Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)
факультет комп'ютерно-інформаційних систем і програмної інженерії
(назва факультету)
кафедри комп'ютерних систем та мереж
(повна назва кафедри)

ПОЯСНЮВАЛЬНА ЗАПИСКА

до дипломної роботи

магістра
(освітній ступінь)

на тему: Методи та засоби автоматизованого розпізнавання мови на базі мікроконтролерної системи

Виконав: студент (ка) 6 курсу, групи СІм-61
спеціальності
123 „Комп'ютерна інженерія"
(шифр і назва спеціальності)

(підпис) Чаплик В. А.
(прізвище та ініціали)

Керівник (підпис) Лусник К. С.
(прізвище та ініціали)

Нормоконтроль (підпис) Гисем Є. В
(прізвище та ініціали)

Рецензент (підпис) Липак Г. І
(прізвище та ініціали)

м. Тернопіль – 2019

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)

Факультет комп'ютерно-інформаційних систем і програмної інженерії
Кафедра комп'ютерних систем та мереж
Освітній ступінь магістр
Напрям підготовки
(шифр і назва)
Спеціальність 123 «Комп'ютерна інженерія»
(шифр і назва)

ЗАТВЕРДЖУЮ
Завідувач кафедри КС
Бухтяк Р.М.
«30» 09 2019 р.

З А В Д А Н Н Я
НА ДИПЛОМНУ РОБОТУ СТУДЕНТУ

Часник Вікторія Анатоліївна
(прізвище, ім'я, по батькові)

1. Тема роботи Методи та засоби автоматизованого розпізнавання мови на базі мікропроцесорної системи

Керівник роботи Лудик Ігор Степанович, доктор філософії
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом по університету від «27» вересня 2019 року № 4/7-854

2. Термін подання студентом роботи 26 грудня 2019

3. Вихідні дані до роботи Методи розпізнавання мови, методи автоматичного розпізнавання мови, автоматичне розпізнавання мови на мікропроцесорних системах, вибір-аналіз дискретних сигналів, методи мінімізації шумових небажаних ефектів

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
1. Вступ, 2. Аналіз існуючих методів та засобів автоматичного розпізнавання мови 3. Вибір оптимального методу та засобів автоматичного розпізнавання мови 4. Реалізація системи розпізнавання мови, 5. Обґрунтування економічної ефективності 6. Охорона праці та безпека в надзвичайних ситуаціях. 7. Економічна. 8. Висновки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Безпека життєдіяльності	Корж Н.Б.		
Економіка	Касьяна О.М.		
Безпека в НС	Стрижак В.С., ст. викл. каф. ОХ		
Охорона праці	Сухіна П.М.	21.11.	23.12.19

7. Дата видачі завдання 30.09.2019

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломної роботи	Термін виконання етапів роботи	Примітка
1	Аналіз існуючих методів та засобів автоматизації розподілених мереж	1.10.2019	Виконав
2	Вибір оптимальних методів та засобів автоматизації розподілених мереж	20.10.2019	Виконав
3.	Розробка системи розподілених мереж	8.11.2019	Виконав
4.	Обґрунтування економічної ефективності	20.11.2019	Виконав
5.	Охорона праці та безпека в надзвичайних ситуаціях	25.11.2019	Виконав
6	Економія	27.11.2019	Виконав
7	Попередній захист дипломної роботи	28.11.2019	Виконав
8	Захист дипломної роботи	26.12.2019	

Студент _____ (підпис) Чесна В.А. (прізвище та ініціали)

Керівник роботи _____ (підпис) Лудик Н.С. (прізвище та ініціали)

АНОТАЦІЯ

Часник В.А. Методи та засоби автоматичного розпізнавання мови на базі мікроконтролерної системи // Дипломна робота // Часник Василь Анатолійович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно – інформаційних систем та програмної інженерії, група СІм – 61 // Тернопіль, 2019 // с. – 116, рис. – 23, табл. – 8, аркушів А1 – 10, додат. – 1, бібліогр. – 96.

Ключові слова: РОЗПІЗНАВАННЯ МОВИ, ВЕЙВЛЕТ-АНАЛІЗ, НЕЙРОННА МЕРЕЖА, МІКРОКОНТРОЛЕРНА СИСТЕМА, STM32.

У дипломній роботі досліджено існуючі методи та засоби автоматичного розпізнавання мови на базі мікроконтролерних систем, визначено їх переваги і недоліки та обґрунтовано актуальність розробки методів і засобів побудови таких систем.

Запропоновано використання нового методу виявлення ознак в мовленнєвому сигналі, у основі якого лежить вейвлет-аналіз попередньо обробленого сигналу. Визначено оптимальний метод автоматичного розпізнавання мови, який здійснюється за допомогою двох блоків: front-end та back-end процесорів. Front-end процесор займається зчитуванням звукового сигналу, його попередньою. фільтрацією, розбиттям на фрейми та накладанням вікна, визначенням початку та кінця мовлення та виділенням ознак за допомогою вейвлет-перетворення. Back-end процесор займається порівнянням представлення звукового сигналу у вигляді вектора ознак, з базою шаблонів відомих представлень мовленнєвих сигналів, використовуючи повнозв'язну штучну нейронну мережу.

Розроблено систему автоматичного розпізнавання мови на базі мікроконтролерної системи на базі плати розробки «STM32 Nucleo» з попереднім моделюванням процесу розпізнавання мови засобами програмного пакету Matlab.

ANNOTATION

Methods and means of automatic speech recognition on microcontroller system // Master thesis // Chasnyk Vasyl Anatoliiovych // Ternopil Ivan Pul'uj National Technical University, Faculty of Computer Information Systems and software engineering, group CIm – 61 // Ternopil, 2019 // p. – 116, fig. – 23, tab. – 8, A1 Sheets – 10, Add. – 1, Ref. – 96.

Keywords: SPEECH RECOGNITION, WAVELET ANALYSIS, NEURAL NETWORK, MICROCONTROLER SYSTEM, STM32.

In the thesis, there were studied the current methods and means of automatic speech recognition on microcontroller systems. There were determined advantages and disadvantages of the mentioned above methods. There were justified the actuality of development of methods and means of such systems construction.

There was proposed the new method of detecting features in a speech signal, which is based on wavelet analysis of a pre-processed signal. The optimal method of automatic speech recognition was determined. This method is implemented using two blocks: front-end and back-end processors. The front-end processor reads the audio signal, filters, splits signal into frames, and overlays window, determines the beginning and end of speech, and extracts speech features using wavelet transform. The back-end processor compares the presentation of the audio signal as a feature vector with the database of patterns of known representations of speech signals, using a fully connected artificial neural network.

There was developed the system of automatic speech recognition on microcontroller system based on the development board «STM32 Nucleo» with preliminary modeling of speech recognition process.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ 8
ВСТУП 9
РОЗДІЛ 1 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ ТА ЗАСОБІВ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВИ 13
1.1. Огляд літератури 13
1.1.1. Мікроконтролерні системи розпізнавання мовлення 19
1.2. Визначення базової моделі розпізнавання мовлення 20
1.3. Класифікація систем розпізнавання мовлення 22
1.3.1. Системи розпізнавання неізольованих слів 23
1.3.2. Системи розпізнавання неперервного мовлення 23
1.3.3. Системи розпізнавання спонтанного мовлення 24
1.4. Процес автоматичного розпізнавання мовлення 24
1.4.1. Методи аналізу мовлення 24
1.4.2. Методи виявлення ознак 25
1.4.3. Методи моделювання 26
1.4.4. Методи співставлення 32
1.5. Застосування автоматичного розпізнавання мовлення 33
1.6. Висновки до розділу 1 35
РОЗДІЛ 2 ВИБІР ОПТИМАЛЬНИХ МЕТОДІВ ТА ЗАСОБІВ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВИ 36
2.1. Front-end процесор 37
2.1.1. Попередня обробка. Розбиття на фрейми та накладання вікна 37
2.1.2. Визначення моментів початку та кінця мовлення 38
2.1.3. Виявлення ознак 39
2.2. Back-end процесор 50
2.2.1. Штучні нейронні мережі 50
2.3. Апаратне забезпечення мікроконтролерної системи 57
2.3.1. Мікроконтролер STM32F446RET 57

2.3.2. Плата розробки на базі мікроконтролера STM32F446RET 59
2.4. Висновки до розділу 2 60
РОЗДІЛ 3 РЕАЛІЗАЦІЯ СИСТЕМИ РОЗПІЗНАВАННЯ МОВИ 61
3.1. Попередня обробка та виявлення ознак..... 61
3.2. Навчання штучної нейронної мережі..... 64
3.3. Результати 65
3.4. Висновки до розділу 3 66
РОЗДІЛ 4 ОБҐРУНТУВАННЯ ЕКОНОМІЧНОЇ ЕФЕКТИВНОСТІ 68
4.1. Визначення стадій технологічного процесу та загальної тривалості проведення НДР 69
4.2. Визначення витрат на оплату праці та відрахувань на соціальні заходи. 71
4.3. Розрахунок витрат на електроенергію 75
4.4. Розрахунок витрат на матеріали 76
4.5. Розрахунок суми амортизаційних відрахувань 76
4.6. Обчислення накладних витрат..... 77
4.7. Складання кошторису витрат та визначення собівартості НДР 77
4.8. Розрахунок ціни НДР..... 78
4.9. Визначення економічної ефективності і терміну окупності капітальних вкладень..... 79
4.10. Висновки до розділу 4 81
РОЗДІЛ 5 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ 82
5.1. Охорона праці..... 82
5.2. Підтримання безпеки життєдіяльності в системі "людина - життєве середовище" 85
5.2.1. Діяльність людини 86
5.2.2. Середовище життєдіяльності людини 87
5.2.3. Природне середовище..... 87
5.2.4. Вплив діяльності людини на навколишнє середовище..... 88
5.2.5. Пожежна небезпека..... 90

5.3. Виробничий і побутовий травматизм, наслідки і заходи щодо його попередження. 91
5.4. Висновки до розділу 5 93
РОЗДІЛ 6 ЕКОЛОГІЯ 95
6.1. Методи узагальнення екологічної інформації 95
6.2. Отримання енергії за рахунок альтернативних джерел 97
6.2.1. Задачі використання НДЕ 98
ВИСНОВКИ 101
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 103
Додаток А Тези конференцій 111

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ПЛІС	програмована логічна інтегральна схема
ПММ	приховані моделі Маркова
НМ	нейронні мережі
МОВ	метод опорних векторів
DTW	dynamic time warping
АЦП	аналого-цифровий перетворювач
ЦАП	цифро-аналоговий перетворювач
ШІМ	широтно-імпульсна модуляція
LED	light emitting diode
COM	communication port
USB	universal serial bus

ВСТУП

Не важко зрозуміти привабливість мовлення як засіб для взаємодії між людиною та комп'ютером. Для більшості людей мова є найшвидшим та найлегшим способом комунікації, на відміну будь-яким іншим способам (наприклад, набору або написання тексту). Таким чином, мовлення є більш природним та інтуїтивно зрозумілим для середньостатистичного користувача комп'ютера, що змушує практично всіх лідерів індустрії вважати, що забезпечення взаємодії людини з комп'ютером через мовлення є перспективним напрямком розвитку машинно-людинних інтерфейсів.

Актуальність теми роботи. Розпізнавання мовлення, також відоме як автоматичне розпізнавання мовлення, визначається як процес перетворення мовного сигналу в набір слів певним алгоритмом, який може бути реалізований як системна програма, або процес перетворення акустичного сигналу в набір слів.

Сучасні комп’ютерні системи обробки мови досягнули великої складності та високої точності внаслідок використання складних алгоритмів обробки сигналів та появи потужних обчислювальних засобів. Тим не менш, ріст продуктивності вбудованих процесорів (які також називаються мікроконтролерами) дозволив реалізувати розпізнавання мови в споживчій електроніці, де вартість має найбільше значення.

Застосування дешевих мікроконтролерів для автоматичного розпізнавання мови накладає обмеження на складність системи внаслідок малого об’єму оперативної та постійної пам’яті і невеликої обчислювальної потужності. Тому вибір параметрів та методів її реалізації є досить складною задачею.

В опублікованих роботах, у цьому напрямі, пропонувалось застосування спеціалізованих інтегральних схем (ASIC) або окремих спеціалізованих процесорів. Недоліком таких систем є висока вартість та складність розробки. Також пропонувалось використання недорогих стандартних

мікроконтролерів, однак достатню точність розпізнавання, у таких системах, досягнуто лише для дикторів, на мовленні яких системи навчено.

На основі проведеного аналізу існуючих методів та засобів автоматичного розпізнавання мови поставлено за мету створити недорогу систему розпізнавання мови на базі мікроконтролера STM32F446. Користуючись стандартними інструментами розробки і застосувавши новий метод попередньої обробки голосового сигналу для мікроконтролера та навчання класифікаційної нейронної мережі, необхідно запрограмувати мікроконтролер початкового рівня для автоматичного розпізнавання мови з достатньою точністю.

Метою роботи є дослідження існуючих методів та засобів автоматичного розпізнавання мовлення на базі мікроконтролерної системи та створення недорогої системи розпізнавання мови на базі мікроконтролера STM32F446, застосувавши новий метод попередньої обробки голосового сигналу для мікроконтролера та навчання класифікаційної нейронної мережі, для автоматичного розпізнавання мовлення з достатньою точністю.

Для досягнення вказаної мети в роботі поставлено наступні **задачі**:

- аналіз наукових публікацій щодо автоматичного розпізнавання мови, зокрема розпізнавання мови на базі мікроконтролерних систем;
- дослідження алгоритмічного та математичного забезпечення автоматичного розпізнавання мови;
- формалізація алгоритмічного та математичного забезпечення запропонованих методів та засобів автоматичного розпізнавання мови;
- побудова моделі автоматичного розпізнавання мови на базі запропонованих методів та засобів;
- порівняння існуючих систем автоматичного розпізнавання мови на базі мікроконтролерних систем з побудованою системою;
- апробація запропонованих методів та засобів автоматичного розпізнавання мови на базі мікроконтролерної системи.

Об'єкт дослідження: процес автоматичного розпізнавання мови на базі мікроконотролерної системи.

Предмет дослідження: методи і засоби автоматичного розпізнавання мови на базі мікроконотролерної системи.

Методи дослідження. Для вирішення поставлених задач використано наступні методи: аналіз та узагальнення – при проведені аналізу існуючих методів та засобів автоматичного розпізнавання мови; теорії аналізу часових рядів, теорії машинного навчання та добування даних – при розробці запропонованого методу розпізнавання мови; проектування та програмування – при побудові системи автоматичного розпізнавання мови на базі мікроконтролерної системи; експеримент та вимірювання – для апробації запропонованих методів та засобів.

Наукова новизна одержаних результатів. Запропоновано поліпшений метод виявлення ознак у мовленнєвому сигналі, який базується на використанні вейвлет-аналізу сигналу за допомогою вейвлету Добеші четвертого порядку, з розбиттям сигналу на одинадцять частотних смуг.

Для реалізації системи автоматичного розпізнавання мови використано мікроконтролерну систему на базі мікроконтролера STM32F446.

Практичне значення одержаних результатів. Можливим є впровадження запропонованих методів та засобів автоматичного розпізнавання мови на базі мікроконтролерної системи та їх реалізація у вигляді системи, яка дає змогу автоматизувати керування іншими системами за допомогою набору голосових команд.

Публікації. Результати дослідження апробовано на VIII Міжнародній науково-технічній конференції молодих учених та студентів. «Актуальні задачі сучасних технологій» (27-28 листопада 2019 р.) Тернопільського національного технічного університету імені Івана Пулюя та на VII науково-технічній конференції Тернопільського національного технічного університету імені Івана Пулюя «Інформаційні моделі, системи та технології» (11-12 грудня 2019 року) у вигляді тез конференцій.

Структура роботи. Робота складається з пояснювальної записки та графічної частини. Пояснювальна записка складається із вступу, шести розділів, висновків, переліку посилань та додатку. Обсяг роботи: пояснювальна записка – 116 арк. формату А4, графічна частина – 10 аркушів формату А1.

РОЗДІЛ 1
АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ ТА ЗАСОБІВ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВИ

Не важко зрозуміти привабливість мовлення як засіб для взаємодії між людиною та комп'ютером. Для більшості людей мова є найшвидшим та найлегшим способом комунікації, на відміну будь-яким іншим способам (наприклад, набору або написання тексту) [1]. Отже, мовлення є більш природним та інтуїтивно зрозумілим для середньостатистичного користувача комп'ютера, що змушує практично всіх лідерів індустрії вважати, що забезпечення взаємодії людини з комп'ютером через мовлення є перспективним напрямку розвитку машинно-людинних інтерфейсів [2].

1.1. Огляд літератури

Машинне розпізнавання мовлення з’явилось на початку 1920-х. Перша машина «Radio Rex», яка використовувалась для розпізнавання мовлення, була виготовлена в 1920 [3]. В 1950-х, більшість систем розпізнавання мовлення досліджували спектральний резонанс під час кожної голосної частини мовлення, які були відокремлені від мовного сигналу за допомогою банків фільтрів та логічних схем [4]. К. Девіс, Р. Бідульф та С. Балашек розробили систему розпізнавання відокремлених цифр для одного диктора, яка базувалась на обчислені спектральних резонансів голосних ділянок кожної цифри [5]. В цей період в 1959, в лабораторіях Масачусетського технологічного університету Л. Форгі та Дж. Форгі було розроблено ще одну систему розпізнавання голосних ділянок слів, де десять голосних були вбудовані в формат /b/-vowel/t/ та розпізнавались незалежно від диктора [6]. Спектральна інформація отримувалась за допомогою аналізатора на основі банку фільтрів і апроксимації резонансів голосового тракту, було зроблено для

визначення голосного який вимовляється. Основна складність у розпізнаванні мовлення полягала в нерівномірності шкали часу в мовних подіях.

На початку 1960-х рр. Т. Мартін та його колеги з лабораторій авіоніки Повітряних сил США розробили набір елементарних методів нормалізації часу, що базувались на здатності точно визначати моменти початку та кінця мовлення, що значно знизило варіативність оцінки розпізнавання [7]. Паралельно, у Радянському Союзі Т. Вінцюком пропонувалось використання динамічних методів програмування для вирівнювання часу (загальновідоме як динамічне викривлення часу), включаючи алгоритми розпізнавання нерозділених слів. У 1960 р. було розроблено різну апаратуру спеціального призначення, оскільки комп'ютери ще не були досить швидкими для цієї задачі [8]. Японська система, яку описали Дж. Сузукі та К. Наката з лабораторії радіодосліджень у Токіо, була апаратним розпізнавачем голосних [9]. Було сформовано складний аналізатор спектра на основі банку фільтрів разом з логікою, що пов'язує виходи кожного каналу спектрального аналізатора (зваженим чином) до схеми вибору голосних, і далі, для вибору голосного використовувалася логічна вибору за мажоритарним принципом. Досягненням у 1960-х роках стало революційне дослідження Д. Редді в області безперервного розпізнавання мовлення шляхом динамічного відстеження фонем [10]. Врешті-решт, це дослідження залишається надзвичайно успішною програмою дослідження розпізнавання мовлення в університеті Карнегі Меллон та все ще є світовим лідером у системах неперервного розпізнавання мовлення.

