витрати на роботи, які виконують сторонні підприємства, установи і організації», прийmemo  $H_{cn} = 30\%$ .

$$B_{cn} = (45768,10 + 3078,53) \cdot 30 / 100\% = 14653,99 \text{ грн.}$$

#### 4.3.9 Інші витрати

До статті «Інші витрати» належать витрати, які не знайшли відображення у зазначених статтях витрат і можуть бути віднесені безпосередньо на собівартість досліджень за прямими ознаками.

Витрати за статтею «Інші витрати» розраховуємо як 50...100% від суми основної заробітної плати дослідників та робітників за формулою:

$$I_{\text{в}} = (Z_{\text{o}} + Z_{\text{р}}) * \frac{H_{\text{ів}}}{100\%} \quad (4.16)$$

де  $H_{\text{ів}}$  – норма нарахування за статтею «Інші витрати», прийmemo  $H_{\text{ів}} = 55\%$ .

$$I_{\text{в}} = (45768,10 + 3078,53) \cdot 55 / 100\% = 26865,64 \text{ грн.}$$

#### 4.3.10 Накладні (загальновиробничі) витрати

Стаття "Накладні (загальновиробничі) витрати" включає витрати на управління організацією, винаходи та раціоналізацію, навчання кадрів, найм робочої сили, банківські послуги, освоєння виробництва продукції, науково-технічну інформацію, рекламу і інше.

Витрати за статтею «Накладні (загальновиробничі) витрати» розраховуємо як 100...150% від суми основної заробітної плати дослідників та робітників за формулою:

$$B_{\text{нзв}} = (Z_{\text{o}} + Z_{\text{р}}) * \frac{H_{\text{нзв}}}{100\%} \quad (4.17)$$

де  $H_{\text{нзв}}$  – норма нарахування за статтею «Накладні (загальновиробничі) витрати», прийmemo  $H_{\text{нзв}} = 105\%$ .

$$B_{\text{нзв}} = (45768,10 + 3078,53) \cdot 105 / 100\% = 51288,96 \text{ грн.}$$

Витрати на проведення науково-дослідної роботи на тему «Автоматизована система ідентифікації особи за її голосом» розраховуємо як суму всіх попередніх статей витрат за формулою:

$$V_{\text{заг}} = Z_o + Z_p + Z_{\text{дод}} + Z_n + M + K_v + V_{\text{спец}} + \\ + V_{\text{прг}} + A_{\text{обл}} + V_e + V_{\text{сп}} + I_e + V_{\text{нзв}} \quad (4.18)$$

$$V_{\text{заг}} = 45768,10 + 3078,53 + 5373,13 + 11928,35 + 4577,32 + 13414,82 + 22876,50 + 20705,00 + \\ 9943,36 + 1246,07 + 9769,33 + 14653,99 + 26865,64 + 51288,96 = 241489,08 \text{ грн.}$$

Загальні витрати  $ZV$  на завершення науково-дослідної (науково-технічної) роботи та оформлення її результатів розраховується за формулою:

$$ZV = \frac{V_{\text{заг}}}{\eta} \quad (4.19)$$

де  $\eta$  – коефіцієнт, який характеризує етап (стадію) виконання науково-дослідної роботи, прийmemo  $\eta = 0,9$ .

$$ZV = 241489,08 / 0,9 = 268321,20 \text{ грн.}$$

#### 4.4 Розрахунок економічної ефективності науково-технічної розробки при її можливій комерціалізації потенційним інвестором

В ринкових умовах узагальнюючим позитивним ефектом для потенційного інвестора від можливого впровадження науково-технічної розробки є збільшення чистого прибутку. Дослідження "Автоматизована система ідентифікації особи за її голосом" передбачає комерціалізацію протягом 4-х років на ринку.

В цьому випадку майбутній економічний ефект буде формуватися на основі таких даних:

$\Delta N$  – збільшення кількості споживачів пристрою, у періоди часу, що аналізуються, від покращення його певних характеристик;

Показник	1-й рік	2-й рік	3-й рік	4-й рік
Збільшення кількості споживачів, осіб	1000	1500	1500	950

$N$  – кількість споживачів які використовували аналогічний пристрій у році до впровадження результатів нової науково-технічної розробки, прийmemo 15500 осіб;

$C_0$  – вартість пристрою у році до впровадження результатів розробки, прийmemo 3800,00 грн;

$\pm \Delta C_0$  – зміна вартості пристрою від впровадження результатів науково-технічної розробки, прийmemo 281,50 грн.

Можливе збільшення чистого прибутку у потенційного інвестора  $\Delta \Pi_i$  для кожного із 4-х років, протягом яких очікується отримання позитивних результатів від можливого впровадження та комерціалізації науково-технічної розробки, розраховуємо за формулою [52].

$$\Delta \Pi_i = (\pm \Delta C_0 * N * C_0 * \Delta N)_i * \lambda * \rho * (1 - \frac{\vartheta}{100}) \quad (4.20)$$

де  $\lambda$  – коефіцієнт, який враховує сплату потенційним інвестором податку на додану вартість. У 2023 році ставка податку на додану вартість складає 20%, а коефіцієнт  $\lambda=0,8333$ ;

$\rho$  – коефіцієнт, який враховує рентабельність інноваційного продукту).  
Прийmemo  $\rho=42\%$ ;

$\vartheta$  – ставка податку на прибуток, який має сплачувати потенційний інвестор, у 2023 році  $\vartheta=18\%$ ;

Збільшення чистого прибутку 1-го року:

$$\Delta \Pi_1 = (281,50 \cdot 15500,00 + 4081,50 \cdot 1000) \cdot 0,83 \cdot 0,42 \cdot (1 - 0,18/100\%) = 2413948,68 \text{ грн.}$$

Збільшення чистого прибутку 2-го року:

$$\Delta\Pi_2 = (281,50 \cdot 15500,00 + 4081,50 \cdot 2500) \cdot 0,83 \cdot 0,42 \cdot (1 - 0,18/100\%) = 4164006,08 \text{ грн.}$$

Збільшення чистого прибутку 3-го року:

$$\Delta\Pi_3 = (281,50 \cdot 15500,00 + 4081,50 \cdot 4000) \cdot 0,83 \cdot 0,42 \cdot (1 - 0,18/100\%) = 5914063,49 \text{ грн.}$$

Збільшення чистого прибутку 4-го року:

$$\Delta\Pi_4 = (281,50 \cdot 15500,00 + 4081,50 \cdot 4950) \cdot 0,83 \cdot 0,42 \cdot (1 - 0,18/100\%) = 7022433,18 \text{ грн.}$$

Приведена вартість збільшення всіх чистих прибутків  $ПП$ , що їх може отримати потенційний інвестор від можливого впровадження та комерціалізації науково-технічної розробки:

$$ПП = \sum_{i=1}^T \frac{\Delta\Pi_i}{(1+\tau)^t} \quad (4.21)$$

де  $\Delta\Pi_i$  – збільшення чистого прибутку у кожному з років, протягом яких виявляються результати впровадження науково-технічної розробки, грн;

$T$  – період часу, протягом якого очікується отримання позитивних результатів від впровадження та комерціалізації науково-технічної розробки, роки;

$\tau$  – ставка дисконтування, за яку можна взяти щорічний прогнозований рівень інфляції в країні,  $\tau=0,15$ ;

$t$  – період часу (в роках) від моменту початку впровадження науково-технічної розробки до моменту отримання потенційним інвестором додаткових чистих прибутків у цьому році.

$$\begin{aligned} ПП &= 2413948,68/(1+0,15)^1 + 4164006,08/(1+0,15)^2 + 5914063,49/(1+0,15)^3 + \\ &+ 7022433,18/(1+0,15)^4 = 2099085,81 + 3148586,83 + 3888592,74 + 4015098,96 = 1315 \\ &1364,35 \text{ грн.} \end{aligned}$$

Величина початкових інвестицій  $PV$ , які потенційний інвестор має вкласти для впровадження і комерціалізації науково-технічної розробки:

$$PV = k_{\text{інв}} \quad (4.22)$$

де  $k_{\text{інв}}$  – коефіцієнт, що враховує витрати інвестора на впровадження науково-технічної розробки та її комерціалізацію, приймаємо  $k_{\text{інв}} = 2$ ;

$ZB$  – загальні витрати на проведення науково-технічної розробки та оформлення її результатів, приймаємо 268321,20 грн.