За рахунок досліджень, проведених в 1970-х роках, В. Величком і Н. Загоруйком в Росії [11], Г. Сакое і С. Чіба в Японії [12], та Ф. Ітакура в Сполучених Штатах, технології розпізнавання мовлення ізольованого слова або дискретного розпізнавання висловів набули життєздатності та практичності. Новаторська пропозиція використання ідей розпізнавання образів у розпізнаванні мовлення було висунуто в російськими дослідженнях, японські дослідники показали застосування методів динамічного

програмування. Робота Ф. Ітакури представляє, як ідеї лінійного прогнозного кодування можуть бути використані для системи розпізнавання мовлення за допомогою відповідного вимірювання відстані на основі спектральних параметрів лінійного прогнозного кодування [13]. У лабораторіях компанії IBM було досліджено розпізнавання мовлення з великим словником для трьох різних завдань, а саме: мова «New Raleigh» для простих запитів до баз даних [14], мова лазерних патентних текстів для транскрибування лазерних патентів [15], та завдання офісної кореспонденції під назвою «Tangora» для диктування простих нагадувань.

Проблема розпізнавання мовлення, з нероздільними словами, була головною темою досліджень 1980-х років. Основною метою була розробка надійної системи, здатної розпізнавати слова при звичайному мовлені (наприклад, цифр) на основі відповідності об’єднаному шаблону окремих слів. М. Ласрі вперше розробив систему розпізнавання мовлення на основі ознак на початку 1980 року. В своїй роботі він досліджував мовні спектрограми букв і цифр [16]. Було сформульовано та впроваджено широкий спектр алгоритмів, які базуються на узгодженні зведеної структури окремих слів, включаючи дворівневий підхід динамічного програмування Г. Сакое з «Nippon Electric Corporation» [17], метод однопрохідного проходу Дж. Бріддла та М. Брауна в Спільному дослідницькому підрозділі мови у Великобританії [18], підхід на основі побудови рівня К. Майєрса та Л. Рабінера [19], та підхід на основі синхронної побудови кадрів К. Лі та Л. Рабінера в лабораторії компанії Bell [20]. У деяких лабораторіях (в першу чергу компанії IBM, Інститут оборонних аналізів та «Dargon systems») досліджувалась методологія прихованих моделей Маркова. В кінці 1970-х років метод розпізнавання мовлення, який використовує поєднання миттєвих кепстральних коефіцієнтів та коефіцієнтів їхніх поліномів першого та другого порядку, які тепер називаються Δ коефіцієнтами та Δ-кепстральними коефіцієнтами, як основні спектральні ознаки для розпізнавання мовлення [21]. У 1980-х роках була знову запропонована ідея застосування нейронних мереж для розпізнавання

мовлення. Нейронні мережі вперше були введені в 1950-х роках, але вони зазнали різних труднощів через практичні проблеми. У 1980-х роках було досягнуто більш глибокого врахування сильних та слабких сторін технології, а також розуміння взаємозв'язку цієї технології з класичними методами класифікації [22] [23] [24]. В даний час більшість систем розпізнавання мовлення базуються на статистичній роботі, розробленій у 1980 році, та значне розширення було здійснено у 1990-х роках. Ряд нововведень відбувся у сфері розпізнавання мовлення у 90-х роках. Складна задача розпізнавання шаблонів була трансформована в задачу оптимізації, що стосується мінімізації емпіричної помилки розпізнавання [25]. Ця основна парадигматична зміна була викликана визнанням того, що функції розподілу мовного сигналу неможливо точно вибрати або визначити, і що теорія рішень Байеса стає непридатною в цих умовах. По суті, мета проектування розпізнавача повинна полягати в тому, щоб досягти найменшої помилки розпізнавання, а не забезпечити найкраще пристосування функції розподілу до заданого (відомого) набору даних, як це відбувається при застосуванні критерієм Байеса. Ця концепція мінімізації помилок породила ряд методів, таких як дискримінаційна підготовка та методи на основі ядра. Було досліджено багато різних методик для підвищення надійності систем розпізнавання мовлення, яка зменшується через невідповідність між умовами навчання та тестування через фонові шуми, голосову індивідуальність, мікрофони, канали передачі, реверберацію в кімнаті, тощо. Ключові методи включають максимальну ймовірність лінійної регресії [26], декомпозицію моделі [27], паралельний склад моделі [28] та структурний апостеріорний максимум [29].

У 21 столітті було розроблено варіаційну методику оцінювання та кластеризації Байеса [30]. У розпізнаванні мовлення, проблема адаптивного навчання була вирішена Дж. Річарді [31], також ним був запропонований алгоритм активного навчання для автоматичного розпізнавання мовлення. У 2007 році було проаналізовано різницю акустичних особливостей між спонтанним та читаним мовленням за допомогою широкомасштабної бази

мовних даних [32]. Вся схема кепстрального аналізу була реалізована апаратно за допомогою ядер інтелектуальної власності, щоб прискорити процес розпізнавання. Звідси, було запроваджено системи розпізнавання мовлення на основі програмованих логічних інтегральних схем (ПЛІС). Дослідниками в [33] була розроблена програмна реалізація на ПЛІС «Virtex-4» сімейства розпізнавачів литовських слів. Використання цього програмного забезпечення компанії «Microblaze» спільно з ядрами інтелектуальної власності для обробки сигналів дозволило збільшити процес розпізнавання слів у 1,55 рази [34], але це все одно було недостатньо для роботи в режимі реального часу. При розпізнаванні мовлення з використанням програмованих логічних інтегральних схем використовується дві методи, а саме: паралелізація та конвеєрна техніка. Це є гнучкою архітектурою для створення систем на відміну від реалізацій на спеціалізованих інтегральних схемах. Рішення на основі вбудованих процесорів мають середню успішність розпізнання в 80% та обмежений розмір словника в 32 [35] або 75 [36] команд. Основна проблема такого розпізнавача полягає у забезпеченні виконання в реальному часі функції вилучення ознак (особливо на етапах порівняння). Методи вилучення та узгодження особливостей, які реалізовані на ПЛІС [37] [38] [39] [40] або GPU [41] [42] можуть працювати незалежно на відміну від попереднього підходу.

Робота, виконана М. Давелом та його колегами [43] [44], показала, що «розробники» з обмеженим лінгвістичним досвідом можуть створювати точні моделі «шляхом використання процесу розробки з належною підтримкою інструментів». У березні 2008 року було введено найперший багатомодельний мовний додаток карт Google для мобільних пристроїв. Більшість сучасних систем розпізнавання мовлення використовують приховані моделі Маркова для вирішення часової мінливості мови та моделей Гауссової суміші, щоб вирішити, наскільки кожен стан ПММ вписується в кадр або коротке вікно кадрів коефіцієнтів, що представляє акустичний вхід. Альтернативний спосіб оцінювання придатності полягає у використанні нейронної мережі, яка

приймає кілька кадрів коефіцієнтів як вхідні дані і генерує ймовірності над станами ПММ як вихід. Показано, що глибокі нейронні мереж, які мають кілька прихованих шарів і навчаються за допомогою нових методів, показують кращі результати ніж у додатку карт Google для мобільних пристроїв в різних тестах розпізнавання мовлення, іноді з великим запасом [45]. Для досягнення найкращої точності розпізнавання високотехнологічні глибокі нейронні мережі приймаються як акустичні моделі. Завдяки значно меншим розмірам слідів, глибока НМ може забезпечити поліпшення точності в порівнянні з акустичними моделями у додатку карт Google для мобільних пристроїв [46]. Раніше розпізнавання мовлення на портативних комп’ютерах та смартфонах досліджувалося в рамках програми «Transtac», спонсорованої Агентством передових оборонних дослідницьких проектів США, де для телефонів були розроблені системи перекладу мовлення [47] [47] [48]. Сьогоднішня сучасна технологія обробки мовлення - це майбутнє поле пошуку Google за допомогою голосу. Можливість здійснювати пошук в Інтернеті просто, називаючи їх на своїх мобільних телефонах, вперше з'явилася на телефонах iPhone та Android [49]. В Google, голосовий пошук на японській мові призвів до значного спрощення, та пришвидшив процес роботи з новими мовами в загальному. На сьогодні японська та корейська є одними з найбільших міжнародних мов. Голосовий пошук на китайському мандарині та кантонській, який було запущено Google у 2009 та 2010 роках, описано у [50] та [51]. У вересні 2013 року деякі аспіранти та дослідники Техаського університету в Далласі розробили нові системи, які можуть розпізнавати мовленнєві голоси, незважаючи на умови, які можуть ускладнити розпізнавання голосу, наприклад, шепіт, розмова через різні емоції чи розмова з закритим носом. У нещодавньому конкурсі щодо визнання диктора, організатори від Національного інституту стандартів та технологій надіслали понад 80 мільйонів випробувань голосової перевірки з додаванням шуму - природних фонових звуків або штучних комп'ютерних звуків - до більш ніж 50 університетів, науково-дослідних лабораторій та компаній у всьому світі.

Команди повинні були визначити, чи є записи мовлення від певних дикторів чи ні. Нещодавно в 2014 році з'явилася програма JUSTSPEAK - універсальне рішення для управління голосом для невізуального доступу до операційної системи Android. В порівнянні з уже існуючими системами, JUSTSPEAK пропонує два внески. По-перше, вона дозволяє системне голосове управління будь-яких програм на Android. По-друге, вона забезпечує більш винахідливий і природний інтерфейс з підтримкою декількох голосових команд в одній і тій же вимові [52]. Служби автоматичного розпізнавання мовлення Google використовуються в рамках JUSTSPEAK з метою розпізнавання розмовної мови користувачів [46].

1.1.1. Мікроконтролерні системи розпізнавання мовлення

Застосування мікроконтролерів для автоматичного розпізнавання мови накладає обмеження на складність системи внаслідок малого об’єму оперативної та постійної пам’яті і невеликої обчислювальної потужності. Однак, внаслідок стрімкого росту популярності компактних пристроїв які потребують реалізації розпізнавання мовлення, було здійснено значну кількість досліджень та реалізацій таких систем, зокрема.

У роботі К. Накамури та інших [53] пропонувалось використання спеціалізованої інтегральної схеми для реалізації системи розпізнавання мови. Процес розпізнавання мовлення такої системи полягав у використанні аналізу вхідного сигналу на основі кодування з лінійним предиктором. В результаті такого аналізу отримувався вектор ознак, який шляхом векторного квантування перетворювався на вхідний символ ПММ, яка здійснювала класифікацію вхідного сигналу. Перевагами роботи К. Накамури та ін. є висока продуктивність такої системи та високий рівень інтеграції, однак недоліками є висока кінцева вартість та складність розробки.

Процес розпізнавання мовлення у системі Ші Ю. на основі мікроконтролера на базі ядра «8051» [54] поділявся на два блоки: front-end процесор та back-end процесор. Front-end процесор здійснював зчитування голосового сигналу та обчислював кепстральні коефіцієнти лінійного

передбачення за алгоритмом Шчура. отриманий вектор коефіцієнтів кодується дельта-модуляцією зі змінною крутістю, після чого він порівнюється з шаблонами, закодованими таким самим чином. Перевагами такої системи є невелика вартість, однак використання непродуктивного мікроконтролера наклало обмеження розмір словника, швидкість та точність розпізнавання.

Система автоматичного розпізнавання мовлення, запропонована Б. Бурчардом [55], аналогічно до [54], умовно поділялась на два блоки: front-end процесор та back-end процесор. Front-end процесор здійснював зчитування голосового сигналу, однак виділення ознак здійснювалось шляхом аналізу Фур'є на основі якого здійснювався дискримінаційний аналіз. Back-end процесор здійснював класифікацію отриманого вектору ознак використовуючи ПММ. Внаслідок використання продуктивного і водночас дорогого сигнального процесора вдалось досягнути високої точності та швидкості розпізнавання, однак вартість запропонованої системи була висока.

1.2. Визначення базової моделі розпізнавання мовлення

Розпізнавання мовлення, також відоме як автоматичне розпізнавання мовлення, визначається як процес перетворення мовного сигналу в набір слів певним алгоритмом, який може бути реалізованим як системна програма, або процес перетворення акустичного сигналу в набір слів [56] [57]. Автоматичне розпізнавання мовлення є галуззю з одним з найбільших темпів розвитку з області науки і інженерії мовних технологій. Дослідження обробки мовлення та комунікації більшою мірою спонукалось бажанням людей побудувати механічну модель, яка б змогла повторити можливості вербальної комунікації людини. Головною метою систем автоматичного розпізнавання мовлення є розробка нових технік та систем голосового вводу машин. Математичне представлення системи автоматичного розпізнавання мовлення є прямим рівнянням, яке містить акустичний front-end блок, блок акустичної моделі, блок моделі мови та пошуковий блок, що зображено на рис. 1.1.

Одним з традиційних підходів до неперервного розпізнавання мовлення з великим словником є припущення простої ймовірнісної моделі представлення мови, де вибраний набір слів $\widehat{W}$, генерує спостережену акустичну послідовність A, з ймовірністю $P(W|\ A)$, де W – мовна модель.

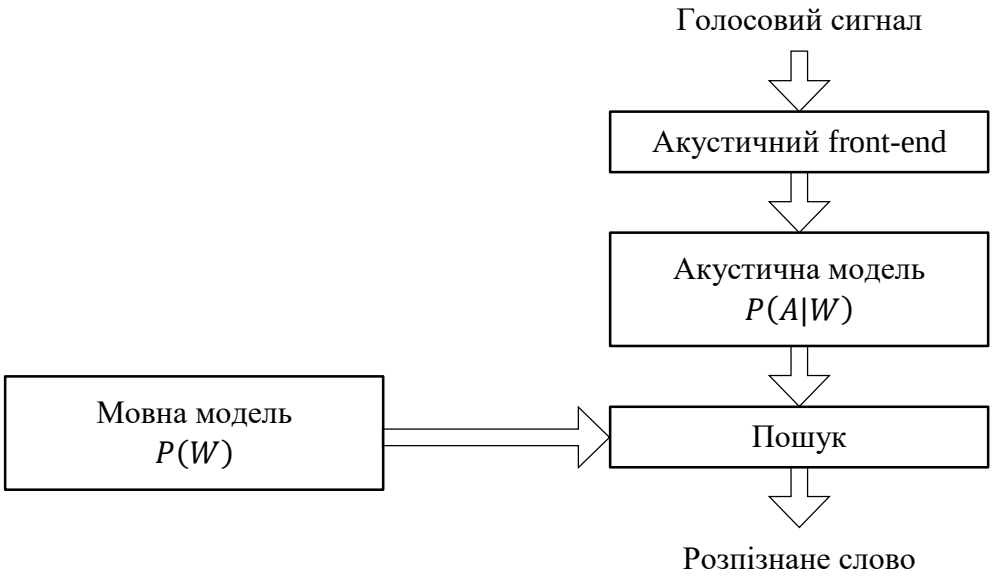


Рис. 1.1. Базова модель розпізнавання мовлення

Завданням є декодування стрічки слова, базуючись на спостереженій акустичній послідовності, таким чином щоб декодована стрічка мала максимальну апостеріорну ймовірність:

$$\widehat{W} = \underset{W}{\operatorname{argmax}}\, P(W|A) \tag{1.1}$$

Використовуючи правило Байеса, рівняння (1.1) можна записати як:

$$P(W|A) = \frac{P(A|W)P(W)}{P(A)} \tag{1.2}$$

Оскільки ймовірність акустичної спостереженої послідовності $P(A)$ є незалежною від W, виразу (1.1) можна записати як:

$$\widehat{W} = \underset{W}{\operatorname{argmax}} P(A|W)P(W) \tag{1.3}$$

Перший терм виразу (1.3) $P(A|W)$, зазвичай називається акустичною моделлю, оскільки він визначає ймовірність послідовності акустичного спостереження, з умовою на стрічці слова. Звідси обчислюється $P(A|W)$. Для систем розпізнавання мовлення з великим словником є необхідною розробка статистичних моделей для підсловесних мовних одиниць (використання лексикону для опису будови слова), та далі обумовлювати послідовності слів та визначати ймовірність з акустичної моделі через стандартні методи конкатенації. Другий терм виразу (1.3) $P(W)$, називається мовною моделлю. Він ілюструє ймовірність пов’язану з обумовленою послідовністю слів. Така модель мови може включати як і синтаксичні так і семантичні обмеження мовлення та задачі його розпізнавання.

1.3. Класифікація систем розпізнавання мовлення

Системи розпізнавання мовлення можна класифікувати на декілька різних класів, описуючи, які типи висловлювань вони мають здатність розпізнавати.

Системи розпізнавання ізольованих слів. При розпізнавані ізольованих слів, кожне висловлення повинне бути виділене періодами тишини перед і після мовлення. Одночасно приймається одне слово або одне висловлювання. Ці системи мають стан "слухати" або "не слухати", коли диктору доводиться робити паузу між висловлюваннями. Ізольовані висловлювання може бути найкращою назвою для цього класу слів [58]. Розпізнавання ізольованих слів засноване на принципі, що сигнал у заданому інтервалі запису складається з ізольованого слова, перед яким передує тиша або інший фоновий шум. Отже, передбачається, що мовні сегменти можуть бути надійно відмежовані від мовчазних сегментів [59]. Система розпізнавання ізольованих слів

розглядається як система, де окреме висловлювання трактується відповідно до двох неявних припущень. Перше припущення передбачає, що мовлення, яке слід розпізнавати, повинне складатися з одного слова або фрази, і сказане має бути визнане повною сутністю без явних знань щодо фонетичного змісту слова або фрази. Друге припущення полягає в тому, що кожне вимовлене висловлювання повинно мати точно визначені кінцеві точки, тобто початок і кінець. Одне з застосувань, де можна використовувати цей тип розпізнавання мовлення - це системи командування та управління.