$$PV = k_{\text{інв}} = 2 \cdot 268321,20 = 536642,41 \text{ грн.}$$

Абсолютний економічний ефект  $E_{\text{абс}}$  для потенційного інвестора від можливого впровадження та комерціалізації науково-технічної розробки становитиме:

$$E_{\text{абс}} = \text{ПП} - PV \quad (4.23)$$

де ПП – приведена вартість зростання всіх чистих прибутків від можливого впровадження та комерціалізації науково-технічної розробки, 13151364,35 грн;

$PV$  – теперішня вартість початкових інвестицій, 536642,41 грн.

$$E_{\text{абс}} = \text{ПП} - PV = 13151364,35 - 536642,41 = 12614721,94 \text{ грн.}$$

Внутрішня економічна дохідність інвестицій  $E_{\text{в}}$ , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$E_{\text{в}} = \sqrt[T_{\text{ж}}]{1 + \frac{E_{\text{абс}}}{PV}} - 1 \quad (4.24)$$

де  $E_{\text{абс}}$  – абсолютний економічний ефект вкладених інвестицій, 12614721,94 грн;

$PV$  – теперішня вартість початкових інвестицій, 536642,41 грн;

$T_{\text{ж}}$  – життєвий цикл науково-технічної розробки, тобто час від початку її розробки до закінчення отримання позитивних результатів від її впровадження, 4 роки.



$$E_B = \sqrt[3]{1 + \frac{E_{abc}}{PV}} - 1 = (1 + 12614721,94/536642,41)^{1/4} = 1,22.$$

Мінімальна внутрішня економічна дохідність вкладених інвестицій  $\tau_{min}$

$$\tau_{min} = d + f \quad (4.25)$$

де  $d$  – середньозважена ставка за депозитними операціями в комерційних банках; в 2023 році в Україні  $d=0,1$ ;

$f$  – показник, що характеризує ризикованість вкладення інвестицій, прийmemo 0,25.

$\tau_{min} = 0,1 + 0,25 = 0,35 < 1,22$  свідчить про те, що внутрішня економічна дохідність інвестицій  $E_B$ , які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки вища мінімальної внутрішньої дохідності. Тобто інвестувати в науково-дослідну роботу за темою «Автоматизована система ідентифікації особи за її голосом» доцільно.

Період окупності інвестицій  $T_{ок}$  які можуть бути вкладені потенційним інвестором у впровадження та комерціалізацію науково-технічної розробки:

$$T_{ок} = \frac{1}{E_B} \quad (4.26)$$

де  $E_B$  – внутрішня економічна дохідність вкладених інвестицій.

$$T_{ок} = 1 / 1,22 = 0,82 \text{ р.}$$

$T_{ок} < 3$ -х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

#### 4.5 Висновки до розділу

Згідно проведених досліджень рівень комерційного потенціалу розробки за темою «Автоматизована система ідентифікації особи за її голосом» становить 33,7 бала, що, свідчить про комерційну важливість проведення даних досліджень (рівень комерційного потенціалу розробки вище середнього).

При оцінюванні за технічними параметрами, згідно узагальненого коефіцієнту якості розробки, науково-технічна розробка переважає існуючі аналоги приблизно в 1,19 рази.

Також термін окупності становить 0,82 р., що менше 3-х років, що свідчить про комерційну привабливість науково-технічної розробки і може спонукати потенційного інвестора профінансувати впровадження даної розробки та виведення її на ринок.

Отже можна зробити висновок про доцільність проведення науково-дослідної роботи за темою «Автоматизована система ідентифікації особи за її голосом».

## ВИСНОВКИ

В процесі виконання магістерської кваліфікаційної роботи отримані такі результати.

У Вступі до роботи показано актуальність розв'язання проблеми розробки автоматизованої системи ідентифікації особи за її голосом, продиктовано широкою затребуваністю систем доступу до сучасних інформаційних ресурсів.

В Розділі 1 проведено огляд існуючих програмних продуктів для ідентифікації і верифікації мовця за його голосом, проаналізовано їх переваги і недоліки. Проведений аналіз показав переваги застосування нейромережевого підходу для побудови системи ідентифікації особи на основі нейромережевого класифікатора.

В Розділі 2 розглянуто математичні основи методів обробки сигналу мовлення, які застосовуються для виділення встановлених інформативних ознак індивідуальних для мовця. Проведений експериментальний аналіз впливу типу ознак дозволяє вибрати в якості них мел-спектральні коефіцієнти і частоту основного тону сигналу мовлення особи..

В Розділі 3 розроблено структуру програмного забезпечення автоматизованої системи ідентифікації особи за її голосом. Визначено, що вона складається з програм виділення спектральних характеристик сигналу мовця, програми виділення частоти основного тону, програми навчання нейромережевого класифікатора і розпізнавання голосу.

В Економічній частині магістерської роботи проведена оцінка рівня комерційного потенціалу виконаної розробки автоматизованої системи ідентифікації особи за її голосом, яка показала її інвестиційну привабливість.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. М.М. Биков, В.В. Ковтун, В.О. Гаврилюк. Основи інтелектуальних технологій. - Частина 1. Технології розпізнавання. Навчальний посібник для студентів спеціальності 174 – Автоматизація та робототехніка. – Вінниця: ВНТУ, 2023. – 218 с.
2. Тарасов Ю. Контрольно-пропускний режим для підприємств. Захист інформації // Конфідент, 2002. - № 1. - с. 55-61.
3. An Overview and Analysis of Voice Authentication Methods [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.semanticscholar.org/paper/An-Overview-and-Analysis-of-Voice-Authentication-Shoup-Talkar/572af444f0382b8e7e156ab36192da95a3b8dec>.
4. Campbell J. P., Speaker Recognition: A Tutorial / J. P. Campbell // Proceedings of the IEEE. 1997. V. 85, N 9. P. 1437-1462.
5. Doddington George. Speaker recognition: Identifying people by voice ТИЕР. – 1985. – № 11 (73), С. 129–146
6. Ли У.А. Методы автоматического распознавания речи: в 2 кн. : Перев. с англ. / У. А. Ли, Е. П. Нейбург, Т. Б. Мартин и др.. – М.: Мир. 1983. – Кн. 1. – 328 с.
7. У.А. Методы автоматического распознавания речи: в 2 кн. : Перев. с англ. / У. А. Ли, Е. П. Нейбург, Т. Б. Мартин и др.. – М.: Мир. 1983. – Кн. 2. – 328 с.
8. Pervushin Ye. A. Obzor osnovnykh metodov raspoznavaniya diktorov [Review of the main speaker recognition methods] / Ye. A. Pervushin // Matematicheskiye struktury i modelirovaniye.[Mathematical Structures and Modeling] –2011. – Вып. 24. – P. 41-54.
9. Андреев А. Siri, Alexa, Google Assistant: огляд віртуальних помічників. – [Електронний ресурс]. Режим доступу: [://apix-drive.com/ua/blog/reviews/siri-alexa-google-assistant](http://apix-drive.com/ua/blog/reviews/siri-alexa-google-assistant)

10. Іващенко, Г.С. Методи голосової ідентифікації / Г.С. Іващенко, Д.І. Коваль // Проблеми інформатизації. Том 3. – 2021. – с. 94.
11. Martin A., Przybocki M. The NIST 1999 Speaker Recognition Evaluation – An Overview // Digital Signal Processing. 2000. V. 10. – P. 643-701.
12. Bonastre J.-F., Fredouille C., et al. A tutorial on text-independent speaker verification // EURASIP J. Appl. Signal Processing, 2004. No. 4. P. 430–451.
13. T N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P.Ouellet. Front-End Factor Analysis For Speaker Verification // IEEE Transactions on Audio, Speech, and Language Processing, 13(4):788–798, May 2011.
14. S. Liu and K. C. Sim, "On combining DNN and GMM with unsupervised speaker adaptation for robust automatic speech recognition", 2014 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP). – 2014 – p. 195-199.
15. Сорокін В. Н. Верифікація диктора по спектрально-часових параметрах мовного сигналу / В. Н. Сорокін, А. І. Ципліхин // Інформаційні процеси. - 2010. –Т.10. - № 2. - С. 87–104.
16. Корнієнко, О. О. (2018). Метод відображення мовних сигналів у задачі розпізнавання мовця. Технічні науки та технології, 3(9). – 2017. – с. 129-137.
17. Биков М.М., Ковтун В.В. Дослідження ефективності ознак розпізнавання мовців при використанні згортальних нейромереж // Оптико-електронні інформаційно-енергетичні технології, Вінниця. – 2016. - №2(32). – с. 22-28.
18. Биков М.М., Ковтун В.В. Аналіз ефективності ідентифікації мовця за частотою основного тону // Вісник Хмельницького національного університету. – 2004. – №2. – Ч.1. – Т.2(60). – с. 20-23.

19. Gaussian Mixture Models: What are they & when to use? [Электронный ресурс]. – Режим доступа до ресурсу: <https://vitalflux.com/gaussian-mixture-models-what-are-they-when-to-use>.
20. NIST Speaker Recognition Evaluation [Электронный ресурс]. – Режим доступа до ресурсу: <https://www.nist.gov/itl/iad/mig/speaker-recognition>.
21. Sorokin V. N. Speaker verification using the spectral-temporal parameters of a speech / V. N. Sorokin, A. I. Tsyplikhin // *Informatsionnyye protsessy*. [ Informational processes] – 2010. –Т.10. – № 2. – P. 87–104
22. Petrovsky, A. Instantaneous harmonic analysis: techniques and applications to speech signal processing / A. Petrovsky, E. Azarov // *International Conference on Speech and Computers*. – 2014. – №8773. – p. 24–33.
23. Reynolds D. Experimental evaluation of features for robust speaker identification // *IEEE Trans. Speech Audio Processing*, 1994. Vol. 2. No. 4. P. 639–643.
24. Burget L., Matejka P., Glembek O., et al. Analysis of feature extraction and channel compensation in GMM speaker recognition system // *IEEE Trans. Audio Speech Language Processing*, 2007. Vol. 15. Iss. 7. P. 1979–1986.
25. Kenny P., Ouellet P., Dehak N., et al. A study of interspeaker variability in speaker verification // *IEEE Trans. Audio Speech Language Processing*, 2008. Vol. 16. Iss. 5. P. 980–988.
26. Vogt R., Sridharan S. Explicit modeling of session variability for speaker verification // *Computer Speech Language*, 2008. Vol. 22 (I). P. 17–38.
27. Reynolds D. A., Quatieri T. F., Dunn R. B. Speaker verification using adapted GMM // *Digital Signal Processing*, 2000. No. 10. P. 19–41.
28. Campbell W., Sturim D., Reynolds D. Support vector machines using GMM supervectors for speaker verification // *IEEE Signal Processing Lett.*, 2006. Vol. 13. No. 5. P. 308–311.
29. Vapnik V. *The nature of statistical learning theory*. — Springer, 1995.

30. Белозьорова Я.А. Ідентифікація диктора на основі кратномасштабного аналізу // Інженерія програмного забезпечення No1(29). – 2017. – с. 15-24.
31. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер. – М.: Радио и связь, 2001. 496 с.
32. Фант Г. Анализ и синтез речи / Г. Фант.–Новосибирск: Наука, 1970.–325 с.
33. Фланаган Дж. Анализ, синтез и восприятие речи / Дж. Фланаган. – М.: Связь, 1968. – 392 с.
34. Патент 55863А UA, МКІ G10L15 15/00 Пристрій для виділення ознак мовних сигналів. Биков М.М., Грищук Т.В., Ковтун В.В., №2002075729; заявл. 11.07.2002; опубл. 15.04.2003. – 4 с.
35. Скучик Е. Основы акустики. Т.1 / Е. Скучик. – М.: Изд-во “Мир”, 1976. - 520 с.
36. S. Paulose, D. Mathew and A. Thomas, "Performance evaluation of different modeling methods and classifiers with MFCC and IHC features for speaker recognition", Procedia computer science, vol. 115. – 2017.– p. 55-62.
37. J. Martinez, H. Perez, E. Escamilla and M. M. Suzuki, "Speaker recognition using mel frequency cepstral coefficients (MFCC) and vector quantization (VQ) techniques", 22nd International Conference on Electrical Communications and Computers (CONIELECOMP 2012). – 2012. – p. 248-251.
38. L. Muda, B. KM and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques", Journal of Computing, vol. 2, no. 3. – 2010. – p. 138-143.
39. Z. Wang, L. Chen, L. Wang and G. Diao, "Recognition of audio depression based on convolutional neural network and generative antagonism network model", IEEE Access, 2020.
40. C. Li, X. Ma, B. Jiang, X. Li, X. Zhang, X. Liu, et al., Deep speaker: an end-to-end neural speaker embedding system, vol. 650 – 2017.

41. Ganchev, T. Comparative evaluation of various MFCC implementations on the speaker verification task / T. Ganchev, N. Fakotakis, G. Kokkinakis // Proceedings of the SPECOM. – 2005. – p. 191–194.
42. Z. Saquib, N. Salam, R.P. Nair, N. Pandey and A. Joshi, "A survey on automatic speaker recognition systems" in Signal Processing and Multimedia, Berlin, Heidelberg:Springer. – 2010. – p. 134-145.
43. N. Singh, A. Agrawal and R.A. Khan, "Automatic speaker recognition: current approaches and progress in the last six decades", Global J Enterp Inf Syst., vol. 9, no. 3. – 2017. – p. 45-52.
44. A. Revathi, R. Ganapathy and Y. Venkataramani, "Text independent speaker recognition and speaker independent speech recognition using iterative clustering approach", International Journal of Computer science & Information Technology (IJCSIT), vol. 1.2. – 2009. – p. 30-42.
45. Y. Lukic, C. Vogt, O. Dürr and T. Stadelmann, "Speaker identification and clustering using convolutional neural networks", 2016 IEEE 26th international workshop on machine learning for signal processing (MLSP). – 2016. – p. 1-6.
46. Царьов Р.Ю., Лемеха Т.М. Біометричні технології: навч. посіб. [для вищих навчальних закладів]. Одеса: ОНАЗ ім. О.С. Попова, 2016. 7-27, 57-67 с.
47. Дедів, І.Ю. Обґрунтування методу голосової ідентифікації особи / І.Ю. Дедів, Л.Є. Дедів, С. Макар // Матеріали IV Міжнародної науково-технічної конференції «Теоретичні та прикладні аспекти радіотехніки, приладобудування і комп'ютерних технологій» – 2019. – с. 90-91.
48. Scheffer, N. Towards noise-robust speaker recognition using probabilistic linear discriminant analysis / N. Scheffer, M. Graciarena, L. Ferrer, L. Burget, Y. Lei // International Conference on Acoustics, Speech, and Signal Processing (ICASSP). – Kyoto, 2012. – p. 4253–4256.



49. Kinnunen T. Spectral Features for Automatic Text-independent Speaker Recognition / T. Kinnunen. – Helsinki: Gaudeamus, 2013.
50. Srinivasan. A Speaker Identification and Verification using Vector Quantization and Mel Frequency Cepstral Coefficients. Research Journal of Applied Sciences, Engineering and Technology 4(1). – 2012. – p. 33-40.
51. П. Бідюк, Сучасні методи біометричної ідентифікації / П. Бідюк, В. Бондарчук // Правове, нормативне та метрологічне забезпечення системи захисту інформації в Україні: науково-технічний збірник. – 2009. – №1 (18). – с. 137-146
52. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. – Вінниця : ВНТУ, 2021. – 42 с.
53. Кавецький В. В. Економічне обґрунтування інноваційних рішень: практикум / В. В. Кавецький, В. О. Козловський, І. В. Причепа – Вінниця : ВНТУ, 2016. – 113 с.

**ДОДАТКИ**

**ПРОТОКОЛ  
ПЕРЕВІРКИ КВАЛІФІКАЦІЙНОЇ РОБОТИ  
НА НАЯВНІСТЬ ТЕКСТОВИХ ЗАПОЗИЧЕНЬ**

Назва роботи: «Автоматизована система ідентифікації особи за її голосом»

Тип роботи: Магістерська кваліфікаційна робота  
(БДР, МКР)

Підрозділ КСУ, ФІТА  
(кафедра, факультет)

**Показники звіту подібності Unicheck**

Оригінальність 92,8% Схожість 7,2%

Аналіз звіту подібності (відмітити потрібне)

У Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.

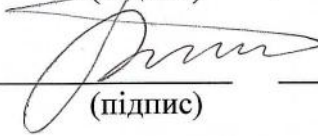
Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і відсутності самостійності її автора. Роботу направити на розгляд експертної комісії кафедри.

Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Особа, відповідальна за перевірку  Володимир ДУБОВОЙ  
(підпис) (прізвище, ініціали)

Ознайомлені з повним звітом подібності, який був згенерований системою Unicheck щодо роботи.

Автор роботи  Дмитро РИБАК  
(підпис) (прізвище, ініціали)

Керівник роботи  Микола БИКОВ  
(підпис) (прізвище, ініціали)

Додаток Б

ВНТУ

ЗАТВЕРДЖУЮ



Т.в.о. зав. кафедри КСУ

д.т.н., доцент, Марія ЮХИМЧУК

---

«6» шостня 2023 р.

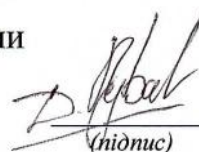
## ТЕХНІЧНЕ ЗАВДАННЯ

на виконання магістерської кваліфікаційної роботи

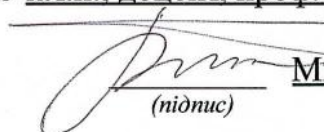
“Автоматизована система ідентифікації особи за її голосом”

08-33.МКР.012.00.000 ТЗ

Студент групи

  
(підпис)2 АКІТ-22мДмитро РИБАК

Ім'я ПРІЗВИЩЕ

Керівник к.т.н., доцент, проф. кафедри КСУ  
(підпис)Микола БИКОВ

Ім'я ПРІЗВИЩЕ

Вінниця 2023

1. Назва та галузь застосування

1.1. Назва – Автоматизована система ідентифікації особи за її голосом.

1.2. Галузь застосування – Автоматика та управління

2. Підстава для проведення розробки:

Тема магістерської кваліфікаційної роботи затверджена наказом по ВНТУ №247 від “18” вересня 2023 року

3. Мета та призначення розробки.

Метою магістерської кваліфікаційної роботи є підвищення ефективності функціонування автоматизованих систем управління за рахунок автоматизації процесу ідентифікації персоналу за голосом.

4. Джерела розробки.

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

1. An Overview and Analysis of Voice Authentication Methods [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.semanticscholar.org/paper/An-Overview-and-Analysis-of-Voice-Authentication-Shoup-Talkar/572af444f0382b8e7e156ab36192da95a3b8dec4>

2. Биков М.М., Ковтун В.В. Дослідження ефективності ознак розпізнавання мовців при використанні згортальних нейромереж // Оптико-електронні інформаційно-енергетичні технології, Вінниця. – 2016. - №2(32). – 22-28 с.

3. Биков М.М., Ковтун В.В. Аналіз ефективності ідентифікації мовця за частотою основного тону // Вісник Хмельницького національного університету. – 2004. – №2. – Ч.1. – Т.2(60). – С. 20-23.

5. Вимоги до розробки.

5.1. Перелік головних функцій:

- введення звукового файлу (описів образів);
- виділення ознак індивідуальності мовця;
- обчислення схожості і ідентифікація мовця.

5.2. Основні технічні вимоги до розробки.

5.2.1. Вимоги до програмної платформи:

- WINDOWS XP\7;
- MATLAB 7.0;
- C++;
- Visual Studio 2017.

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;
- розмір капіталовкладень;
- економічний ефект;
- інші економічні переваги у порівнянні з аналогами.

## 7. Стадії та етапи розробки.

### 7.1 Пояснювальна записка:

1	Варіантний аналіз об'єкта автоматизації	12.09.2023 р.
2	Постановка задачі і розробка технічного завдання	22.09. 2023 р.
3	Математичне обґрунтування вибраного методу ознакового опису голосу мовця	28.09. 2023 р.
4	Розробка алгоритмічного та програмного забезпечення системи	3.10. 2023 р.
5	Машинний експеримент	12.11. 2023 р.
6	Оформлення пояснювальної записки, графічного матеріалу і презентації	20.11. 2023 р.
7	Апробація результатів дослідження	05.11.2023 р.
8	Публікації	12.11.2023 р.
9	Графічні матеріали:	
	Плакати демонстраційні	11.11.2023 р.
	Схеми алгоритмів і програм	12.11.2023 р.
	Схеми системи	13.11.2023 р.
	UML діаграми	14.11.2023 р.
	Результати машинного експерименту	17.11. 2023 р.
10	Захист МКР	20.12.2023 р.

## 8. Порядок контролю і приймання.

- 8.1. Хід виконання проекту контролюється керівником роботи, консультантами з економічної частини. Рубіжний контроль провести до «11» грудня 2023 р.
- 8.2. Атестація роботи здійснюється на попередньому захисті. Попередній захист магістерської кваліфікаційної роботи провести до «11» грудня 2023 р.
- 8.3. Підсумкове рішення щодо оцінки якості виконання магістерської кваліфікаційної роботи приймається на засіданні ДЕК. Захист роботи провести до «20» грудня 2023 р.

## Додаток В – Лістинги програм

## Лістинг програми фільтрації

```

[b,a]=ellip(4,0.1,40,[pr1 pr2]*2/z); % обчислення параметрів фільтра
[H,w]=freqz(b,a,512);           % pr1, pr2 - границі полоси

%plot(w*z/(2*pi),abs(H)); % вивід полоси пропускання на екран
%axis([0 5000 0 1]);

ff=filter(b,a,y);           % отримання відфільтрованого сигналу
fff=abs(ff);                % енергія сигналу

iii=-255;                   % підготовка до циклу
i=0;

while iii < end_cycl        % отримання огинаючої сигналу фільтрів
    iii=iii+256;            % інтервал перерахунку 256 келейок = 20 мс
    iii_e=iii+255;
    i=i+1;
    N=i;                    % кількість крапок вихідного вектора
    Ampl=0;
    for ii=iii:iii_e
        Ampl=Ampl+fff(ii);
    end
    f(nf,i)=Ampl;          % значення енергії в крапках
end

```

## Нейронна мережа для ідентифікації диктора

```

da1=dlmread('da1.dat');
da2=dlmread('da2.dat');
da3=dlmread('da3.dat');
da4=dlmread('da4.dat');
da5=dlmread('da5.dat');
da6=dlmread('da6.dat');
for i=1:72
    ina(i,1)=da1(i);
    ina(i,2)=da2(i);
    ina(i,3)=da3(i);
    ina(i,4)=da4(i);
    ina(i,5)=da5(i);
    ina(i,6)=da6(i);
end;
ina

```

```
do1=dlmread('do1.dat');
do2=dlmread('do2.dat');
do3=dlmread('do3.dat');
do4=dlmread('do4.dat');
do5=dlmread('do5.dat');
do6=dlmread('do6.dat');
for i=1:72
    ino(i,1)=do1(i);
    ino(i,2)=do2(i);
    ino(i,3)=do3(i);
    ino(i,4)=do4(i);
    ino(i,5)=do5(i);
    ino(i,6)=do6(i);
end;
ino
```

```
du1=dlmread('du1.dat');
du2=dlmread('du2.dat');
du3=dlmread('du3.dat');
du4=dlmread('du4.dat');
du5=dlmread('du5.dat');
du6=dlmread('du6.dat');
for i=1:72
    inu(i,1)=du1(i);
    inu(i,2)=du2(i);
    inu(i,3)=du3(i);
    inu(i,4)=du4(i);
    inu(i,5)=du5(i);
    inu(i,6)=du6(i);
end;
inu
```

```
de1=dlmread('de1.dat');
de2=dlmread('de2.dat');
de3=dlmread('de3.dat');
de4=dlmread('de4.dat');
de5=dlmread('de5.dat');
de6=dlmread('de6.dat');
for i=1:72
    ine(i,1)=de1(i);
    ine(i,2)=de2(i);
    ine(i,3)=de3(i);
    ine(i,4)=de4(i);
    ine(i,5)=de5(i);
    ine(i,6)=de6(i);
end;
ine
```

```
di1=dlmread('di1.dat');
di2=dlmread('di2.dat');
di3=dlmread('di3.dat');
```



```

di4=dlmread('di4.dat');
di5=dlmread('di5.dat');
di6=dlmread('di6.dat');
for i=1:72
    ini(i,1)=di1(i);
    ini(i,2)=di2(i);
    ini(i,3)=di3(i);
    ini(i,4)=di4(i);
    ini(i,5)=di5(i);
    ini(i,6)=di6(i);
end;
ini

outdate=[1 0 0 0 0 0;
          0 1 0 0 0 0;
          0 0 1 0 0 0;
          0 0 0 1 0 0;
          0 0 0 0 1 0;
          0 0 0 0 0 1];

neta= newff(minmax(ina),[12 10 6],{ 'purelin' 'purelin'
'purelin'},'trainlm');
neto= newff(minmax(ino),[12 10 6],{ 'purelin' 'purelin'
'purelin'},'trainlm');
netu= newff(minmax(inu),[12 10 6],{ 'purelin' 'purelin'
'purelin'},'trainlm');
nete= newff(minmax(ine),[12 10 6],{ 'purelin' 'purelin'
'purelin'},'trainlm');
neti= newff(minmax(ini),[12 10 6],{ 'purelin' 'purelin'
'purelin'},'trainlm');

% матриці ваг міжнейронних зв'язків до навчання мережі
ab1do = neta.b{1}
aW2do= neta.LW{2,1}
ab2do = neta.b{2}
aW3do= neta.LW{3,2}
ab3do = neta.b{3}

ob1do = neto.b{1}
oW2do= neto.LW{2,1}
ob2do = neto.b{2}
oW3do= neto.LW{3,2}
ob3do = neto.b{3}

ub1do = netu.b{1}
uW2do= netu.LW{2,1}
ub2do = netu.b{2}
uW3do= netu.LW{3,2}
ub3do = netu.b{3}

eb1do = neti.b{1}
eW2do= neti.LW{2,1}
eb2do = neti.b{2}