1.3.1. Системи розпізнавання неізольованих слів

Неізольовані слова (висловлювання) певним чином схожі на окремі висловлювання, але дозволяють окремим висловлювання "збігатися" з певним проміжком у кілька мілісекунд між ними, що називається мовним [60]. Слова, які розділені між собою з невеликим проміжком (мовний код) між ними, відомі як неізольовані слова, а система, що використовується для цих типів слів, відома як система розпізнавання неізольованих слів. Нагадуючи розпізнавання відокремлених слів, розпізнавання неізольованих слів має властивість, що основною одиницею розпізнавання мовлення є слово або фраза значною мірою. [61] досліджував три алгоритми, розроблені для розпізнавання неізольованих слів: Дворівневий підхід DP, підхід рівня побудови та підхід One Pass, і було встановлено, що вони забезпечують найкращу однакову відповідність рядку з однаковою шкалою відповідності для розпізнавання неізольованих слів. Обчислювальна ефективність, вимога до об’єму пам’яті та простота реалізації в апаратах реального часу були чинниками, за якими вони відрізняються. Пізніше [62] розробили дикторозалежну систему розпізнавання неізольованих цифр, шляхом використання необмеженої техніки динамічного викривлення часу, в якій кожна цифра розпізнавалася шляхом оцінки відстані щодо відповідності вхідної розмовної цифри із збереженим шаблоном.

1.3.2. Системи розпізнавання неперервного мовлення

При безперервному розпізнаванні мови комп'ютери визначають вміст в той же час як диктор говорить природною мовою. Ці типи розпізнавачів дуже

важко розробити, оскільки вони використовують спеціальні методи для визначення кінцевих точок висловлювань. У режимі безперервного розпізнавання мовлення слова з'єднуються разом, а не розділяються паузами (кодами мови). Як результат, на ефективність систем безперервного розпізнавання мовлення впливають, з точки зору невідомої граничної інформації (кінцевих точок) про слова, спільна артикуляція, накладання оточуючих фонем і швидкість мовлення.

1.3.3. Системи розпізнавання спонтанного мовлення

Спонтанна мова може розглядатися як природна мова, але яка не повторюється. Система розпізнавання спонтанного мовлення повинна вміти обробляти різноманітні природні особливості мовлення, такі як слова, що накладаються одне на одного та навіть незначні заїкання.

1.4. Процес автоматичного розпізнавання мовлення

Основна мета розпізнавання мовлення полягає в тому, щоб машина мала змогу чути (розпізнавання мовлення), розуміти (обробка природної мови) та відповідати назад (синтез мови) на розмовну інформацію [63] та [64]. У системах автоматичного розпізнавання мовлення зазвичай метою є аналіз, вилучення та класифікація або розпізнавання змісту інформації, яку говорять люди. Система розпізнавання динаміків працює в чотири етапи:

1. аналіз мовлення;
2. вилучення ознак;
3. моделювання;
4. тестування.

1.4.1. Методи аналізу мовлення

Коли вимовляються фрази, особливості вимови диктора (наприклад, довжина, форма та стать голосових шляхів), мовний зміст, акустичне оточення та частота мовлення одночасно впливають на акустику всієї мови [65]. Аналіз мовного сигналу проводиться за допомогою наступних методів:

- Сегментаційний аналіз. Тут мовний сигнал аналізується за допомогою розміру кадру та зсуву в діапазоні 10-30 мс для отримування інформації з мовлення.

- Субсегментаційний аналіз. Цей аналіз в основному використовується для отримання характеристик стану збудження [66];

- Супрасегментаційний аналіз. У цьому випадку характер поведінки мовлення аналізується за допомогою аналізу розміру кадру.

1.4.2. Методи виявлення ознак

Основна мета вилучення ознак - знаходження набору властивостей висловлювання, які мають акустичні кореляції з мовним сигналом, тобто параметри, які певним чином можна обчислити або оцінити за допомогою обробки сигналу. Такі параметри називають ознаками. Для отримання достатньої точності розпізнавання мовлення головним питанням є вибір відповідних ознак мовного сигналу. Для вилучення ознак використовуються два типи аналізу, а саме: методи часового та спектрального аналізу. Сам мовленнєвий сигнал використовується для аналізу в попередньому випадку, в той час, як наступних в основному використовується спектральне представлення мовного сигналу. Під виділенням ознак мається на увазі відстеження необхідного інформаційного вмісту з мовного сигналу та відкидання небажаної інформації. Під час виділення ознак, розміри вхідного вектора зменшуються при збереженні дискримінаційної сили мовного сигналу.

Виділення ознак можна розділити на три етапи. Спектральний аналіз: Тут генеруються «сирі» ознаки, що описують огинаючі спектрів потужності мовних сегментів. Другий крок називається параметричним перетворенням, де складається вектор ознак, що складається з деяких статичних та динамічних ознак. Нарешті, останній етап, відомий як статистичне моделювання, використовується для перетворення векторів ознак у більш надійні та щільні вектори, які передаються back-end процесору. Деякі способи вилучення ознак,

разом із їх властивостями та відповідним методом реалізації наведені в таблиці в додатку Б.

1.4.3. Методи моделювання

Покоління мовних моделей, що використовують вектор ознак, є основною метою методики моделювання. Методи моделювання мовлення класифікуються на два типи: ті, що розпізнають мовлення дикторів та ті, що проводять ідентифікацію дикторів. Остання, автоматично визначає, хто говорить на основі індивідуальної інформації, інтегрованої у висловлювання, в той час як розпізнавання диктора в свою чергу поділяється на два типи, дикторозалежні та дикторонезалежні. У дикторозалежному режимі, розпізнавач витягує всі характеристики диктора в акустичному сигналі [67]. У дикторонезалежному випадку такі особливості не беруться до уваги, і вилучається лише необхідне повідомлення. Нижче згадується кілька методів моделювання, які використовуються в системах розпізнавання мовлення.

1.4.3.1. Акустично-фонетичний підхід

Встановлення акустичного фонетичного підходу ґрунтується на знаходженні звуків мовлення та наданні відповідних міток цим звукам, які припускають, що в розмовній мові існує обмежена кількість фонем, і для класифікації цих фонем використовується набір акустичних властивостей. Найпершим кроком вищезгаданого підходу є спектральний аналіз мовного сигналу разом із виявленням особливостей, який перетворює спектральні вимірювання в ряд характеристик, що описують широкі акустичні властивості різних фонем. На другому етапі, відомому як сегментація та маркування, висловлювання спочатку сегментується на стабільні акустичні області, а потім ще одна фонетична мітка з'єднується з кожною сегментованою частиною. На останньому та заключному кроці дійсне слово або набір слів визначаються із фонетичної послідовності, яка виробляється сегментацією в маркування. Розпізнавання «поломаного» телефону, моделювання Гауссової суміші та метод опорних векторів [68] [69] - це три основні методи, які широко

використовуються для ідентифікації мовлення, однак багато додатків відмовляються від використання акустичного фонетичного підходу [10].

1.4.3.2. Метод розпізнавання шаблонів

В останні роки розпізнавання образів набуло значної популярності і широко використовувалося в різних програмах. Ідея використання розпізнавання образів виникла з прагнення наслідувати поведінку розпізнавання людини. Основна ідея розпізнавання шаблонів полягає в оптимальному вилученні шаблонів на основі певних умов і тим самим розділення даних на різні класи. Більше того, метод розпізнавання образів виконує класифікацію, не маючи уявлення про розподіл вимірювань у різних групах. Метод розпізнавання шаблону став переважаючим методом розпізнавання мовлення протягом останніх шести десятиліть [70]. Важливою особливістю методу розпізнавання шаблонів [71] [72] є те, що він використовує добре сформульований математичний фреймворк та генерує послідовне подання мовних образів для послідовного порівняння шаблонів із набору промаркованих навчальних зразків за допомогою формального алгоритму навчання. Етап порівняння шаблонів цього методу включає пряме порівняння невідомих висловлювань, тобто мовлення, яке слід розпізнати, з кожним можливим шаблоном, вивченим на етапі навчання, для визначення невідомого відповідно до найкращої відповідності шаблону.

1.4.3.3. Методи шаблонного підходу

Основна ідея цього підходу полягає в тому, що набір ідеальних мовних зразків зберігається у вигляді опорних зразків, що представляють словник слів-кандидатів. Потім розпізнавання проводиться шляхом зіставлення векторів ознак невідомого висловлювання (мовлення, яке слід розпізнати) з усіма даними в довідковому шаблоні, щоб досягти найкращої відповідності. Помилки, які виникають через сегментацію або класифікацію менших, акустично більш змінних одиниць, таких як фонеми, можна уникнути, створивши шаблони для цілих слів. Однією з ключових ідей шаблонного підходу є отримання типових послідовностей мовних кадрів для шаблону

(слова) за допомогою певної процедури та покладання на використання локальних спектральних вимірювань відстані для порівняння зразків. Ще одна ідея полягає у використанні певної форми динамічного програмування для узгодження зразків для врахування варіацій частоти розмов у розмовниках, а також для повторень цього ж слова.

1.4.3.4. Метод стохастичного підходу

Стохастичне моделювання вимагає використання імовірнісних моделей для роботи з невизначеною або неповною інформацією. Основним ривком прогресу парадигми розпізнавання мовлення стало впровадження статистичних методів, зокрема стохастичної обробки прихованих моделей Маркова [73] [74] на початку 1970-х років [75]. Більше 30 років потому ця методологія все ще найбільш поширеною. Приховані моделі Маркова - це найбільш застосований стохастичний підхід на сьогодні. ПММ, які давно домінують у світі акустичного моделювання, використовуються для характеристики мовного сигналу як параметричного стохастичного процесу та статистично характеризують варіацію мовної одиниці (слово, фонема або склад). ПММ мають ті ж переваги для розпізнавання дикторів, що і для розпізнавання мовлення. Надзвичайно надійні моделі мовленнєвих подій можуть бути розроблені лише з невеликою кількістю вимог або інформації, що йде з навчальними висловлюваннями. Системи розпізнавання дикторів, засновані на архітектурі ПММ, використовували моделі дикторів, отримані з багатословного речення, одного слова або фонеми. Зазвичай використовувались багатословні словосполучення (наприклад, рядок із семи до десяти цифр), а моделі для кожного окремого слова та для мовчання поєднувались на рівні речення відповідно до заздалегідь визначеної граматики рівня речень [76]. Негативний побічний ефект цього підходу полягає в тому, що ПММ не забезпечують великої точності щодо процесу розпізнавання. Отже, часто виникають проблеми під час аналізу помилок системи ПММ, у спробі покращити її продуктивність. З іншого боку, пильне введення знань значно покращило системи на основі ПММ.

1.4.3.5. Метод динамічного викривлення часу

Метод динамічного викривлення часу - це метод, який вимірює відстань між кожним вхідним кадром та кожним опорним кадром, використовуючи алгоритм динамічного програмування, щоб знайти найдрібніші викривлення шаблону, і визначає найкращу відповідність, мінімізуючи відстань між вхідним кадром та опорним кадром. Автоматичне розпізнавання мовлення – одне з добре відомих застосувань DTW, воно застосовується для розпізнавання з різною швидкістю вимови. Вирівнювання часу різних мовних сигналів є центральною проблемою для вимірювання відстані в розпізнаванні мови. Невеликий зсув призводить до неправильного розпізнавання. Динамічне викривлення часу - це ефективний метод вирішення проблеми вирівнювання часу. Алгоритм DTW фокусується на вирівнюванні двох послідовностей векторів ознак циклічним викривленням осі часу до тих пір, поки не буде знайдено оптимальної відповідності між двома послідовностями. Неперервність відіграє меншу роль у DTW, ніж у інших методах узгодження шаблонів. DTW - алгоритм, який переважно підходить для узгодження послідовностей з відсутньою інформацією, за умови, що є достатньо довгі сегменти для знаходження відповідності.

1.4.3.6. Метод квантування векторів

Векторне квантування - це класична методика квантування від обробки сигналів, яка дозволяє моделювати функції густини ймовірностей шляхом розподілу векторів-прототипів. В основному він використовувався для стиснення даних. Воно працює шляхом поділу великого набору точок (векторів) на відносно невеликі групи, що мають приблизно однакову кількість точок, найближчих до них. Кожна група представлена своєю центральною точкою, як у методі k-середніх та деяких інших алгоритмах кластеризації. Основною метою стиснення даних є зниження швидкості надсилання бітів для передачі або зберігання даних, зберігаючи необхідну надійність даних. Вектор ознак може представляти всі можливі параметри кодування мовлення, включаючи коефіцієнти лінійного прогнозного

кодування та кепстральні коефіцієнти [77]. Системи автоматичного розпізнавання мовлення часто використовують квантування векторів [16]. Для використання ефективних кодових книг для еталонних моделей та пошуку кодових книг використовується векторне квантування. Тестоване висловлювання оцінюється всіма можливими кодовими книгами, а система автоматичного розпізнавання мовлення вибирає слово, чия кодова книга дає найменшу відстань [78]. Оскільки записи в кодовій книзі не впорядковані і можуть походити з будь-якої частини навчальних слів, таким чином, кодові книги не мають жодної конкретної інформації про час. З іншого боку, деякі непрямі тривалі сигнали зберігаються, оскільки записи кодової книги використовуються для мінімізації середньої відстані між усіма тренувальними кадрами, а кадри, відповідні більш довгим акустичним сегментам (наприклад, голосні звуки), частіше зустрічаються у навчальних даних. Часто декількох кодових слів достатньо, щоб представити велику кількість кадрів під час відносно стійких ділянок голосних, таким чином дозволяючи ще більше кодових слів представляти короткі, динамічні частини слів. Отже, що стосується словників подібних слів, використання векторного квантування може мати перевагу перед іншими методами порівняння автоматичного розпізнавання мовлення.

1.4.3.7. Інтелектуальний підхід

Застосування гібридного акустично-фонетичного підходу та розпізнавання образів являє собою інтелектуальний підхід (підхід на основі штучного інтелекту) [16]. Вся інформація щодо лінгвістичної, фонетичної та спектрограми використовується в інтелектуальному підході. Інтелектуальне інженерне проектування використовує пряме та явне включення мовленнєвих знань експертів у систему розпізнавання. Однак, чиста інтелектуальна інженерія розвивалась за рахунок досліджень, що стосувались експертних систем. Через складність у кількісному оцінюванні експертних знань такий підхід має обмежений успіх. Впровадження багатьох рівнів людських знань,

таких як фонетика, фонотактика, синтаксис, семантика та прагматика – все ще є серйозною проблемою.

1.4.3.8. Підхід на основі штучних нейронних мереж

Ще одна технологія, яка була (повторно) впроваджена наприкінці 1980-х, - це концепція штучних нейронних мереж. Нейронні мережі вперше були введені в 1950-х роках, але спочатку вони не мали успіху в отриманні помітних результатів [79]. Порівняно з ПММ, нейронні мережі не припускають статистичних властивостей ознак і мають численні особливості, що робить їх привабливими моделями розпізнавання. Нейронні мережі дозволяють проводити дискримінаційне навчання природним та ефективним чином, коли вони використовуються для оцінки ймовірності сегмента мовної ознаки. Недавній розвиток у розпізнаванні мовлення є і досі предметом суперечок. Знання або обмеження не запрограмовані в окремих одиницях, правилах чи процедурах, а розподіляються між багатьма простими обчислювальними одиницями, що стосується моделей на основі штучних нейронних мереж. Дані моделі в основному залежать від наявності надійних стратегій навчання. Апаратна реалізація моделей штучних нейронних мереж через їх простоту та рівномірність елементів обробки стає більш привабливою. Однак для підготовки даних потрібна велика кількість ітерацій, а в деяких випадках її застосування є дорожчим.

1.4.3.9. Метод опорних векторів

Одним з найбільш перспективних та використовуваних алгоритмів навчання для класифікації та регресії є метод опорних векторів. МОВ ефективно застосовується до багатьох додатків розпізнавання образів. МОВ знаходить розділову поверхню з великим запасом між навчальними зразками двох класів у просторі високої розмірності зображень, що неявно вводиться обчислювально ефективним відображенням ядра, а великий запас передбачає кращу здатність до узагальнення відповідно до теорії статистичного навчання [80]. Машина опорних векторів - це лінійна машина, яка працює у просторі ознак високої розмірності, утвореному нелінійним відображенням N-мірного

вхідного вектора x, у K-вимірний простір ознак $(K > N)$ за допомогою використання відображення $u(x)$ [81]. Алгоритм спочатку був розроблений в 1985-му для проблеми класифікації з відокремленими даними, і цей метод згодом був удосконалений для обробки нероздільних даних. Типовий факт МОВ полягає в тому, що завдання навчання зводиться до квадратичного програмування за допомогою так званих множників Лагранжа. У МОВ функції ядра, які відповідають умовам Мерсера, використовуються для виконання всіх операцій у режимах навчання та тестування. В останні роки МОВ виявились домінуючою технікою для класифікації у багатьох сферах, таких як біоінформатика, розпізнавання рукописних символів, машинний зір тощо. Це узагальнений лінійний класифікатор з функціями пристосування максимальної відстані. Керуючи розмірами вектора моделі МОВ, він контролює складність моделі.

1.4.4. Методи співставлення

Невідоме слово узгоджується з відомим словом за допомогою однієї з наступних методик [82]:

- Співставлення цілих слів. Тут розпізнавач порівнює вхідний цифровий звуковий сигнал із заздалегідь записаним шаблоном слова. Ця техніка вимагає набагато менше обробки, ніж знаходження відповідності між частинами слів, але вона вимагає, щоб користувач (або хтось) записав кожне слово, яке буде розпізнано - іноді кілька сотень тисяч слів. Цілі шаблони слів також потребують великої кількості пам’яті (від 50 до 512 байт на слово) і корисні лише в тому випадку, якщо словник розпізнавання відомий при розробці програми [83];

- Співставлення частин слів. Класифікатор шукає частини слова - як правило, фонеми - і потім здійснює подальше розпізнавання їхніх шаблонів. Ця техніка вимагає більше обробки, ніж знаходження відповідності цілих слів, але її необхідно менше сховища (від 5 до 20 байт на слово). Крім того, вимову слова можна вгадати з англійського тексту, не вимагаючи від користувача заздалегідь вимовляти слово.

В [82] [84] [85] обговорювали, що дослідження в галузі автоматичного розпізнавання мовлення виконувались протягом останніх трьох десятиліть, що лише систем розпізнавання мовлення на основі розпізнавання цілих слів знайшли практичне використання та здобули значних успіхів.

1.5. Застосування автоматичного розпізнавання мовлення

Найважливіший вплив розпізнавання мовлення полягає в тому, чи можна повністю інтегрувати сприятливі технології з додатками. Як успішно інтегрувати мову в програми, часто залежить від характеру інтерфейсу користувача та програми. Елегантний користувальницький інтерфейс вимагає ретельного розгляду конкретної групи користувачів програми та доставки програми, яка працює ефективно та економічно. Одною винятковою проблемою у програмах розпізнавання мовлення є те, що розпізнавання мовлення (як і розуміння) є незадовільним. Тому важливо, щоб додатки використовували необхідні інтерактивні методи обробки помилок, щоб мінімізувати їх вплив. Таким чином, розробники додатків повинні мати належні знання про сильні та слабкі сторони основних мовленнєвих технологій та створити відповідне місце для ефективного використання технологій розпізнавання та розуміння мовлення.