```

```

eW3do= neti.LW{3,2}
eb3do = neti.b{3}

ib1do = neti.b{1}
iW2do= neti.LW{2,1}
ib2do = neti.b{2}
iW3do= neti.LW{3,2}
ib3do = neti.b{3}

% значення виходів нейронної мережі до навчання
aout_do= sim(neta,ina)
oout_do= sim(neto,ino)
uout_do= sim(netu,inu)
eout_do= sim(nete,ine)
iout_do= sim(neti,ini)

% встановлення параметрів процесу навчання
neta.trainParam.epochs = 100; % максимальна кількість епох
neta.trainParam.show = 10; % показувати результати навчання
через кожні 10 епох
neta.trainParam.goal = 1e-3; % похибка навчання

neto.trainParam.epochs = 100; % максимальна кількість епох
neto.trainParam.show = 10; % показувати результати навчання
через кожні 10 епох
neto.trainParam.goal = 1e-3; % похибка навчання

netu.trainParam.epochs = 100; % максимальна кількість епох
netu.trainParam.show = 10; % показувати результати навчання
через кожні 10 епох
netu.trainParam.goal = 1e-3; % похибка навчання

nete.trainParam.epochs = 100; % максимальна кількість епох
nete.trainParam.show = 10; % показувати результати навчання
через кожні 10 епох
nete.trainParam.goal = 1e-3; % похибка навчання

neti.trainParam.epochs = 100; % максимальна кількість епох
neti.trainParam.show = 10; % показувати результати навчання
через кожні 10 епох
neti.trainParam.goal = 1e-3; % похибка навчання

% Навчання нейронної мережі
tic

figure(1);
[neta,tr]=train(neta,ina,outdate);
aout_posle=sim(neta,ina);
figure(2);
[neto,tr]=train(neto,ino,outdate);
oout_posle=sim(neto,ino);
figure(3);
[netu,tr]=train(netu,inu,outdate);

```

```

uout_posle=sim(netu, inu);
figure(4);
[nete, tr]=train(nete, ine, outdate);
eout_posle=sim(nete, ine);
figure(5);
[neti, tr]=train(neti, ini, outdate);
iout_posle=sim(neti, ini);
toc
% матриці ваг міжнейронних зв'язків після навчання мережі
ab1posle = neta.b{1}
aW2posle= neta.LW{2,1}
ab2posle = neta.b{2}
aW3posle= neta.LW{3,2}
ab3posle = neta.b{3}

ob1posle = neto.b{1}
oW2posle= neto.LW{2,1}
ob2posle = neto.b{2}
oW3posle= neto.LW{3,2}
ob3posle = neto.b{3}

ub1posle = netu.b{1}
uW2posle= netu.LW{2,1}
ub2posle = netu.b{2}
uW3posle= netu.LW{3,2}
ub3posle = netu.b{3}

eb1posle = neti.b{1}
eW2posle= neti.LW{2,1}
eb2posle = neti.b{2}
eW3posle= neti.LW{3,2}
eb3posle = neti.b{3}

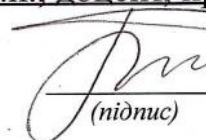
ib1posle = neti.b{1}
iW2posle= neti.LW{2,1}
ib2posle = neti.b{2}
iW3posle= neti.LW{3,2}
ib3posle = neti.b{3}

```

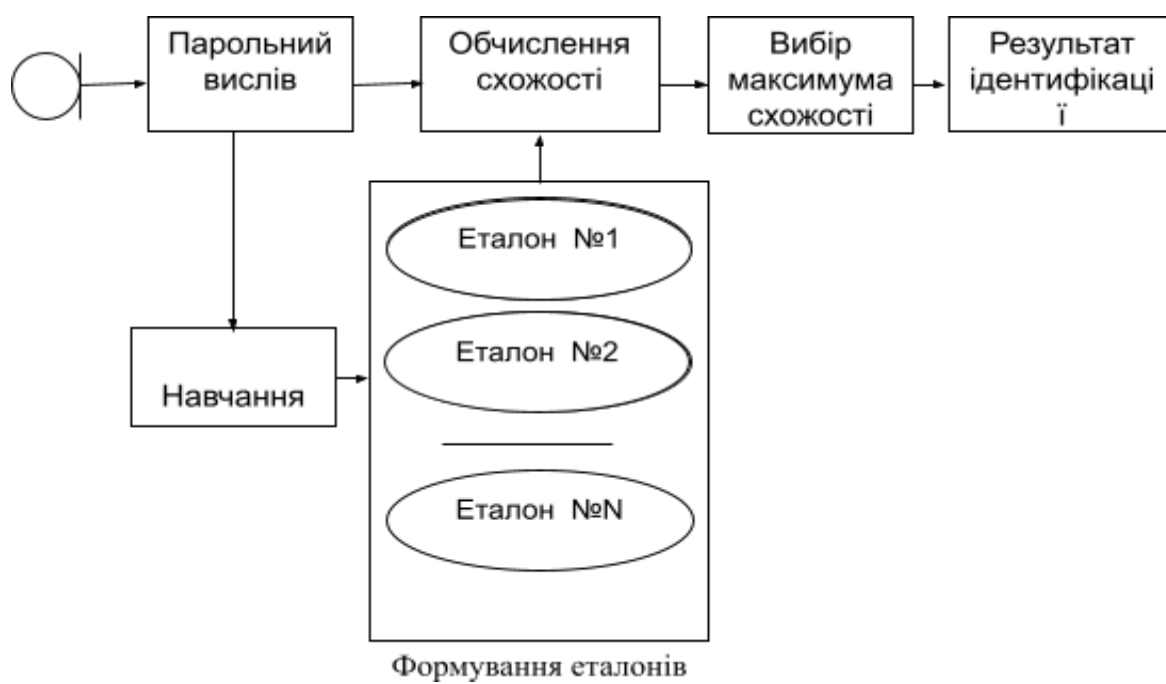
Додаток Г

**ІЛЮСТРАТИВНА ЧАСТИНА****АВТОМАТИЗОВАНА СИСТЕМА ІДЕНТИФІКАЦІЇ ОСОБИ ЗА ЇЇ  
ГОЛОСОМ**

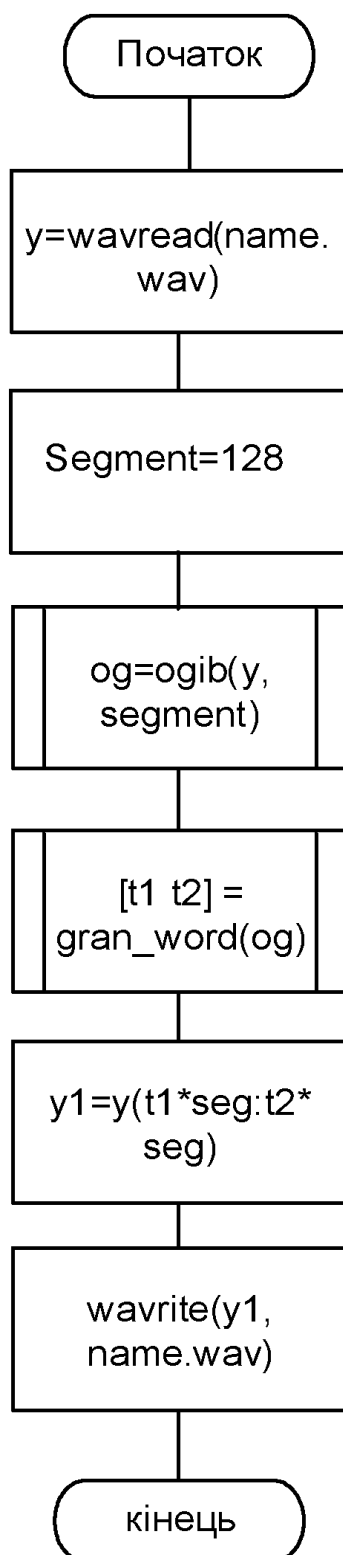
Студент групи

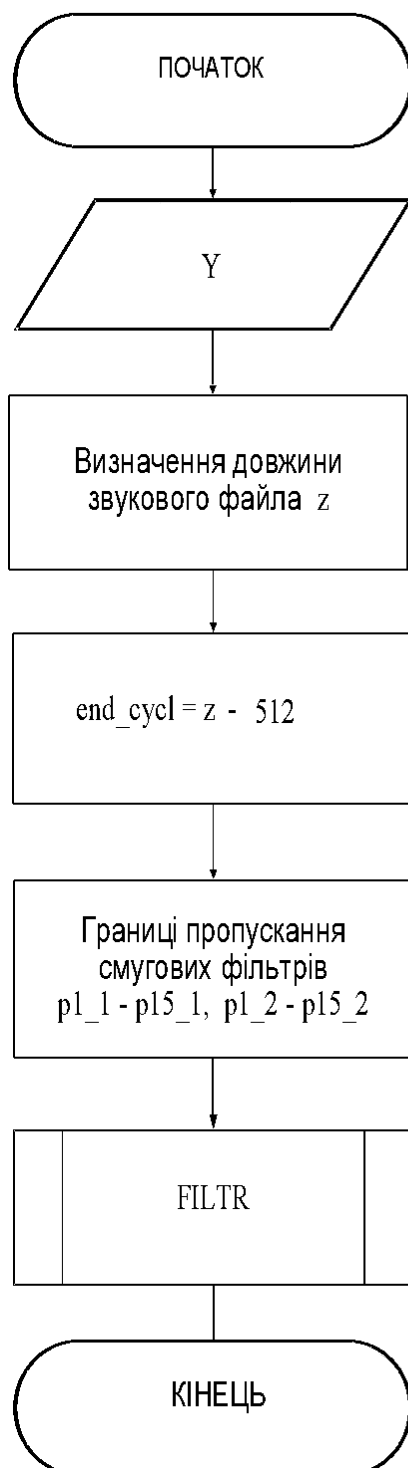
  
(підпис)2 АКІТ-22мДмитро РИБАК  
Ім'я ПРИЗВИЩЕКерівник к.т.н., доцент, проф. кафедри КСУ  
(підпис)Микола БИКОВ  
Ім'я ПРИЗВИЩЕ

## УЗАГАЛЬНЕНА СХЕМА СИСТЕМИ ІДЕНТИФІКАЦІЇ ОСОБИ ЗА ГОЛОСОМ

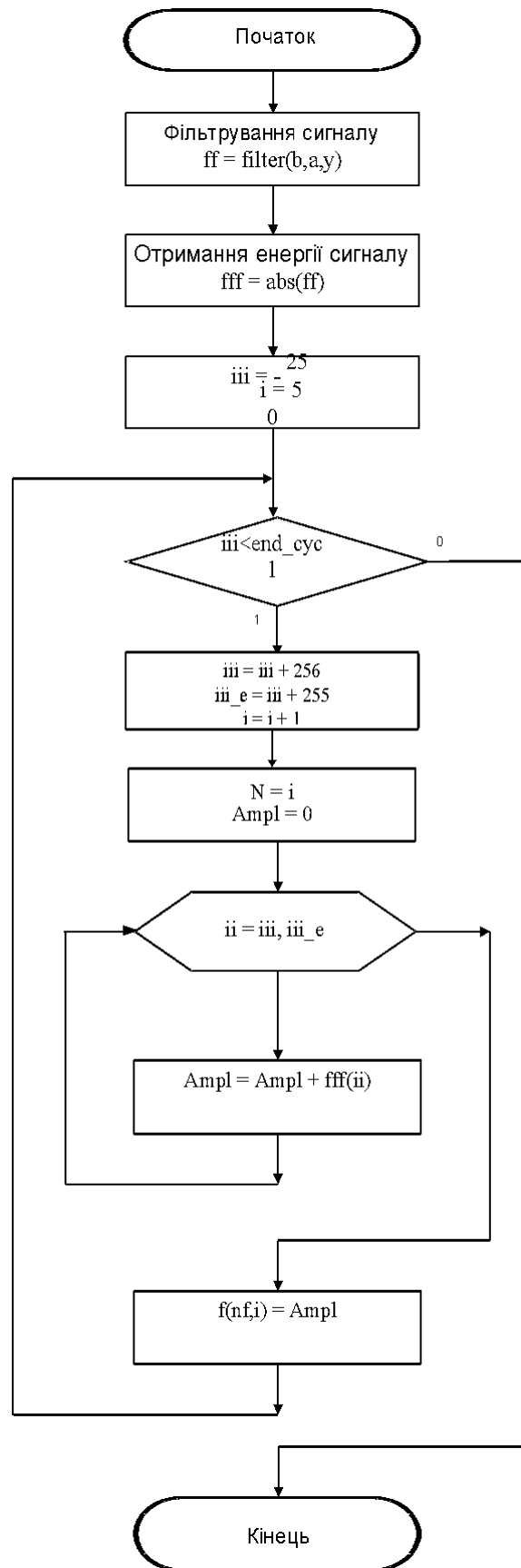


## СХЕМА АЛГОРИТМУ ВИЗНАЧЕННЯ ГРАНИЦЬ ПАРОЛЬНОГО СЛОВА



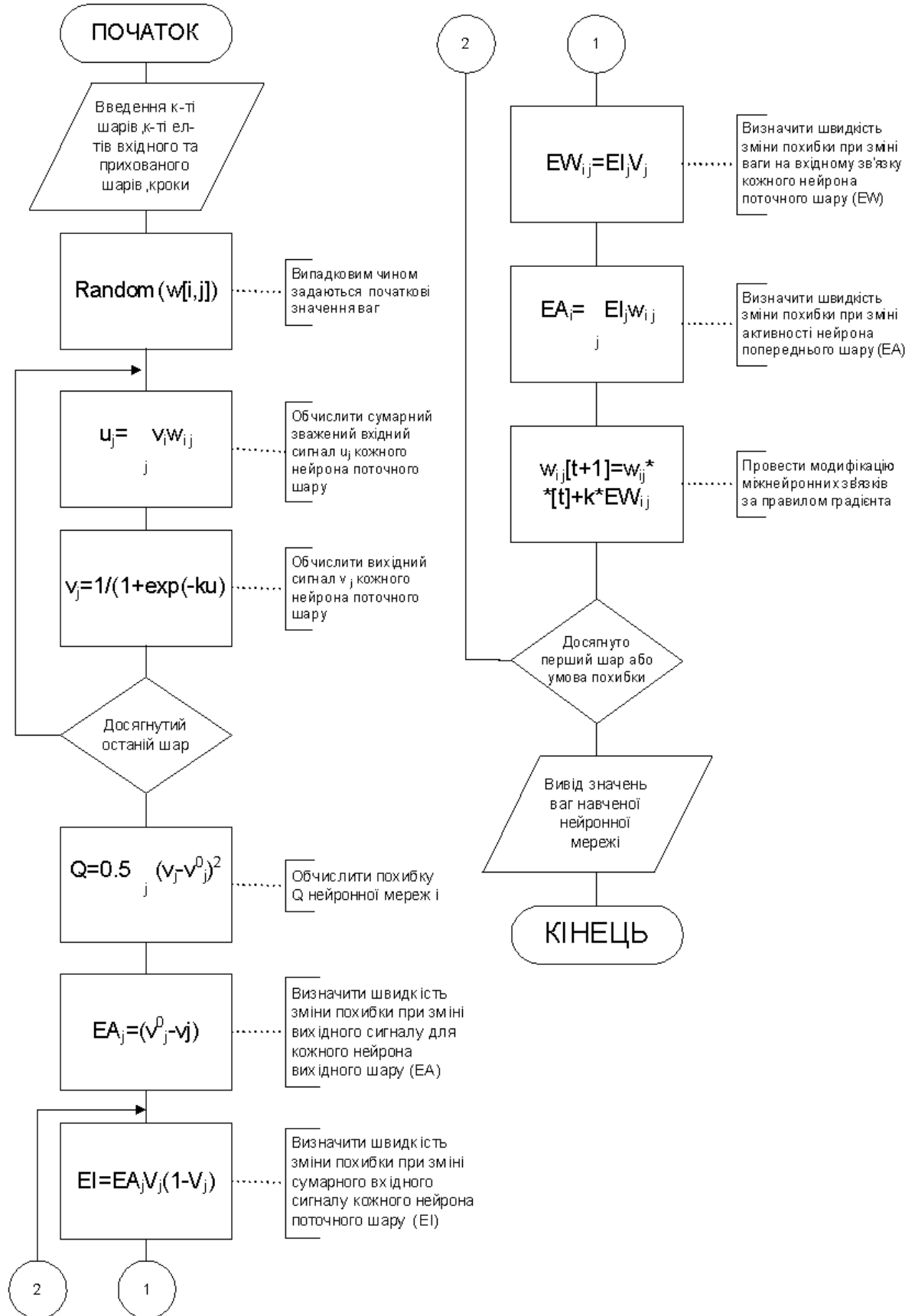
**СХЕМА ПРОГРАМИ ОБРОБКИ СИГНАЛУ МОВЦЯ**