Існує три широкі класи додатків:

1. Центр дзвінків на основі «хмари», IVR (інтерактивна голосова відповідь): сюди входять широко використовувані додатки з інформаційного доступу: від Tellme по телефону до Microsoft Exchange Unified Messaging;

2. Диктування, командування та управління на основі ПК: На ПК є ряд програм диктування. Це інструмент для забезпечення доступності, але ще не готовий до «мейнстріму»;

3. Вбудоване управління командами на пристроях: Існує велика кількість пристроїв, які не мають типової клавіатури ПК або миші, і традиційну програму графічного користувальницького інтерфейсу (GUI) не можна

безпосередньо розширити [86]. Мобільні телефони та автомобільні сценарії також дуже підходять для мовних програм. Через фізичний розмір і обмеженість рук, зайнятих очей, традиційна модель взаємодії додатків GUI потребує значних змін.

Окрім цих, у таблиці 1.1 наведено ще кілька застосувань розпізнавання мовлення.

Таблиця 1.1

Практичні застосування автоматичного розпізнавання мовлення

№	Область	Застосування
1	Освітній сектор	Навчання студентів іноземним мовам, правильній вимові іноземних слів
2	Побутовий сектор	Розумні прилади: холодильники, мікрохвильові печі, пральні машини; автомобілі, засоби охорони
3	Військовий сектор	Високоефективні винищувачі, керування боєм, тренування контролерів авіатрафіку, телефонія та інші області
4	Сектор штучного інтелекту	Робототехніка
5	Медичний сектор	Транскрибування медичних записів
6	Загальний сектор	Автоматичне транскрибування, телеметрія, керування авіатрафіком, продаж товарів, банківські системи
7	Перекладацький сектор	Додатки перекладу мовлення з однієї мови на іншу

Одним з найбільш значущих способів спілкування між людьми є мова і основним середовищем, яке використовується для цього, є мовлення. Широке розгортання програм та служб, що надаються системами автоматичного розпізнавання мовлення, було зустріте низкою практичних обмежень. У даному розділі коротко описано сучасні системи автоматичного розпізнавання мовлення та основні методи розпізнавання мовлення.

Незважаючи на успішне застосування розпізнавання мовлення на ринку та в повсякденному житті, технологія все ще далеко не бездоганна, також не було вирішено всі технічні проблеми.

1.6. Висновки до розділу 1

Одним з найбільш значущих способів спілкування між людьми є мова і основним середовищем, яке використовується для цього, є мовлення. Широке розгортання програм та служб, що надаються системами автоматичного розпізнавання мовлення, було зустріте низкою практичних обмежень. У даному розділі коротко описано сучасні системи автоматичного розпізнавання мовлення, зокрема мікроконтролерні системи та основні методи розпізнавання мовлення.

Незважаючи на успішне застосування розпізнавання мовлення на ринку та в повсякденному житті, технологія все ще далеко не бездоганна, також не було вирішено всі технічні проблеми. Основними проблемами наявних мікроконтролерних систем автоматичного розпізнавання мовлення є висока вартість, складність розробки та низька продуктивність.

На основі проведеного аналізу існуючих методів та засобів автоматичного розпізнавання мови поставлено за мету створити недорогу систему розпізнавання мови на базі мікроконтролера STM32F446, застосувавши новий метод попередньої обробки голосового сигналу для мікроконтролера та навчання класифікаційної нейронної мережі; запрограмувати мікроконтролер початкового рівня для автоматичного розпізнавання мовлення з достатньою точністю.

РОЗДІЛ 2
ВИБІР ОПТИМАЛЬНИХ МЕТОДІВ ТА ЗАСОБІВ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВИ

Розробка системи автоматичного розпізнавання мовлення складається з двох основних частин: розробка front-end процесора та back-end процесора. Front-end процесор займається наступними задачами: зчитування звукового сигналу, його попередня фільтрація, розбиття на фрейми та застосування віконної функції, попередня обробка мовленнєвого сигналу та виділення ознак. Back-end процесор, в свою чергу, займається порівнянням звукового сигналу, представленого front-end процесором, з базою шаблонів відомих мовленнєвих сигналів, також представлених front-end блоком.

Загальна схема системи автоматичного розпізнавання мовлення представлена на рис 2.1.

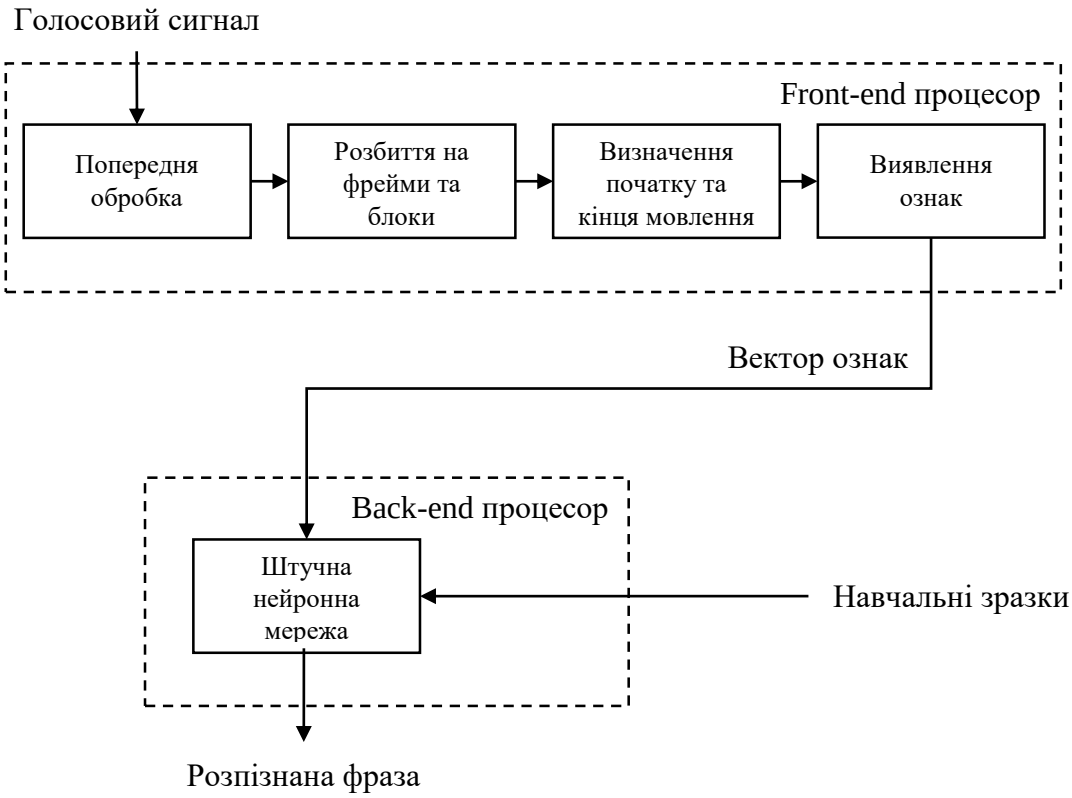


Рис. 2.1. Загальна схема системи автоматичного розпізнавання мовлення

2.1. Front-end процесор

У даній системі front-end процесор виконує функцію зчитування звукового сигналу, його перетворення в цифрову форму, попередню фільтрацію та обробку, визначає моменти початку та кінця мовлення, після чого виділяє з сигналу ознаки, за якими back-end процесор буде здійснювати їх класифікацію.

Загальна схема back-end процесора системи автоматичного розпізнавання мовлення представлена на рис 2.2.

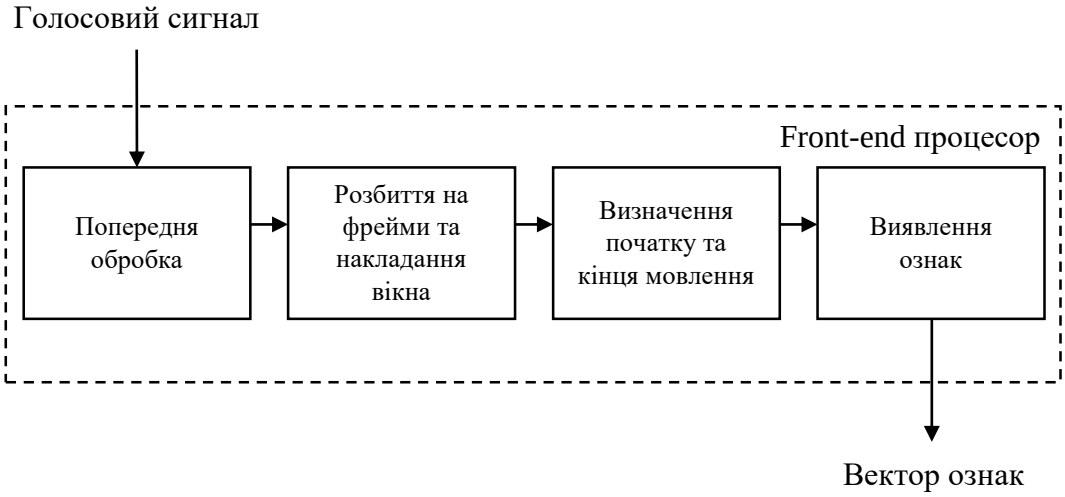


Рис. 2.2. Загальна схема front-end процесора системи автоматичного розпізнавання мовлення

2.1.1. Попередня обробка. Розбиття на фрейми та накладання вікна

Процес попередньої обробки сигналу x включає в себе фільтрацію за допомогою КІХ-фільтру низьких частот:

$$y(n) = \sum_{i=0}^{P} b_i x(n-i)\,, \tag{2.1}$$

де, $x(n)$ – вхідний сигнал, $y(n)$ – вихідний сигнал, P – порядок фільтра, b_i – коефіцієнти фільтра. Після чого отриманий сигнал проріджується у 8 разів.

Далі сигнал розбивається на фрейми довжиною 30 мс з перетином у 5 мс. При частоті дискретизації у 6000 Гц, фрейм довжиною у 30 мс міститиме $N = 180$ відліків сигналу, а перетин складатиме 30 відліків. До кожного фрейму застосовується віконна функція:

$$y(n) = x(n)w(n), \tag{2.2}$$

де $x(n)$ – n-й відлік вхідного фрейму, $y(n)$ – n-й відлік вихідного фрейму, при чому: $0 \leq n \leq N - 1$, а віконна функція $w(n)$ визначається як:

$$w(n) = \frac{25}{46} - \frac{21}{46} cos \frac{2\pi n}{N - 1}. \tag{2.3}$$

2.1.2. Визначення моментів початку та кінця мовлення

Наступним кроком є визначення моментів початку та кінця мовлення. Для кожного фрейму сигналу $s(n)$ з довжиною L відліків обчислюються P_s – потужність та Z_s – кількість нульових моментів:

$$P_s = \frac{1}{L} \sum_{n=1}^{L} s^2(n), \tag{2.3}$$

$$Z_s = \frac{1}{2L} \sum_{n=1}^{L} |\operatorname{sgn} s(n) - \operatorname{sgn} s(n-1)|, \tag{2.4}$$

де $\operatorname{sgn} s(n)$ визначається як:

$$\begin{cases} \operatorname{sgn} s(n) = +1, & \text{якщо } s(n) > 0; \\ \operatorname{sgn} s(n) = -1, & \text{якщо } s(n) \leq 0, \end{cases} \tag{2.5}$$

Далі визначається момент початку та кінця мовлення, або – функція голосової активації $VAD(t)$:

$$\begin{cases} VAD(t) = 1, & \text{якщо } W_s(t) \geq t_W = 0, W_s(m) \geq t_W; \\ VAD(t) = 0 & \text{інакше,} \end{cases} \tag{2.6}$$

де:

$$W_s(t) = P_s(t)\big(1 - Z_s(t)\big), \tag{2.7}$$

$$t_W = \mu_W + \alpha\delta_W, \tag{2.7}$$

де μ_W – середнє значення, δ_W – відхилення, обчислене для фрейму, α – константа, яка обчислюється за:

$$\alpha = 0.2\delta_W^{0.8}. \tag{2.8}$$

Далі сигнал в кожному фреймі нормалізується за наступною формулою:

$$y_i = \frac{x_i - \bar{x}}{\sigma}, \tag{2.9}$$

де, x_i – i-й відлік початкового сигналу, y_i – i-й відлік нормалізованого сигналу, $\bar{x}$ – середнє значення сигналу, σ – стандартне відхилення.

2.1.3. Виявлення ознак

Процес виявлення ознак в сигналі передбачає його аналіз за допомогою певного математичного апарату, в результаті якого сигнал перетворюється в форму, з якої легко виділити необхідні ознаки.

2.1.3.1. Вейвлет-аналіз

Вейвлет перетворення $\mathcal{W}_x(b, a)$ сигналу $x(t)$ визначено як:

$$\mathcal{W}_x(b,a) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-b}{a}\right) dt\,. \tag{2.29}$$

Таким чином, вейвлет-перетворення обчислюється як скалярний добуток $x(t)$ та зміщеними та масштабованими версіями функції $\psi(t)$, так названим вейвлетом.

Якщо вважати $\psi(t)$ імпульсною характеристикою смугового фільтра, то під вейвлет-аналізом можна розуміти аналіз смуговими фільтрами. Варіюючи параметр масштабу a змінюється центральна частота і ширина смуги пропускання. Зміна параметра b при сталому параметрі a впливає на зміщення у часі, так ця зміна може розглядатись як згортка $x(t)$ з оберненим у часі відмасштабованим вейвлетом:

$$\mathcal{W}_x(t,a) = |a|^{-\frac{1}{2}} x(t) * \psi_a(t), \qquad \psi_a(t) = \psi\left(\frac{-t}{a}\right). \tag{2.30}$$

Множник $|a|^{-\frac{1}{2}}$ введено для того всі відмасштабовані функції $|a|^{-\frac{1}{2}}\psi\left(\frac{t}{a}\right)$ з $a \in \mathbb{R}$ мають однакову енергію.

Оскільки аналізуюча функція $\psi(t)$ масштабується та не модулюється, на відміну ядра віконного перетворення Фур'є, вейвелт-аналіз часто називається масштабно-часовим аналізом а не частотно-часовим. Однак, обидва вони пов'язані одне з одним своїми інтерпретаціями через смугові фільтри. На рис. 2.3 зображено приклади ядер віконного перетворення Фур'є як і вейвлет-перетворенням. Як видно з рис. 2.3, варіація затримки у часі b та/або масштабуючого параметра a не впливає на форму ядра вейвлет-перетворення. Однак, часова та частотна роздільна здатність залежать від a. Для високих аналізованих частот (малих значень a), високою є часова локалізація але частотна роздільна здатність є низькою. З іншого боку, для низьких частот, високою є частотна роздільна здатність, при цьому часова є низькою. В той же

час як віконне перетворення Фур'є є аналізом зі сталою смугою пропускання, вейвлет-аналіз можна розуміти як Constant-Q аналіз або октавний аналіз.

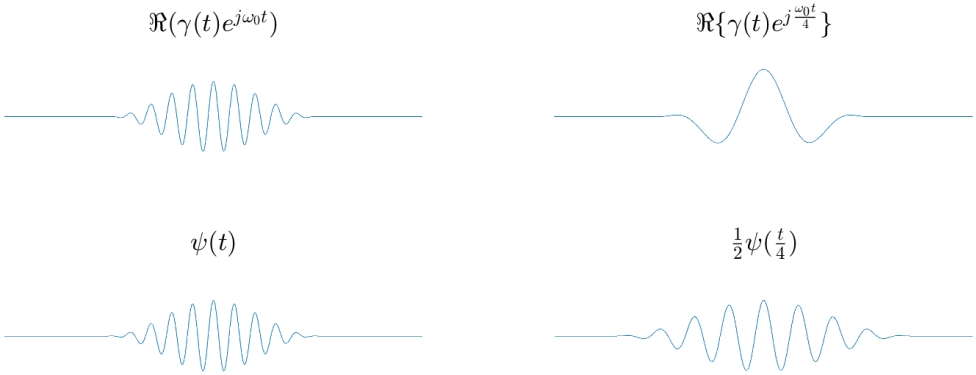


Рис. 2.3. Порівняння аналізуючих ядер віконного перетворення Фур'є як і вейвлет-перетворенням для аналізу високих та низьких частот

Під час використання перетворення для отримання кращого розуміння властивостей сигналу, необхідно забезпечити досконалу реконструкцію сигналу з його представлення. Інакше таке представлення повністю або частково не має ніякого сенсу. Для вейвлет-перетворення, наступною є умова ідеального реконструювання сигналу з його представлення:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty. \tag{2.31}$$

де $\Psi(\omega)$ позначає перетворення Фур'є вейвлета. Ця умова також відома як умова допустимості для вейвлета $\psi(t)$

Очевидно, що для забезпечення умови (2.31) вейвлет має забезпечити виконання наступної умови:

$$\Psi(0) = \int_{-\infty}^{\infty} \psi(t) dt = 0. \tag{2.32}$$

Також, $|\Psi(\omega)|$ має швидко спадати при $|\omega| \to 0$ та при $|\omega| \to \infty$. Таким чином, $\psi(t)$ має мати імпульсну характеристику смугового фільтра. Оскільки імпульсна характеристика виглядає як маленька хвиля, дане перетворення названо вейвлет-перетворенням.

Апарат дискретного вейвлет-перетворення в основному був розроблений Майером Малатом та Добеші для ортонормального випадку [87] [88] [89] [90]. Оскільки біортогональні вейвлети формально підпадають під даний апарат [91] [92], апарат дискретного вейвлет-перетворення буде описано для більш загального біортогонального випадку.