## СХЕМА ПРОГРАМИ СПЕКТРАЛЬНОГО АНАЛІЗУ СИГНАЛУ МОВЦЯ

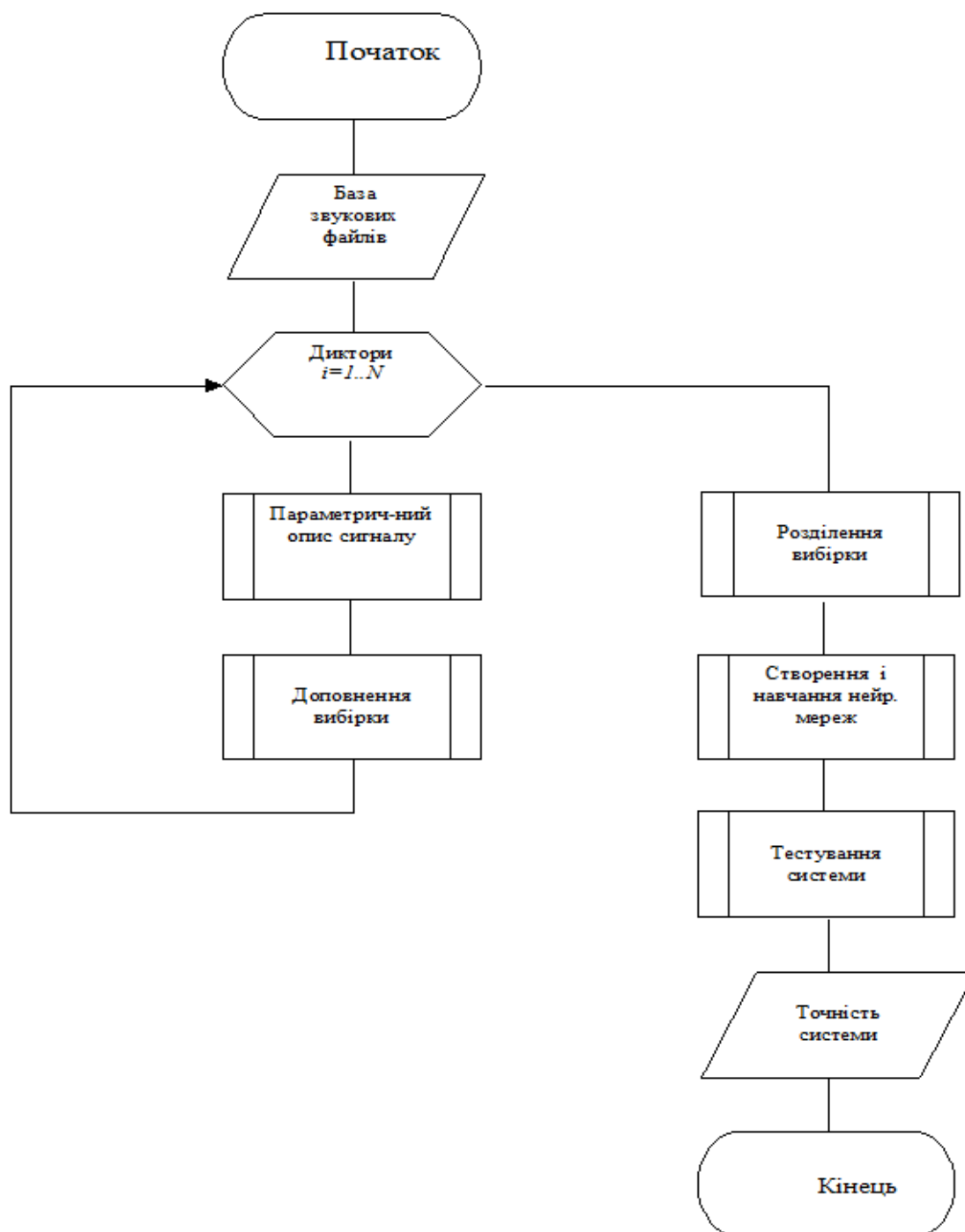




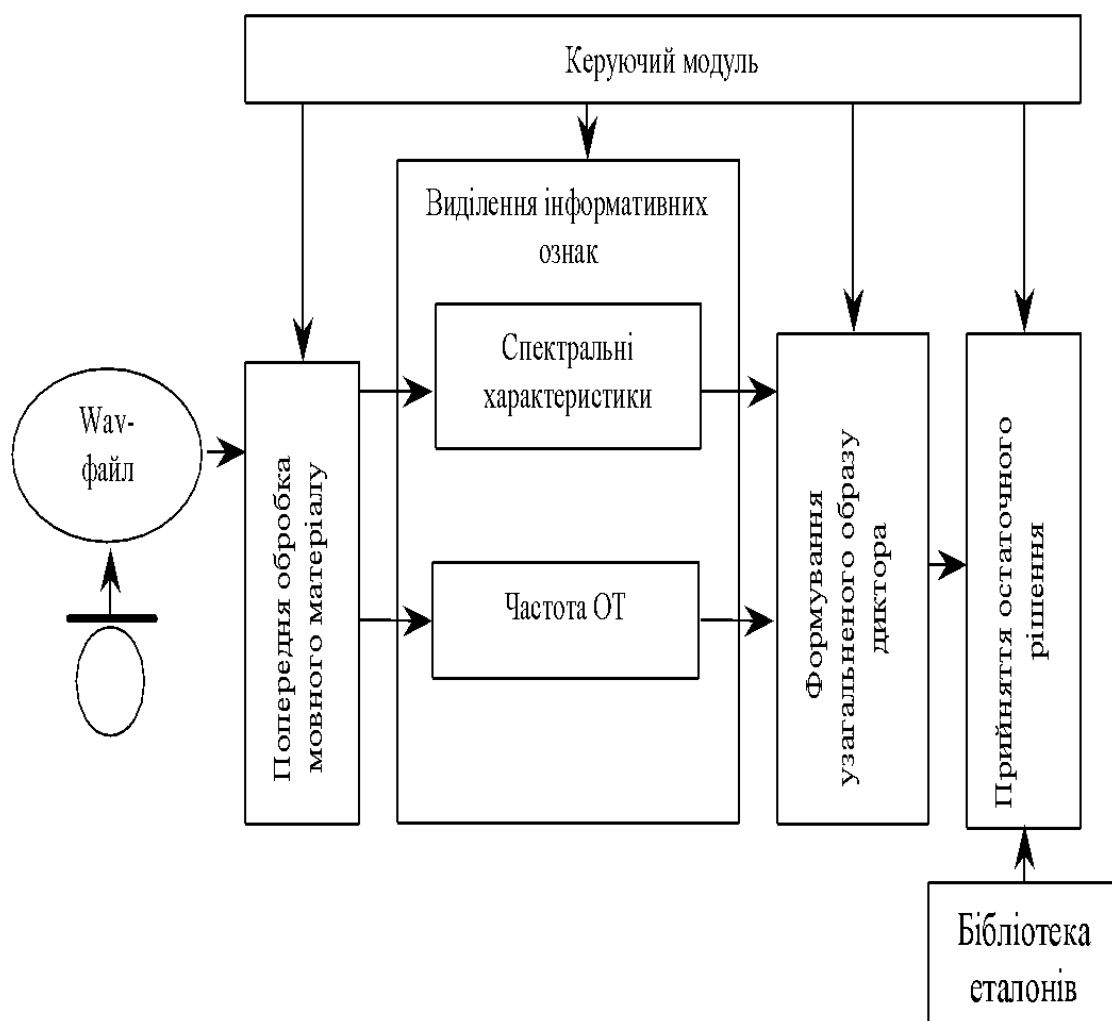
**СХЕМА АЛГОРИТМУ НАВЧАННЯ НЕЙРОМЕРЕЖЕВОГО КЛАСИФІКАТОРА**