У даному випадку робиться припущення що набори:

$$\begin{aligned} \psi_{mn}(t) &= 2^{-\frac{m}{2}}\psi(2^{-m}t - n), \\ \tilde{\psi}_{mn}(t) &= 2^{-\frac{m}{2}}\tilde{\psi}(2^{-m}t - n), \end{aligned} \qquad n \in \mathbb{Z} \tag{2.33}$$

є базисом для $L_2(\mathbb{R})$ тим самим задовольняючи умову біортогональності. Сигнал представлено наступним чином:

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} d_m(n)\psi_{mn}(t)\,, \tag{2.34}$$

де

$$d_m(m) = \mathcal{W}_x^{\tilde{\psi}}(2^m n, 2^m) = \langle x, \tilde{\psi}_{mn} \rangle, \qquad m, n \in \mathbb{Z}. \tag{2.35}$$

Оскільки базис складається з лінійно незалежних функцій, $L_2(\mathbb{R})$ може бути представлений як пряма сума підпросторів:

$$L_2(\mathbb{R}) = \ \ldots \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \ldots \tag{2.36}$$

де

$$W_m = span\{\psi(2^{-m}t - n), n \in \mathbb{Z}\}, \qquad m \in \mathbb{Z}. \tag{2.37}$$

Кожен підпростір W_m покриває певну частотну смугу. Для піддіапазонних сигналів з (2.34) отримується:

$$y_m(t) = \sum_{n=-\infty}^{\infty} d_m(n)\psi_{mn}(t), \qquad y_m(t) \in W_m. \tag{2.38}$$

Кожний сигнал $x(t) \in L_2(\mathbb{R})$ може бути представлений як:

$$x(t) = \sum_{m=-\infty}^{\infty} y_m(t), \qquad y_m(t) \in W_m\,. \tag{2.39}$$

Далі визначаються підпростори $V_m, m \in \mathbb{Z}$ як пряма сума V_{m+1} та W_{m+1}:

$$V_m = V_{m+1} \oplus W_{m+1}. \tag{2.40}$$

Тут можна зробити припущення, що підпростори V_m містять низькочастотні сигнали та що смуга пропускання сигналів, яка розміщена у V_m зменшується зі збільшенням m.

З (2.37) (2.36) та (2.40) виводяться наступні властивості:

1. Є послідовність вкладених підпросторів:

$$\ldots \subset V_{m+1} \subset V_m \subset V_{m-1} \subset \cdots; \tag{2.41}$$

2. Масштабування $x(t)$ у два рази $(x(t) \to x(2t))$ робить відмасштабований сигнал $x(2t)$ елементом наступного, більшого підпростору і навпаки:

$$x(t) \in V_m \Leftrightarrow x(2t) \in V_{m-1}; \tag{2.42}$$

3. Якщо сформувати послідовність функцій $x_m(t)$ проекцією $x(t) \in L_2(\mathbb{R})$ на підпростори V_m, дана послідовність збігається до $x(t)$:

$$\lim_{m \to -\infty} x_m(t) = x(t), \qquad x(t) \in L_2(\mathbb{R}), \qquad x_m(t) \in V_m. \tag{2.43}$$

Таким чином, сигнал може бути наближеним з будь-якою точністю.

Через властивість масштабування (2.42) можна припустити, що підпростори V_m перекриваються масштабованою та зсунутою версією однієї функції $\phi(t)$:

$$V_m = span\{\phi(2^{-m}t - n), n \in \mathbb{Z}\}. \tag{2.44}$$

Таким чином, піддіапазонні сигнали $x_m(t) \in V_m$ виражаються як

$$x_m(t) = \sum_{n=-\infty}^{\infty} c_m(n)\phi_{mn}(t), \tag{2.45}$$

де

$$\phi_{mn}(t) = 2^{-\frac{m}{2}}\phi(2^{-m}t - n). \tag{2.46}$$

Функція $\phi(t)$ називається масштабуючою функцією.

Через рівність $V_0 = V_1 \oplus W_1$ функції $\phi_{0n}(t) = \phi(t - n) \in V_0, n \in \mathbb{Z}$ можна записати як лінійну комбінацію базисних функцій для просторів V_1 та W_1. З коефіцієнтами $h_0(2\ell - n)$ та $h_1(2\ell - n)$, $\ell, n \in \mathbb{Z}$ підходом є:

$$\phi_{0n}(t) = \sum_{t} h_0(2\ell - n)\phi_{1\ell}(t) + h_1(2\ell - n)\psi_{1\ell}(t)\,. \qquad (2.47)$$

Рівняння (2.47) також відоме як рівняння декомпозиції, для якого також використовується наступний запис:

$$\sqrt{2}\phi(2t - n) = \sum_{t} h_0(2\ell - n)\phi(t - \ell) + h_1(2\ell - n)\psi(t - \ell)\,. \qquad (2.48)$$

Ввівши відому послідовність $\{c_0(n)\}$ та підставивши (2.47) у (2.45) для $m = 0$, отримується:

$$\begin{aligned} x_0(t) &= \sum_{n} c_0(n)\phi_{0n}(t) \\ &= \sum_{n} c_0(n) \sum_{n} h_0(2\ell - n)\phi_{1\ell}(t) + h_1(2\ell - n)\psi_{1\ell}(t) \\ &= \sum_{\ell} \underbrace{\sum_{n} c_0(n)h_0(2\ell - n)\phi_{1\ell}(t)}_{c_1(\ell)} + \sum_{\ell} \underbrace{\sum_{n} c_0(n)h_1(2\ell - n)\psi_{1\ell}(t)}_{d_1(\ell)} \\ &= x_1(t) + y_1(t), \end{aligned} \qquad (2.49)$$

де $x_0 \in V_0, x_1 \in V_1$ та $y_1 \in W_1$. Цей метод дозволяє обчислити $\{c_{m+1}(\ell)\}$ та $\{d_{m+1}(\ell)\}$ з $\{c_m(n),\ m, n \in \mathbb{Z}\}$:

$$\left.\begin{aligned} c_{m+1}(\ell) &= \sum_{n} c_m(n)h_0(2\ell - n) \\ d_{m+1}(\ell) &= \sum_{n} c_m(n)h_1(2\ell - n) \end{aligned}\right\}, \qquad i, \ell \in \mathbb{Z} \qquad (2.50)$$

Як видно з даного рівняння, послідовності $\{c_{m+1}(\ell)\}$ та $\{d_{m+1}(\ell)\}$ відбуваються з вдвічі нижчою частотою вибірки за $\{c_{m+1}(\ell)\}$. Взагалі, декомпозиція (2.50) є еквівалентною аналізу двоканальним банком фільтрів

$h_0(n)$ та $h_1(n)$. Якщо прийняти що $x_0(t)$ є достатньою апроксимацією $x(t)$, та якщо відомо коефіцієнти $c_0(n)$, є можливим обчислення коефіцієнтів $c_{m+1}(n)$, $d_{m+1}(n)$, $m > 0$, і таким чином значення вейвлет-перетворення, використовуючи дискретний банк фільтрів, як зображено на рис. 2.4. Даний спосіб є найбільш ефективним для обчислення дискретного вейвлет-перетворення сигналу.

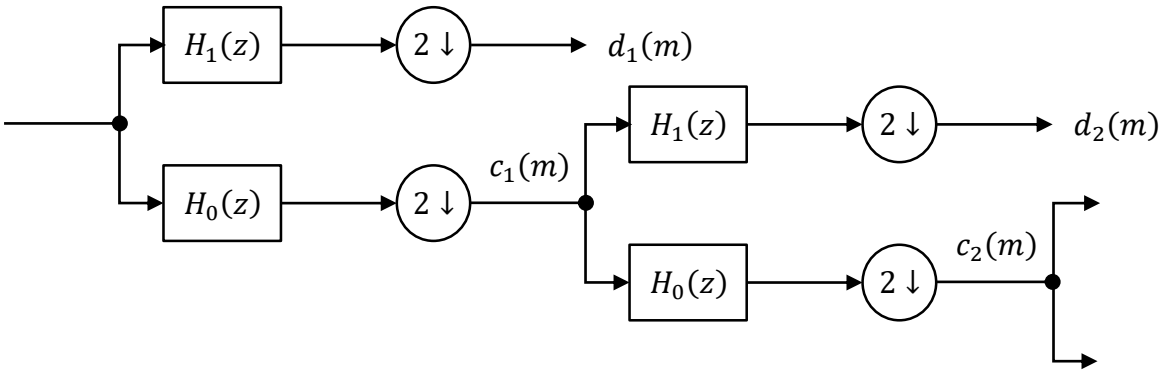


Рис. 2.4. Аналізуючий банк фільтрів для обчислення дискретного вейвлет-перетворення

І. Добеші спроектувала сімейство ортонормальних вейвлетів з максимальною кількістю нульових моментів для заданої опори [90]. Для того щоб контролювати регулярність, розглянуто наступну факторизацію $H_0(e^{j\omega})$:

$$H_0(e^{j\omega}) = \sqrt{2}\left(\frac{1+e^{-j\omega}}{2}\right)^k P(e^{j\omega}). \tag{2.51}$$

Через ортонормальність, прототипом фільтру має задовольняти наступну умову поліфазного представлення:

$$\left|H_0(e^{j\omega})\right|^2 + \left|H_0(e^{j(\omega+\pi)})\right|^2 = 2. \tag{2.52}$$

Оскільки $h_0(n)$ вважається дійсним, терм $\left|H_0\left(e^{j\omega}\right)\right|$ може бути записаний як:

$$\left|H_0\left(e^{j\omega}\right)\right|^2 = \sqrt{2}\left(cos^2\frac{\omega}{2}\right)^k M(cos\,\omega), \tag{2.53}$$

де

$$M(cos\,\omega) = \left|P\left(e^{j\omega}\right)\right|^2. \tag{2.53}$$

Підставивши (2.53) у (2.52) отримується:

$$\left(cos^2\frac{\omega}{2}\right)^k M(cos\,\omega) + \left(sin^2\frac{\omega}{2}\right)^k M(-\,cos\,\omega) = 1. \tag{2.54}$$

Використовуючи такі самі аргументи як і у попередньому випадку, (2.54) можна записати як:

$$\left(cos^2\frac{\omega}{2}\right)^k F\left(sin^2\frac{\omega}{2}\right) + \left(sin^2\frac{\omega}{2}\right)^k F(1-x) = 1, \tag{2.55}$$

або еквівалентно як:

$$(1-x)^k F(x) + x^k F(1-x) = 1, \tag{2.56}$$

де $x = \sin^2\frac{\omega}{2}$. Це, по суті, є тією ж умовою, як і у біортогональному випадку, однак тепер необхідне виконання умови $F\left(\sin^2\frac{\omega}{2}\right) \geq 0\ \forall\ \omega$, оскільки $F\left(\sin^2\frac{\omega}{2}\right) = \left|P\left(e^{j\omega}\right)\right|^2$.

І. Добеші запропонувала використати:

$$F(x) = \sum_{n=0}^{k-1} \binom{k+n-1}{n} x^n + x^k R(1-2x)\,, \qquad (2.57)$$

де $R(x)$ є непарним поліноміалом таким, що $F(x) \geq 0$ для $x \in [0,1]$. Сімейство вейвлетів Добеші виводиться для $R(x) \boxminus 0$ спектральною факторизацією $F(x)$ у $F(x) = P(x)P(x^{-1})$. Для цього, нулі $F(x)$ мають бути обчислені та згруповані у нулі всередині та ззовні одиничного кола. Далі $P(x)$ містить усі нулі всередині одиничного кола. Така факторизація результує у отримані масштабувальних функцій з мінімальним фазовим зсувом. Для фільтрів $H_0(z)$ з, як мінімум, восьми коефіцієнтами, можливою є більш семерична факторизація. Частотні характеристики, однак, є такими ж як і для випадку мінімальної фази.

На рис. 2.5 зображено деякі вейвлети Добеші, відповідні масштабуючі функції та частотні характеристики фільтрів. Можна побачити, що масштабуючі функції та вейвлети стають більш гладкими зі збільшенням довжини фільтра. Для прикладу, деякі вейвлети Добеші з максимальною симетрією, відомі як симлети, зображені на рис. 2.6.

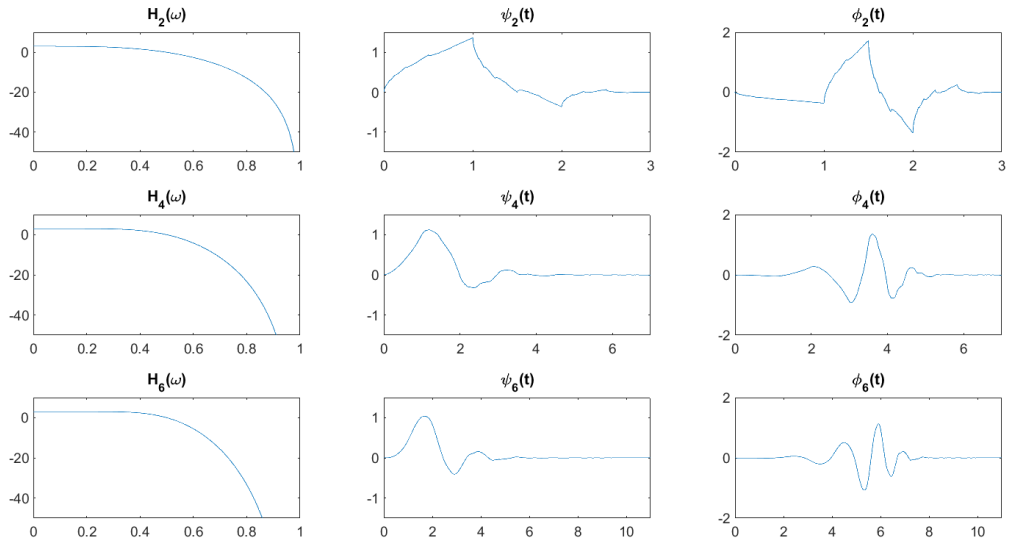


Рис. 2.5. Частотні характеристики фільтрів Добеші та відповідні масштабуючі функції та вейвлети (індекс вказує на довжину фільтра)

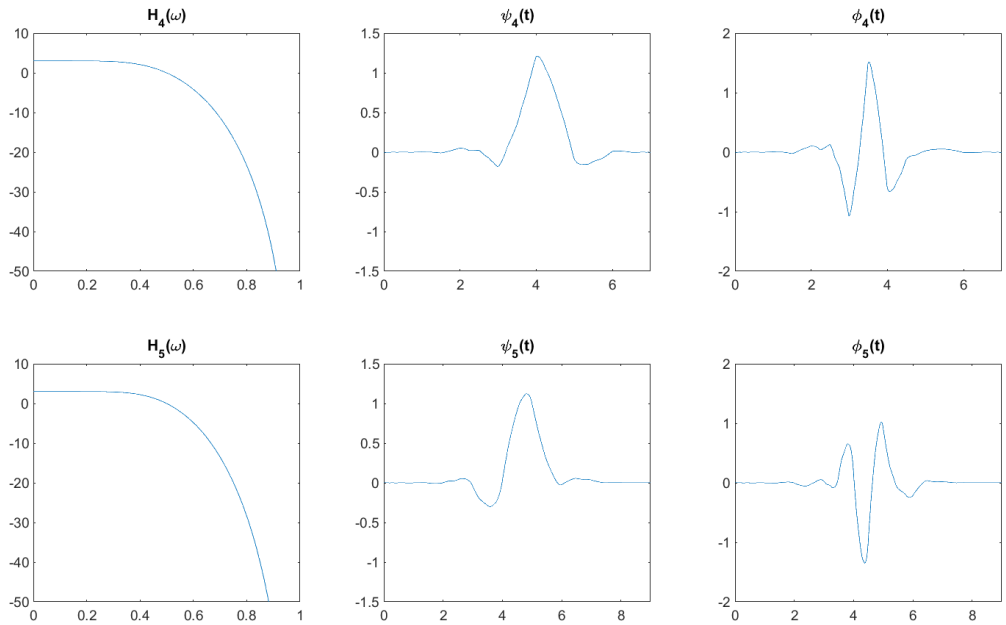


Рис. 2.6. Частотні характеристики фільтрів Добеші з максимальною симетрією та відповідні масштабуючі функції та вейвлети (індекс вказує на довжину фільтра, частотні характеристики еквівалентні зображеним на рис. 2.9)

Для виділення ознак в системі автоматичного розпізнавання мовлення, було обрано за аналізуючий вейвлет Добеші четвертого порядку з розкладом вхідного сигналу на 10 рівнів. Результатом такої декомпозиції є сигнал, розбитий на 11 логарифмічно розміщених частотних смуг. Для кожної з них обчислюються норма N, енергія E, максимум S_i^{max} та мінімум S_i^{min}, для отримання вектору ознак $\vec{F}$ довжиною в 44 елементи:

$$N_i = \sum_{n=1}^{m} |s_i(n)|, \tag{2.58}$$

$$E_i = \sum_{n=1}^{m} s_i(n)^2, \tag{2.59}$$

$$S_i^{max} = max\, s_i(n)\,, \qquad S_i^{min} = min\, s_i(n)\,, \tag{2.60}$$

де $s_i(n)$ – сигнал i-ї частотної смуги довжиною m відліків.

Після обчислення даних параметрів для всіх частотних смуг отримується вектор $\vec{F}$:

$$\vec{F} = \left(N_1, E_1, S_1^{max}, S_1^{min}, N_2, E_2, S_2^{max}, S_2^{min}, \dots, N_{11}, E_{11}, S_{11}^{max}, S_{11}^{min}\right). \tag{2.61}$$

2.2. Back-end процесор

У даній системі back-end процесор виконує функцію класифікацію вхідного звукового сигналу, з попередньою його обробкою front-end процесором. Для цього використовується штучна нейронна мережа (або конективістська система).

2.2.1. Штучні нейронні мережі

Штучний нейрон є базовим структурним елементом будь-якої штучної нейронної мережі. Штучні нейрони представляють собою прості, однотипні елементи, які імітують функціонування нервових клітин мозку. Кожен з таких нейронів має визначений стан у кожен момент часу по аналогії з природними нейронами. Модель штучного нейрона, як і його природнього прототипу складається з входів (які також називаються синапсами), соми (або тіла нейрона) та виходу (аксону) (рис. 2.7).

Стан штучного нейрона s обчислюється як сума добутків вхідних сигналів на відповідні їм ваги:

$$s = \sum_{i=1}^{n} x_i w_i, \tag{2.62}$$

де x_i – i-й вхідний сигнал нейрона, а w_i – i-та відповідна вхідному сигналу вага.

Вихід штучного нейрона є певна функція від його стану, тобто:

$$y = f(s). \tag{2.63}$$

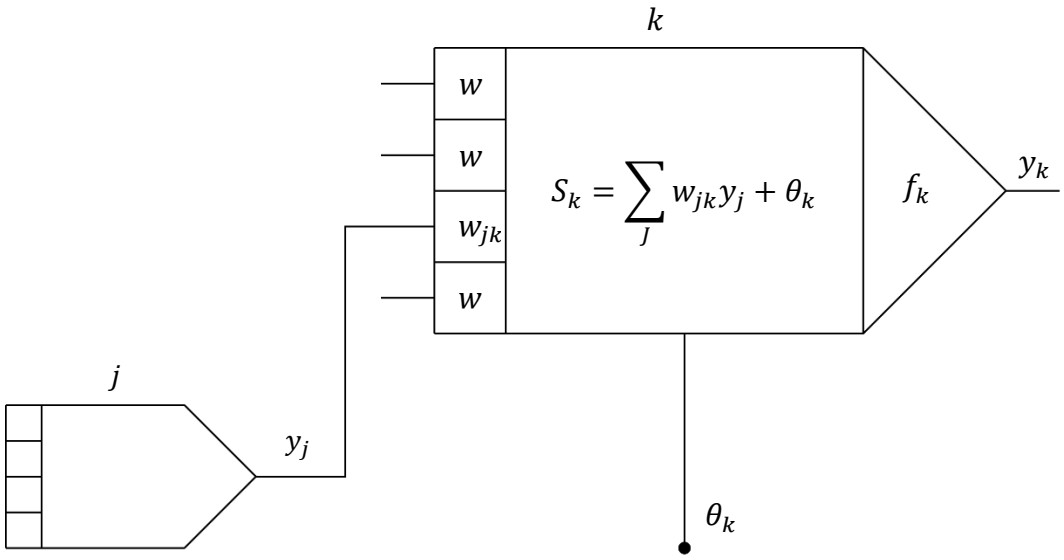


Рис. 2.7. Загальний вигляд моделі штучного штучного нейрона.
y_j – вхідний сигнал, від нейрона j; s_k – зважена сума вхідних сигналів;
f_k – активаційна функція; y_k – вихідний сигнал нейрона

Нелінійну функцію $f(s)$ прийнято називати активаційною функцією, стискуючою функцією або функцією збудження нейрона.

В якості активаційної функції в даній системі розпізнавання мовлення використовується сигмоїдна функція, зображена на рис. 2.8.

Сигмоїдна активаційна функція визначається за формулою:

$$f(x) = \frac{1}{1 + e^{-\alpha x}}, \tag{2.64}$$

де α – коефіцієнт пологості функції, який прийнято рівним 1.

Однією з найбільш поширених структур штучних нейронних мереж є мережа з декількома шарами, у якій всі аксони нейронів певного шару з'єднані зі всіма синапсами наступного. Даний вид штучних нейронних мереж прийнято називати повнозв'язними.

Для навчання штучних мереж такого виду буде використовуватись алгоритм зворотного поширення помилки. Суть алгоритму полягає в розповсюджені помилки від виходів штучної нейронної мережі до її входів.

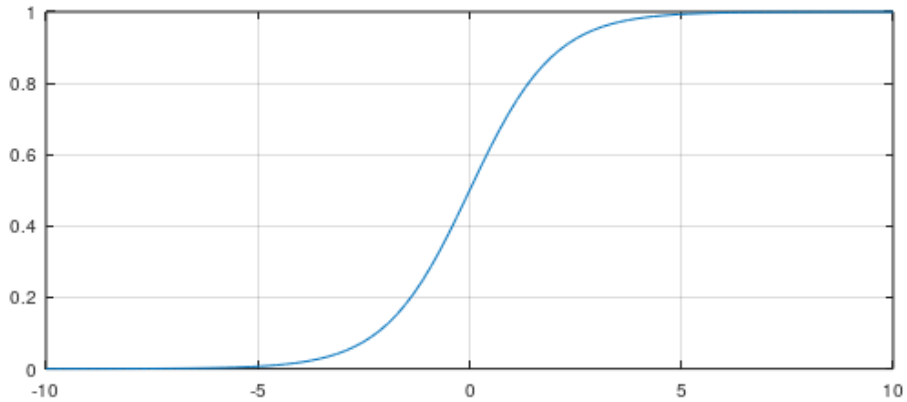


Рис. 2.8. Сигмоїдна активаційна функція

За методом найменших квадратів, за цільову функцію мінімізації помилки у штучній нейронній мережі прийнято:

$$E(w) = \frac{1}{2}\sum_{j,p}\left(y_{j,p}^{(n)} - d_{j,p}\right)^2, \tag{2.65}$$

де $y_{j,p}^{(n)}$ – вихідний стан штучного j-го нейрону n-го шару штучної мережі при поданні на вхід p-го образу; $d_{j,p}$ – ідеальний (цільовий) вихідний стан цього штучного нейрону.

Обчислення суми здійснюється по всіх штучних нейронах вихідного шару і всім його образам, які обробляються даною штучною нейронною

мережею. Мінімізація помилки у нейронній мережі ведеться методом градієнтного спуску, що має на увазі, коригування вагових коефіцієнтів:

$$\Delta w_{ij}^{(n)} = -\eta \frac{\delta E}{\delta w_{ij}}. \tag{2.66}$$

де w_{ij} – коефіцієнт ваги синаптичного зв'язку, з'єднуючого i-й нейрон $n-1$ шару з j-м нейроном того ж шару, η – швидкісний коефіцієнт навчання, $0 < \eta < 1$.

Як показано в [93]:

$$\frac{\delta E}{\delta w_{ij}} = \frac{\delta E}{\delta y_j} \frac{dy_j}{ds_j} \frac{\delta s_j}{\delta w_{ij}}, \tag{2.67}$$

де y_j – вихід j-го нейрона, s_j – зважена сума вихідних сигналів, аргумент функції збудження.

Оскільки терм $\frac{dy_j}{ds_j}$ є похідною цієї функції по її аргументу, звідси, похідна функції збудження має бути неперервною на всій осі абсцис. Через це функція одиничного скачка та інші активаційні функції з розривами не придатні до використання у таких штучних нейронних мережах. Як правило використовуються гладкі функції, такі як гіперболічний тангенс або сигмоїдальна функція. У випадку використання гіперболоїдного тангенса:

$$\frac{dy}{ds} = 1 - s^2. \tag{2.68}$$

Третій терм виразу (2.67) $\frac{\delta s_j}{\delta w_{ij}}$ рівний виходу штучного нейрона $y_i^{(n-1)}$-го шару.

Перший терм (2.67) може бути розкладено наступним чином [93]:

$$\frac{\delta E}{\delta y_j} = \sum_k \frac{\delta E}{\delta y_k} \frac{dy_k}{ds_k} \frac{\delta s_k}{\delta y_j} = \sum_k \frac{\delta E}{\delta y_k} \frac{dy_k}{ds_k} w_{jk}^{(n+1)}, \tag{2.69}$$

де обчислення суми по k здійснюється для всіх штучних нейронів $n+1$ шару. Ввівши змінну:

$$\delta_j^{(n)} = \frac{\delta E}{\delta y_j} \frac{dy_j}{ds_j}, \tag{2.70}$$

отримується рекурентна формула для обчислення значень $\delta_J^{(n)}$ шару n-х величин $\delta_k^{(n+1)}$ старшого $n+1$-го шару:

$$\delta_j^{(n)} = \left[\sum_k \delta_k^{(n+1)} w_{jk}^{(n+1)} \right] \frac{dy_j}{ds_j}. \tag{2.71}$$

Для вихідного шару:

$$\delta_l^{(n)} = \left(y_l^{(n)} - d_l \right) \frac{dy_l}{ds_l}. \tag{2.72}$$

Розгорнувши (2.66):

$$\Delta w_{ij}^{(n)} = -\eta \delta_j^{(n)} y_i^{(n-1)}. \tag{2.73}$$

Для того, щоб процес корекції ваг мав інерційність, яка згладжує різкі стрибки під час переміщення по поверхні цільової функції, (2.73) доповнюються значенням різниці ваг з попередньої ітерації:

$$\Delta w_{ij}^{(n)}(t) = -\eta\left(\mu\Delta w_{ij}^{(n)}(t-1) + (1-\mu)\delta_j^{(n)} y_i^{(n-1)}\right), \tag{2.74}$$

де μ – інерційний коефіцієнт; t –поточна ітерація

Отже, алгоритм за яким здійснюється навчання нейронної мережі за допомогою методу зворотного поширення помилки можна сформулювати наступним чином:

1. При поданні на вхід штучної нейронної мережі, підчас передачі сигналів від входів до виходів, значення цих сигналів обчислюються за наступною формулою:

$$s_j^{(n)} = \sum_{i=0}^{m} y_i^{(n-1)} w_{ij}^{(n)}, \tag{2.75}$$

де m – кількість нейронів в $n-1$ шарі з урахуванням нейрона зі сталим виходом, рівним $+1$, який задає зміщення; $y_i^{(n-1)} = x_{ij}^{(n)}$ – i-й вхід нейрона j n-го шару:

$$y_j^{(n)} = f\left(s_j^{(n)}\right), \tag{2.76}$$

де f – сигмоїдальна активаційна функція;

$$y_q^{(0)} = I_q, \tag{2.77}$$

де I_q – q-й компонент вхідного вектору образу.

2. Здійснюється розрахунок $\delta^{(n)}$ для вихідного шару за формулою (2.72), і також зміну ваг $\Delta w^{(n)}$ шару n по формулі (2.57) або (2.58).

3. Здійснюється розрахунок по формулах (2.71) і (2.73), відповідно, $\delta^{(n)}$ та $\Delta w^{(n)}$ для інших шарів, $n = N - 1, \dots, 1$.

4. Здійснюється коригування всіх ваг в нейронній мережі

$$w_{ij}^{(n)}(t) = w_{ij}^{(n)}(t-1) + \Delta w_{ij}^{(n)}(t). \tag{2.78}$$

5. Якщо штучна мережа здійснює класифікацію з достатньою точність навчання закінчується, інакше здійснюється перехід до кроку 1.

Для здійснення класифікації вхідного голосового сигналу вибрано повнозв'язну штучну нейронну мережу, з сигмоїдальною активаційною функцією нейронів з коефіцієнтом пологості рівним 1, з 44 нейронами на вхідному шарі, 19 на першому, 17 на другому, 15 на третьому та 15 на вихідному шарі (рис. 2.9).

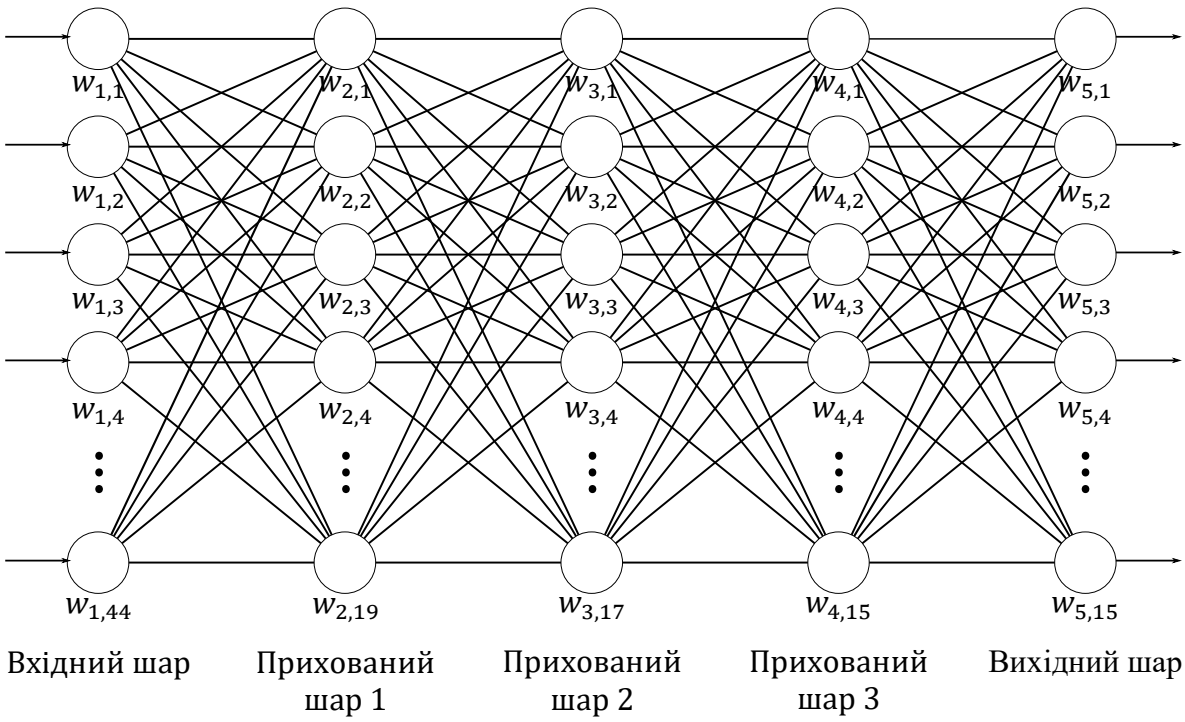


Рис. 2.9. Структура обраної штучної класифікаційної нейронної мережі

2.3. Апаратне забезпечення мікроконтролерної системи

Реалізації системи автоматичного розпізнавання мовлення здійснюється на базі мікроконтролера STM32F446.

2.3.1. Мікроконтролер STM32F446RET

Мікроконтролер STM32F446 базується на основі високопродуктивного Arm® Cortex®-M4 32-бітного RISC ядра, яке може оперувати на частотах до 180 МГц. Ядро Cortex-M4 містить блок обчислень з плаваючою комою з одинарною точністю та підтримує всі інструкції операцій з числами з плаваючою комою одинарної точності Arm®. Також ядро реалізує повний набір інструкцій цифрового опрацювання сигналів та блок захисту пам’яті, який підвищує захищеність додатків.

Мікроконтролери STM32F446 містять в собі високошвидкісні вбудовані пам’яті (FLASH пам’ять, розміром до 512 Кбайт та до 128 Кбайт SRAM пам’яті), до 4 Кбайт запасної SRAM пам’яті та широкий вибір покращених портів вводу-виводу та периферії, підключеної до двох APB шин, дві AHB шини та 32-бітну мульти-AHB матрицю.

Всі мікроконтролери містять три 12-бітні АЦП, два ЦАП, тактовий генератор реального часу з низьким споживанням, дванадцять 16-бітних ШІМ тактових генераторів для керування двигунами, два 32-бітних таймери загального призначення.

Також мікроконтролери STM32F446 включають набір стандартних та нестандартних комунікаційних інтерфейсів:

- До двох I2C інтерфейсів;
- Чотири SPI, три I2S у режимі full simplex для забезпечення якості аудіо, периферія I2S може тактуватись від окремого аудіо PLL або від зовнішнього тактового генератора для можливості синхронізації;
- Чотири USART порти та два UART;

- Порт USB OTG у режимі full-speed та порт USB OTG у режимі high-speed з можливістю вибору режиму full-speed, обидва з окремими лініями живлення;
- Два порти CAN;
- Два порти SAI, які можуть тактуватись від окремого аудіо PLL;
- Інтерфейс SDIO/MMC;
- Інтерфейс камери;
- Інтерфейс HDMI-CEC;
- SPDIF ресивер (SPDIFRx);
- Порт QuadSPI.

Нестандартні інтерфейси включають в себе SDIO, інтерфейс гнучкого керування пам’яттю, інтерфейс камери для CMOS сенсорів.

Мікроконтролери STM32F446 можуть оперувати при температурах від -40 до +105 градусів Цельсія та при живлені від джерел з напругою від 1.7 В до 3.6 В.

Живлення мікроконтролера може знижуватись до 1.7 В з використанням зовнішнього супервізора джерела живлення. Також є доступним широкий набір функції енергозбереження, що дозволяє проектувати пристрої на базі даного мікроконтролера з низьким енергоспоживанням.

Описані вище функції та особливості мікроконтролерів STM32F446 робить можливим їх застосування у широкому спектрі областей:

- контролери електричних двигунів;
- медичне обладнання;
- індустріальне обладнання: програмовані логічні контролери, інвертори, роз’єднувачі кіл;
- принтери та сканери
- системи тривоги, відео інтерком, та системи підігріву, вентилювання та кондиціонування;
- домашні аудіо пристрої.

2.3.2. Плата розробки на базі мікроконтролера STM32F446RET

Плати розробки «STM32 Nucleo-64» (рис 2.10) надає доступний та гнучкий спосіб побудови прототипів з мікроконтролерами STM32, варіюючи комбінацію обчислювальної потужності, вартості та кількості доступних функцій.

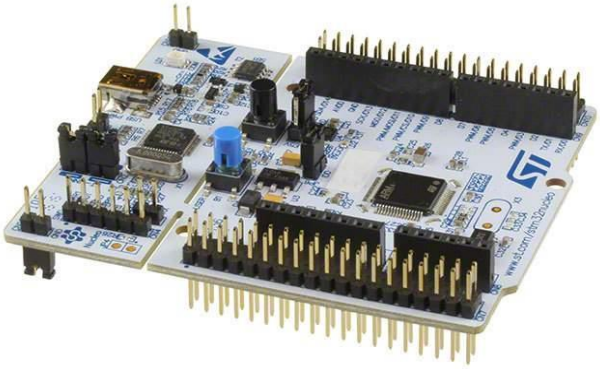


Рис. 2.10. Загальний вигляд плати розробки STM32F446 Nucleo

Плати розробки «STM32 Nucleo-64» надає змогу для використання:

- мікроконтролера STM32F446 в форматі LQFP64;
- трьох світлодіодів: – світлодіод USB (LD1), програмований світлодіод (LD2), світлодіод живлення LED (LD3)
- двох кнопок: USER та RESET
- двох типів плат розширення: Arduino™ Uno V3 сумісних та ST morpho сумісних;
- гнучкого живлення: через шину USB або від зовнішнього джерела з напругами: 3.3 В, 5 В або 7-12 В;
- вбудованого програматора та відладчика ST-LINK/V2-1 з конектором SWD з можливістю використання з зовнішніми пристроями;
- можливості змінювати тип USB пристрою на: віртуальний COM порт, пристрій збереження даних, порт відладки;

- широкого набору безкоштовних програм та бібліотек доступних з пакетом STM32Cube MCU;
- платформи Arm® Mbed.

2.4. Висновки до розділу 2

Процес розпізнавання мовлення здійснюється за допомогою двох блоків: front-end та back-end процесорів.

Front-end процесор займається зчитуванням звукового сигналу, його попередньою. фільтрацією, розбиттям на фрейми та накладанням вікна, виділення ознак за допомогою вейвлет-перетворення.

Back-end процесор, в свою чергу, займається порівнянням представлення звукового сигналу у вигляді вектора ознак, з базою шаблонів відомих представлень мовленнєвих сигналів, використовуючи повнозв’язну штучну нейронну мережу, з сигмоїдальною активаційна функція нейронів.

Апаратна реалізація системи розпізнавання мовлення здійснюється на базі плати розробки «STM32 Nucleo-64» на мікроконтролері STM32F446.

РОЗДІЛ 3
РЕАЛІЗАЦІЯ СИСТЕМИ РОЗПІЗНАВАННЯ МОВИ

Для аналізу ефективності функціонування запропонованої системи розпізнавання мовлення було здійснено реалізацію її основних частин у програмі Matlab. Зокрема, змодельовано процес попередньої обробки звукового сигналу та процедури виявлення ознак майбутньої системи, на основі яких здійснено процес навчання штучної нейронної мережі. Після чого отримано код програми для процедур виявлення ознак та навченої нейронної мережі, який були використані для створення програмного забезпечення мікроконтролерної системи на базі плати розробки «STM32 Nucleo-64».