## СХЕМА АЛГОРИТМУ РОЗПІЗНАВАННЯ НЕЙРОМЕРЕЖЕВИМ КЛАСИФІКАТОРОМ



## СТРУКТУРА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ СИСТЕМИ

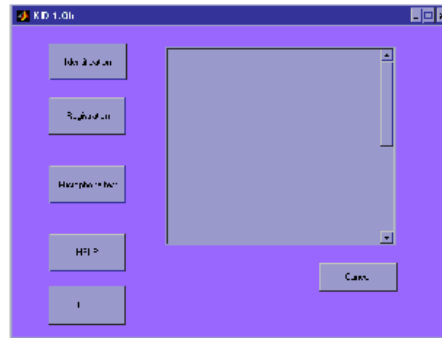


## ДІАЛОВОВІ ВІКНА ІНТЕРФЕЙСУ ПРОГРАМИ ІДЕНТИФІКАЦІЇ

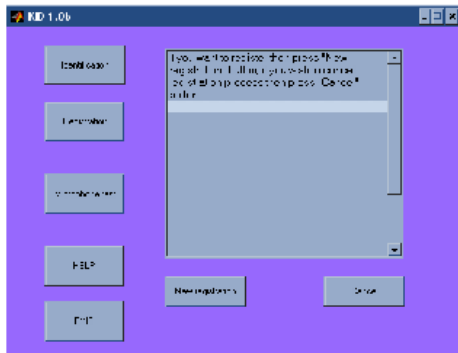
«Registration» - меню реєстрації нового користувача

«Identification» - процедура розпізнавання дисгорів

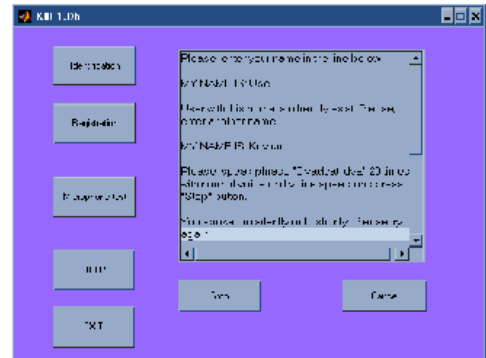
«Microphone test» - тестування мікрофону



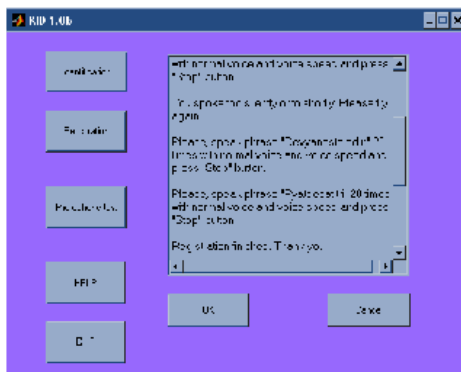
Головне вікно програми ідентифікації



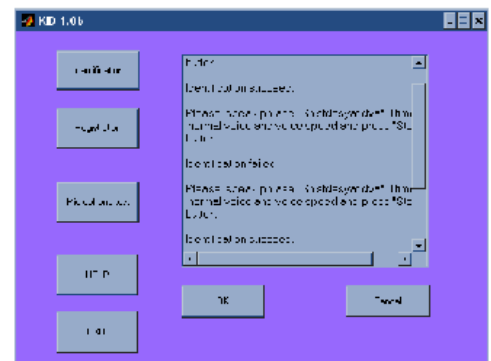
Початок реєстрації нового користувача програми ідентифікації



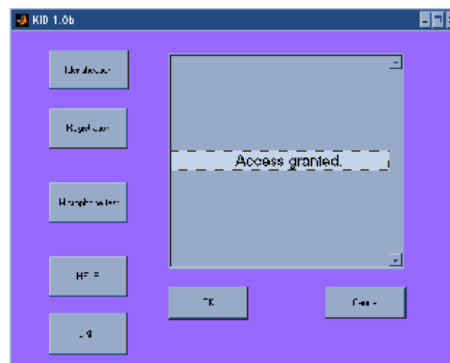
Процедура реєстрації нового користувача та формування ідентифікатора



Виконання процедури реєстрації нового користувача

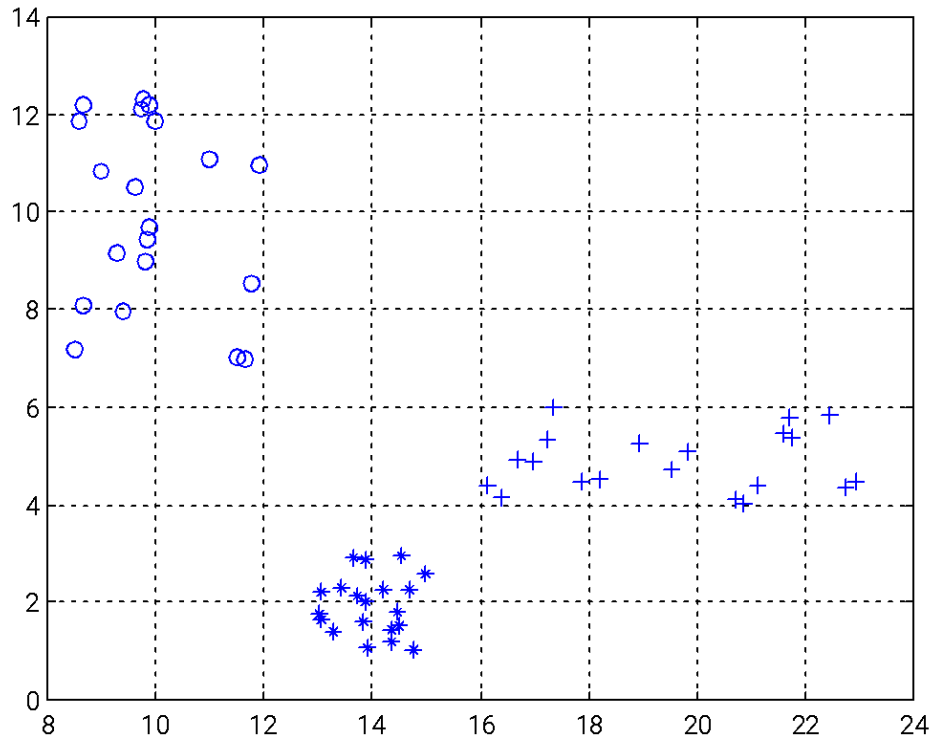


Процедура реєстрації нового користувача



Процедура реєстрації нового користувача

## Образи, що відповідають дикторам у 2-вимірному просторі ознак



**Результати експерименту з навчання та тестування  
нейромережі для ідентифікації особи за короткочасним  
спектром і частотою основного тону**

$N_2$	Використаний метод навчання мережі	$t$ , с	$T$ , %
10	Gradient descent backpropagation	124	82
	Gradient descent with momentum backpropagation	114	82
	Levenberg-Marquardt backpropagation	96	63
	RPROP backpropagation	103	83
20	Gradient descent backpropagation	142	84
	Gradient descent with momentum backpropagation	128	83
	Levenberg-Marquardt backpropagation	107	64
	RPROP backpropagation	126	86
30	Gradient descent backpropagation	149	84
	Gradient descent with momentum backpropagation	132	84
	Levenberg-Marquardt backpropagation	112	63
	RPROP backpropagation	126	86