3.1. Попередня обробка та виявлення ознак

Моделювання попередньої обробки вхідного сигналу здійснюється шляхом попереднім масштабуванням його частоти дискретизації, таким чином, щоб вона стала рівною 48000 Гц. Відмасштабований сигнал пропускається через створений засобами Matlab FIR-фільтр (рис. 3.1) з частотою зрізу у 6000 Гц, після чого проріджується у 8 раз (рис. 3.2).

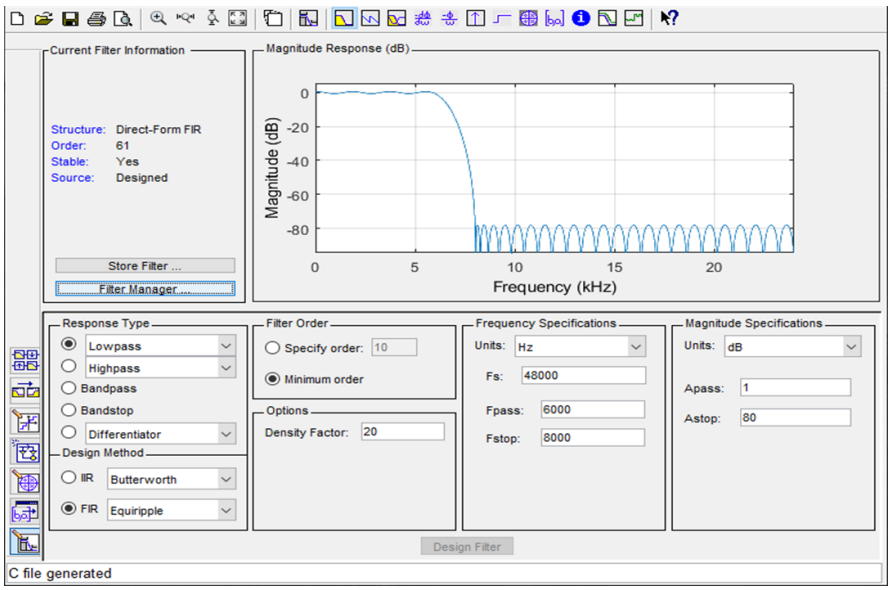
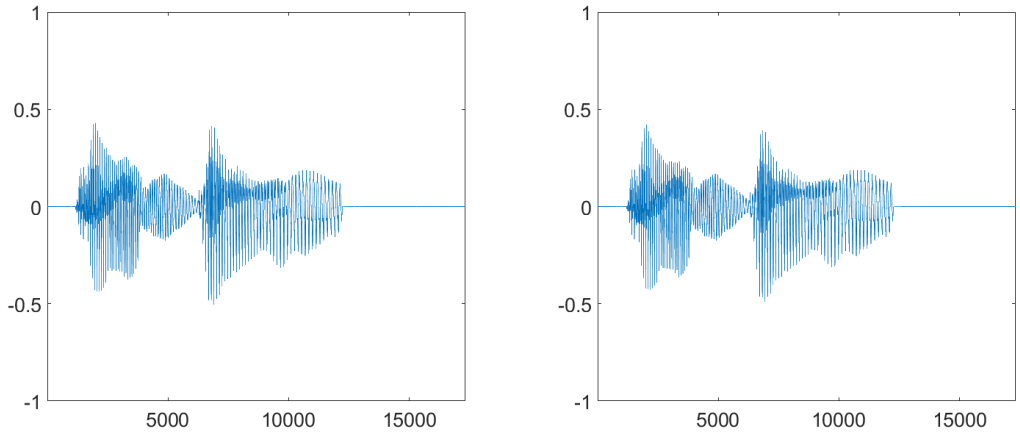


Рис. 3.1. Створення FIR-фільтра засобами Matlab

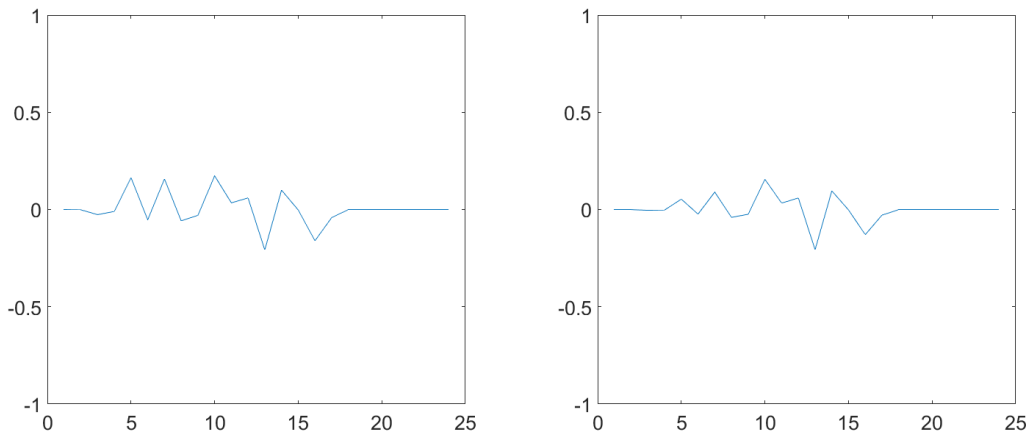


а) б)

Рис. 3.2. Графіки звукових сигналів слова «один».

а) – сигнал до фільтрації, б) – сигнал після фільтрації та проріджения

Таким чином отримується сигнал з частотою дискретизації у 6000 Гц. Здійснюється поділ сигналу на фрейми довжиною у 30 мс з перетином у 5 мс, на які накладається віконна функція (2.3) (рис. 3.3).



а) б)

Рис. 3.3. Накладання віконної функції на один фрейм сигналу.

а) – фрейм до накладання вікна, б) – фрейм після накладання вікна

Для кожного з фреймів обчислюється функція голосової активації за формулою (2.6). На рис. 3.4 зображено значення функції голосової активації для вхідного сигналу.

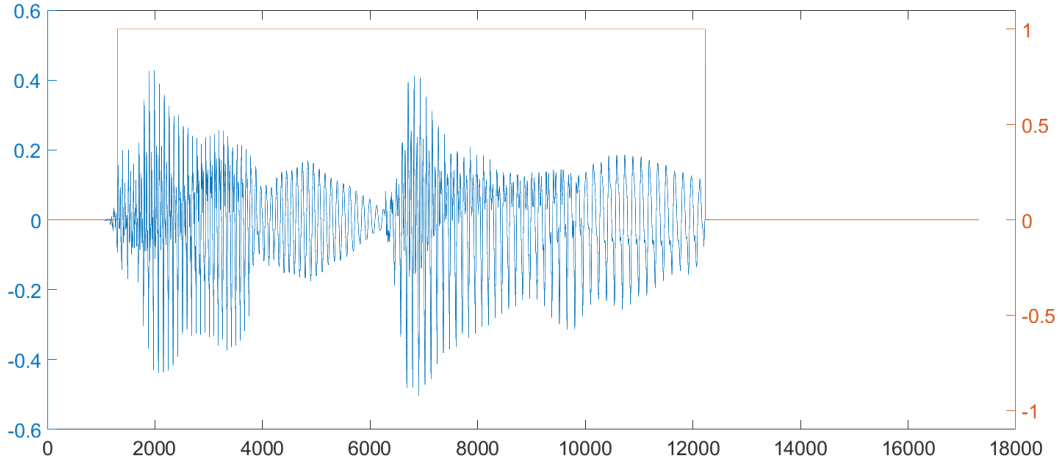
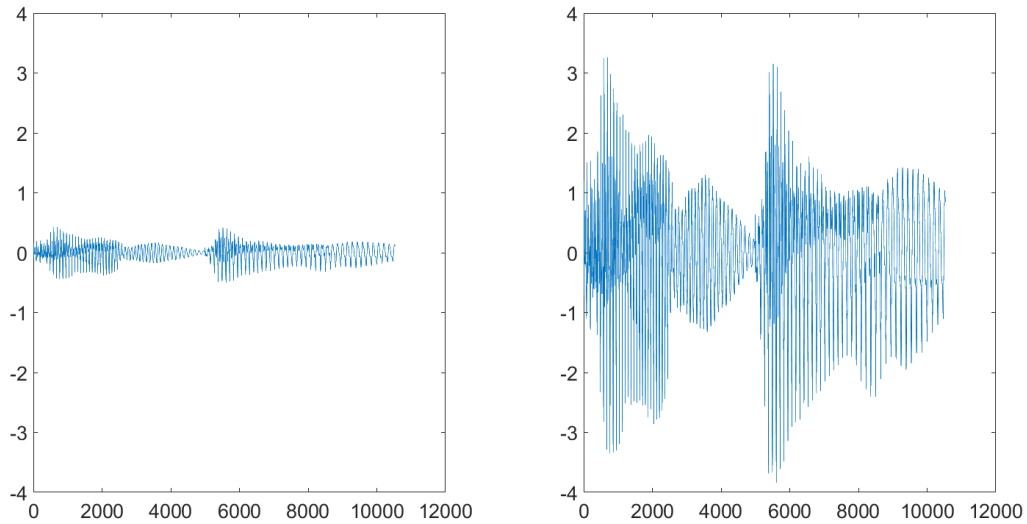


Рис. 3.4. Накладені графіки вхідного звукового сигналу слова «один» та його значення функції голосової активації.

Після визначення моментів початку та кінця мовлення здійснюється нормалізація фреймів сигналу, для яких значення функції голосової активації рівна одиниці (рис. 3.5).



а) б)

Рис. 3.5. Графіки сегментів звукових сигналів слова «один».

а) – сигнал до нормалізації, б) – сигнал після нормалізації

Над нормалізованим сигналом здійснюється вейвлет перетворення (рис. 3.6), використовуючи вейвлет Добеші четвертого порядку та 10 рівнів аналізу; здійснюється обчислення параметрів: норми, енергії, максимуму та мінімуму, для отримання вектору ознак довжиною в 44 елементи (рис. 3.7).

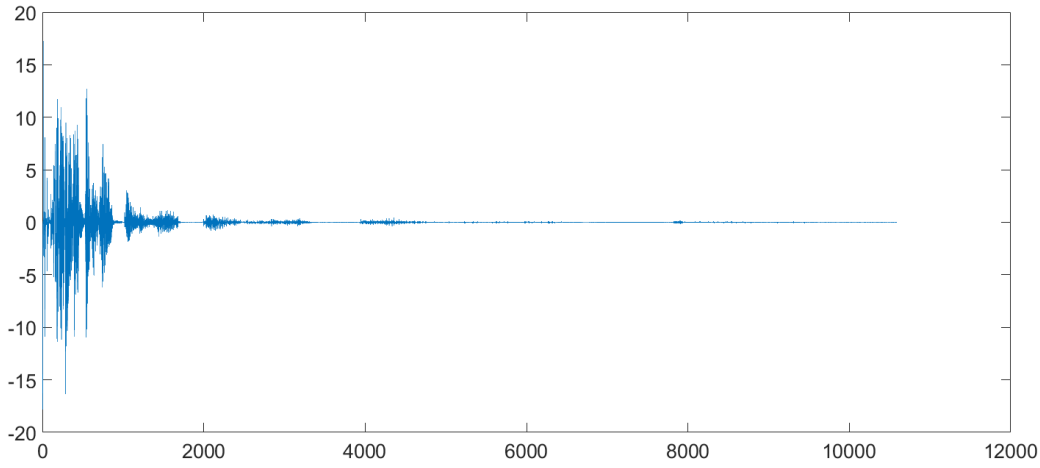


Рис. 3.6. Вейвлет-розклад звукового сигналу вимови слова «один»

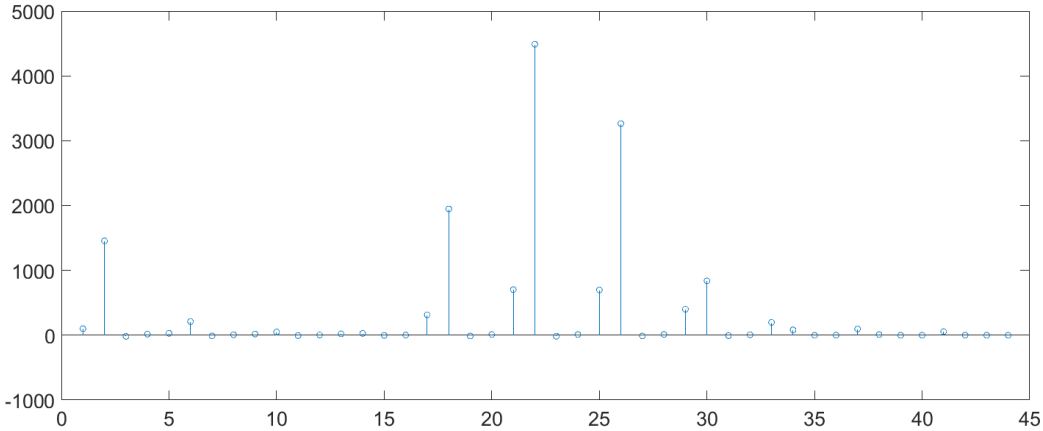


Рис. 3.7. Вектор ознак звукового сигналу вимови слова «один»

3.2. Навчання штучної нейронної мережі

Для навчання штучної нейронної мережі (рис. 3.8) використано базу записів 15 команд з 100 варіантами вимови кожної. Приклади сигналів вимов

наведені на рис. 3.4. Для навчання використано 80% від усіх варіантів кожної команди. Перевірка точності розпізнання здійснюється на всіх записах.

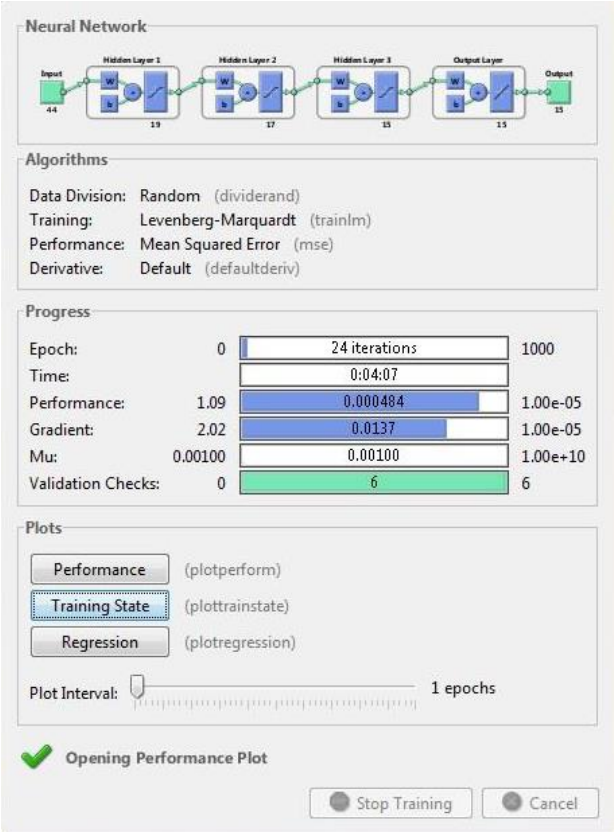


Рис. 3.8. Навчання штучної нейронної мережі

Після завершення навчання, отримано навчену нейронну мережу.

3.3. Результати

На основі проведеного моделювання, запрограмовано мікроконтролерну систему для реалізації запропонованого методу розпізнавання мовлення. Після чого, було проведено оцінку роботи даної системи за такими параметрами, як точність розпізнавання та швидкість розпізнавання. За точність розпізнавання прийнято відношення кількості записів у навчальній базі до кількості правильно розпізнаних слів. За швидкість розпізнавання прийнято час між завершенням вимови слова та моментом видачі системою результату розпізнавання.

Після чого отриману оцінку було порівняно з аналогічними системами автоматичного розпізнавання мовлення на базі мікроконтролерних систем.

Результати проведення оцінки роботи системи та порівняння оцінок наведено в таблицях 3.1 та 3.2 відповідно.

Таблиця 3.1

Результати класифікації системою розпізнавання мовлення

Слово	Кількість варіантів	Коректно класифіковано	Швидкість розпізнання, мс	Середня точність
Нуль	100	100	259,4	100%
Один	100	97	270,7	97%
Два	100	96	223,7	96%
Три	100	100	247,5	100%
Чотири	100	96	377	96%
П’ять	100	99	317,2	99%
Шість	100	99	349	99%
Сім	100	100	227,6	100%
Вісім	100	99	297,7	99%
Дев’ять	100	98	440,8	98%
Старт	100	96	302,7	96%
Стоп	100	99	280,2	99%
Вліво	100	98	327,6	98%
Вправо	100	98	372	98%
Назад	100	97	331,2	97%
Всього	1500	1472	308,3	98,1%

3.4. Висновки до розділу 3

За вибраними у попередньому розділі методами та засобами автоматичного розпізнавання мовлення, було здійснено моделювання процесу розпізнавання мовлення та здійснено його реалізацію на базі мікроконтролерної системи.

Процес моделювання здійснювався засобами програмного пакету Matlab. У процесі моделювання було створено моделі попередньої фільтрації сигналу, розбиття сигналу на фрейми, виявлення голосової активації та виявлення ознак, допомогою вейвлет-перетворення та класифікаційної штучної нейронної мережі.

Отримані моделі було реалізовано на мікроконтролерній системі на базі плати розробки «STM32 Nucleo-64». Після чого було здійснено оцінку роботи отриманої системи та здійснено порівняння отриманих результатів з аналогічними у схожих систем розпізнавання мовлення.

У результаті порівняння було виявлено, що запропонована система показує кращі результати точності розпізнавання та швидкості розпізнавання ніж аналогічні системи. Таким чином можна зробити висновок про доцільність використання запропонованої системи.

Таблиця 3.2

Результати порівняння систем розпізнавання мовлення

Система	Швидкість розпізнання, мс	Середня точність
Запропонована система	308,30	98,10%
Система на базі мікроконтролера [53]	341,30	96,94%
Система на базі мікроконтролера на базі ядра 8051 [54]	356,19	97,81%
Система на базі мікроконтролера [55]	458,79	96,82%

РОЗДІЛ 4
ОБҐРУНТУВАННЯ ЕКОНОМІЧНОЇ ЕФЕКТИВНОСТІ

Економічне обґрунтування дипломної роботи магістра є метою даного розділу. Даний розділ дозволяє встановити доцільність проведення науково–дослідних робіт і економічно обґрунтувати доцільність застосування тих чи інших засобів. Саме проведення економічних розрахунків, спрямованих на визначення економічної ефективності науково–дослідницької роботи (НДР) і прийняття рішення про її подальший розвиток та впровадження або ж недоцільність проведення відповідної розробки.

Метою дипломної роботи магістра є дослідження методів та засобів автоматичного розпізнавання мови на базі мікроконтролерної системи.

В економічній частині дипломного проекту будуть проведені такі етапи розрахунку вартості НДР:

- описати технологічний процес розробки із зазначенням трудомісткості кожної операції;
- визначити суму витрат на оплату праці основного і допоміжного персоналу, включаючи відрахування на соціальні заходи;
- визначити суму матеріальних затрат;
- обчислити витрати на електроенергію для науково– виробничих цілей;
- нарахувати суму амортизаційних відрахувань;
- визначити суму накладних витрат;
- скласти кошторис та визначити собівартість НДР;
- розрахувати ціну НДР; - визначити економічну ефективність та термін окупності продукту.

На основі отриманих розрахунків будуть розроблені техніко-економічні показники проектованого виробництва.

Як відомо, розробка надійної і ефективної інформаційної системи вимагає значних затрат часу. Слід зауважити, що затрати часу залежать від кваліфікації розробника і його можливостей. Розробник повинен у достатній мірі володіти

навиками програмування, вміти адекватно застосовувати математичний апарат, бути добре обізнаним з об'єктом дослідження.

4.1. Визначення стадій технологічного процесу та загальної тривалості проведення НДР

Для оцінки тривалості виконання окремих робіт використовують нормативи часу або попередній досвід. До таких нормативів відносять тривалість написання операцій (команд), які в деяких підприємствах становлять: для одної операції від 30 хвилин до 1,6 годин та 8 годин для п'яти операцій (тривалість зміни).

У разі їх відсутності звертаються до експертних оцінок по встановленню тривалості кожного етапу (стадії):

при трьох оцінках:

$$T_{\text{вс}} = \frac{(t_{min} + 4t_{\text{н.й}} + t_{max})}{6}, \tag{4.1}$$

при двох оцінках:

$$T_{\text{вс}} = \frac{(3t_{min} + 2t_{max})}{5}, \tag{4.2}$$

де $T_{\text{вс}}$ – очікуване (середнє) значення тривалості виконання етапу (стадії); t_{min}, $t_{\text{н.й}}$, t_{max} – відповідно мінімальна, найбільш імовірна і максимальна оцінки тривалості виконання етапу (стадії).

Розробку даної інформаційної системи можна поділити на такі етапи:

- постановка задачі;
- проведення огляд публікацій авторів, які займались питанням дослідження методів та засобів автоматичного розпізнавання мови на базі мікроконтролерних систем;

- прийняття рішень щодо вибору оптимального шляху розв'язання поставленої задачі;
- аналіз математичної моделі інформаційної системи;
- обґрунтування використання обраних методів та засобів автоматичного розпізнавання мови на базі мікроконтролерних системи;
- розробка архітектури та алгоритмічного забезпечення системи автоматичного розпізнавання мови з використанням запропонованих методів та моделі;
- розробка макету системи автоматичного розпізнавання мови;
- написання програмного забезпечення для мікроконтролерної системи;
- налаштування середовища розробки і роботи вже готової програми;
- тестування та оцінка завадостійкості системи;
- написання і оформлення документації (електронної та паперової).

Для зручного представлення і визначення загальної тривалості проведення НДР доцільно дані витрат часу по окремих операціях технологічного процесу звести у таблицю 4.1.

Витрати часу наукового керівника на виконання окремих стадій (етапів) при недостатній кількості інформації доцільно приймати в межах 5% сумарних витрат часу інженерів на виконання цих стадій (етапів).

Таблиця 4.1

Основні етапи і час їх виконання у НДР

№ з/п	Етап	Середній час виконання етапу, год	
		інженер	керівник
1	Постановка задачі	2	1
2	Проведення огляд публікацій авторів, які займались питанням дослідження методів та засобів автоматичного розпізнавання мови на базі мікроконтролерних систем	20	8
3	Прийняття рішень щодо вибору оптимального шляху розв'язання поставленої задачі	5	3

Продовження таблиці 4.1

№ з/п	Етап	Середній час виконання етапу, год	
		інженер	керівник
4	Аналіз математичної моделі інформаційної системи	1	1
5	Обґрунтування використання обраних методів та засобів автоматичного розпізнавання мови на базі мікроконтролерних системи	7	2
6	Розробка архітектури та алгоритмічного забезпечення системи автоматичного розпізнавання мови з використанням запропонованих методів та моделі	45	5
7	Розробка макету системи автоматичного розпізнавання мови	6	2
8	Написання програмного забезпечення для мікроконтролерної системи	25	2
9	Налаштування середовища розробки і роботи вже готової програми	1	1
10	Тестування та оцінка завадостійкості системи	4	2
11	Написання і оформлення документації (електронної та паперової)	84	3
разом		200	30

Отже, сумарний час виконання операцій технологічного процесу інженером становить 200 годин, а керівником 30 годин [94].

4.2. Визначення витрат на оплату праці та відрахувань на соціальні заходи

Заробітна плата працівника незалежно від виду підприємства визначається його особистим трудовим вкладом, залежить від кінцевих результатів роботи підприємства, регулюється податками і максимальними розмірами не обмежується. Розміри, порядок нарахування і виплати заробітної плати регулюються чинним законодавством України, відповідними указами і постановами, галузевими інструкціями. Розмір заробітної плати залежить від

складності та умов виконуваної роботи, професійно-ділових якостей працівника, результатів його праці та господарської діяльності підприємства. Заробітна плата складається з основної та додаткової оплати праці.

Основна заробітна плата нараховується на виконану роботу за тарифними ставками, відрядними розцінками чи посадовими окладами і не залежить від результатів господарської діяльності підприємства.

Додаткова заробітна плата – це складова заробітної плати працівників, до якої включають витрати на оплату праці, не пов'язані з виплатами за фактично відпрацьований час. Нараховують додаткову заробітну плату залежно від досягнутих і запланованих показників, умов виробництва, кваліфікації виконавців. Джерелом додаткової оплати праці є фонд матеріального стимулювання, який створюється за рахунок прибутку.

Основна заробітна плата складається із прямої заробітної плати та доплати, яка при укрупнених розрахунках становить 25% – 35% від прямої заробітної плати. При розрахунку заробітної плати кількість робочих днів в місяці слід приймати – 25,4 дні/міс., що відповідає 203,2 год./міс. Розмір місячних окладів керівника та інженерів слід приймати згідно існуючих на даний час норм. Основна заробітна плата розраховується за формулою:

$$З_{осн} = T_С \cdot K_Г, \tag{4.3}$$

де $Т_с$ – тарифна ставка, грн., $К_г$ - кількість відпрацьованих годин.

Посадові оклади (тарифні ставки) за розрядами Єдиної тарифної сітки визначаються шляхом множення окладу (ставки) працівника 1 тарифного розряду на відповідний тарифний коефіцієнт. У разі коли посадовий оклад (тарифна ставка) визначені у гривнях з копійками, цифри до 0,5 відкидаються, від 0,5 і вище – заокруглюються до однієї гривні.

Законом України "Про Державний бюджет України на 2019 рік" від 23.11.2018 р. №2629 – VIII із змінами, внесеними згідно із Законом № 2696-VIII від 28.02.2019, ВВР, 2019, № 14, ст.66 та № 149-IX від 02.10.2019,

установлено у 2019 році мінімальну заробітну плату: у місячному розмірі: з 1 січня - 4173 гривні; у погодинному розмірі: з 1 січня - 25,13 гривні. Приймемо 65 грн. для інженера, для керівника — 81 грн.

Тарифні ставки: керівник проекту – 81 грн./год., інженер – 65 грн./год.

Тоді скориставшись формулою 4.3 розрахуємо основну заробітну плату для інженера та керівника проекту.

Керівник проекту:

$$З_{осн} = 81 \cdot 30 = 2430 \text{ грн.}$$

Інженер:

$$З_{осн} = 65 \cdot 200 = 13000 \text{ грн.}$$

Додаткова заробітна плата становить 10–15% від суми основної заробітної плати:

$$З_{дод} = З_{осн} \cdot K_{допл}, \tag{4.4}$$

де $K_{допл}$ – коефіцієнт додаткових виплат працівникам 0,1.

Керівник проекту:

$$З_{осн} = 2430 \cdot 0{,}1 = 243 \text{ грн.}$$

Інженер:

$$З_{осн} = 13000 \cdot 0{,}1 = 1300 \text{ грн.}$$

Звідси загальні витрати на оплату праці ($В_{ОП}$) визначаються за формулою, і становлять:

$$В_{ОП} = З_{осн} + З_{дод}, \quad (4.5)$$

Керівник проекту:

$$В_{оп} = 2430 + 243 = 2673 \text{ грн.}$$

Інженер:

$$В_{оп} = 13000 + 1300 = 14300 \text{ грн.}$$

Таким чином загальна сума становить 16973 грн.

Крім того, слід визначити відрахування на соціальні заходи:

- податок на доходи фізичних осіб: 18%;
- військовий збір 1,5%;
- єдиний соціальний внесок 22%.

У сумі зазначені відрахування становлять 41,5 %.

Отже, сума відрахувань на соціальні заходи розраховуємо за формулою:

$$В_{с.з} = ФОП \cdot 0{,}415, \quad (4.6)$$

де ФОП – фонд оплати праці в гривнях.

Тоді, сума відрахувань на соціальні заходи буде становити:

$$В_{с.з} = 16973 \cdot 0{,}415 = 7043{,}80 \text{ грн.}$$

Проведені розрахунки витрат на оплату праці зведено у таблицю 4.2.

Таблиця 4.2

Зведені розрахунки витрат на оплату праці

№ п/п	Категорія працівників	Основна заробітна плата, грн.			Додаткова заробітна плата, грн.	Нарах. на ФОП, грн.	Всього витрати на оплату праці, грн.
		Тариф на ставка, грн.	К-сть відпрацьов. год.	Фактичн о нарах. з/пл., грн.			
1	2	3	4	5	6	7	8
А	Б	1	2	3	4	5	6
1	Керівник проекту	81	30	2430	243	1109,295	3782,295
2	Інженер	65	200	13000	1300	5934,5	20234,5
Разом				15430	1543	7043,795	24016,795

4.3. Розрахунок витрат на електроенергію

Затрати на електроенергію 1-ці обладнання визначаються за формулою:

$$З_E = W \cdot T \cdot S, \tag{4.7}$$

де W – необхідна потужність, кВт; T – кількість годин роботи обладнання; S – вартість кіловат-години електроенергії.

Згідно з постановою НКРЕ України від 05.10.2018 року № 1177 вартість електроенергії становить 308,25 коп./кВт·год.

Середня потужність комп’ютера – 242 Вт з підключеним маршрутизатором і комутатором, кількість годин роботи обладнання згідно таблиці 4.1 – 230 год.

$$З_E = 0{,}242 \cdot 223 \cdot 3{,}0825 = 166{,}36 \text{ грн.}$$

4.4. Розрахунок витрат на матеріали

Результати розрахунку затрат на матеріали зводяться в таблицю 4.3.

Таблиця 4.3

Визначення величини затрат на матеріали

Найменування матеріальних ресурсів	Одиниця виміру	Норма витрат	Ціна за одиницю грн	Затрати матеріалі в грн	Транс-портно-заготі-вельні витрати, грн	Загальна сума витрат на матеріали, грн
Папір А4	Пачка	1	90	90	-	90
Ватман	Штук	10	11	110	-	110
STM32F446 Nucleo	Штук	1	610	610	40	650
Провідники	Пачка	1	40	40	-	40
Разом						890

4.5. Розрахунок суми амортизаційних відрахувань

Характерною особливістю застосування основних фондів у процесі виробництва є їх відновлення. Для відновлення засобів праці у натуральному виразі необхідне їх відшкодування у вартісній формі, яке здійснюється шляхом амортизації.

Амортизація – це процес перенесення вартості основних фондів на вартість новоствореної продукції з метою їх повного відновлення. Для заміщення зношеної частини основних засобів виробництва підприємства роблять амортизаційні відрахування, тобто відрахування певних грошових сум відповідно до розмірів фізичного і морального зносу засобів виробництва.

Комп’ютери та оргтехніка належать до четвертої групи основних фондів. Для цієї групи річна норма амортизації дорівнює 60 % (квартальна – 15 %).

Для визначення амортизаційних відрахувань застосовуємо формулу:

$$А = \frac{Б_в \cdot Н_а}{100}, \quad (4.8)$$

де А – амортизаційні відрахування за звітний період, грн., $Б_в$ – балансова вартість комп'ютера, на початок звітного періоду, грн., $Н_а$ – норма амортизації, %.

Для роботи використовується один ноутбук (вартість якого становить 10000 грн.), який працює 230 години.

$$A = \frac{10000 \cdot 15\%}{100\%} = 1500 \text{ грн.}$$

4.6. Обчислення накладних витрат

Накладні витрати пов'язані з обслуговуванням виробництва, утриманням апарату управління підприємства (фірми) та створення необхідних умов праці.

Накладні витрати можуть становити 20% від суми основної та додаткової заробітної плати працівників:

$$Н_в = В_{о.п} \cdot 0{,}2, \quad (4.9)$$

$$Н_в = 16973 \cdot 0{,}2 = 3394{,}6 \text{ грн.}$$

4.7. Складання кошторису витрат та визначення собівартості НДР

Результати проведених вище розрахунків зведемо у таблицю 4.4. Собівартість ($С_в$) НДР розрахуємо за формулою:

$$С_в = В_{о.п} + В_{с.з} + З_{м.в} + З_е + T_в + A + Н_в, \quad (4.10)$$

$$С_в = 16973 + 7043{,}80 + 890 + 166{,}36 + 1500 + 3394{,}6 = 29967{,}76 \text{ грн.}$$

Таблиця 4.4

Кошторис витрат на НДР

Зміст витрат	Сума, грн.	У % до загальної суми
1	2	3
Витрати на оплату праці (основну і додаткову заробітну плату)	16973	56,6%
Відрахування на соціальні заходи	7043,80	23,5%
Матеріальні витрати	890	3%
Витрати на електроенергію	166,36	0,6%
Амортизаційні відрахування	1500	5%
Накладні витрати	3395,6	11,3%
Собівартість	29967,76	100%

4.8. Розрахунок ціни НДР

Ціну НДР можна визначити за формулою:

$$Ц = \frac{C_в \cdot (1 + P_{рен}) + K \cdot В_{н.і}}{K} \cdot (1 + ПДВ), \qquad (4.11)$$

де $Р_{рен}$ – рівень рентабельності, 30%; К – кількість замовлень; $В_{н.і}$ – вартість носія інформації, грн., ПДВ – ставка податку на додану вартість (20%)

Оскільки розробка є прикладною, і використовуватиметься тільки для одного підприємства, то для розрахунку ціни не потрібно вказувати коефіцієнти К та $В_{і.н}$, оскільки їх в даному випадку не потрібно.

Тоді, формула для обчислення ціни розробки буде мати вигляд:

$$Ц = C_в \cdot (1 + P_{рен}) \cdot (1 + ПДВ), \quad (4.12)$$

Звідси ціна на проект складе:

$$Ц = 29967{,}76 \cdot (1 + 0{,}3) \cdot (1 + 0{,}2) = 46749{,}71 \text{ грн.}$$

Таким чином ціна рівна 46749,71 грн.

Визначимо величину прибутку за формулою

$$П = Ц - С_в, \quad (4.13)$$

Звідси величину прибутку складе:

$$П = 46749{,}71 - 29967{,}76 = 16781{,}95 \text{ грн.}$$

4.9. Визначення економічної ефективності і терміну окупності капітальних вкладень

Ефективність виробництва – це узагальнене і повне відображення кінцевих результатів використання робочої сили, засобів та предметів праці на підприємстві за певний проміжок часу. Економічна ефективність ($Е_р$) полягає у відношенні результату виробництва до затрачених ресурсів:

$$Е_р = \frac{П}{С_в}, \quad (4.14)$$

де П – прибуток; $С_в$ – собівартість.

$$E_р = \frac{16781,95}{29967,76} = 0,56.$$

Поряд із економічною ефективністю розраховують термін окупності капітальних вкладень ($T_р$):

$$T_р = \frac{1}{E_р}, \quad (4.15)$$

$$T_р = \frac{1}{0,59} = 1,69 \text{ роки.}$$

Про доцільність розробки програми можна сказати при врахуванні критеріїв, які наведено у таблиці 4.5

Таблиця 4.5

Техніко-економічні показники НДР

№ п/п	Показник	Значення
1	Собівартість, грн.	29967,76
2	Плановий прибуток, грн.	16781,95
3	Ціна, грн.	46749,71
4	Економічна ефективність	0,56
5	Термін окупності, рік	1,79

4.10. Висновки до розділу 4

У результаті проведення розрахунків можна зробити висновок: розробка матиме оптимальну економічну ефективність 0,56 і термін окупності становитиме не більше двох років (1,79 роки). Варто зазначити, що дані розрахунки носять номінальний характер і основна їх мета оцінити приблизну вартість дослідження та створення даного продукту. Номінальний характер розрахунків зумовлений тим, що даний програмний продукт має дослідницьке призначення.

РОЗДІЛ 5
ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

5.1. Охорона праці

У роботі досліджуються методи та засоби автоматичного розпізнавання мови на базі мікроконтролерної системи. Оскільки дане дослідження виконується з використанням електронно-обчислювальних машин необхідним є дотримання відповідних правил охорони праці.

Основним законодавчим актом з охороні праці при роботі з персональними комп'ютерами в Україні є НПАОП 0.00-7.15-18 «Вимоги щодо безпеки та захисту здоров'я працівників під час роботи з екранними пристроями». Даний акт призначений для запобігання впливу на працівників шкідливих і небезпечних факторів, пов'язаних із зоровою й нервово-емоційною напругою, вимушеною сталістю робочої пози при локальній напрузі рук на фоні обмеженої загальної м'язової активності (гіподинамії) під впливом комплексу фізичних факторів: шуму, електростатичного поля, електромагнітних випромінювань, що не іонізують і іонізують повітря, а також електричної напруги.

Згідно НПАОП 0.00-7.15-18 облаштування робочих місць, обладнаних екранними пристроями, повинно забезпечувати:

- належні умови освітлення приміщення і робочого місця, відсутність відблисків;
- оптимальні параметри мікроклімату;
- належні ергономічні характеристики основних елементів робочого місця, а також враховувати небезпечні і шкідливі фактори.

Матеріали для оздоблення приміщень з ЕОМ повинні відповідати вимогам до них органів державного санітарно-епідеміологічного нагляду.

Забороняється застосовувати для оздоблення полімерні матеріали: деревинно-стружкові плити, шпалери, що миються, рулонні синтетичні

матеріали, шаруватий паперовий пластик тощо, що виділяють у повітря шкідливі хімічні речовини.

Для внутрішнього оздоблення приміщень з ПК мають застосовуватися дифузно-відбивні матеріали з коефіцієнтом відбиття:

- для стелі – 0,7-0,8;
- для стін – 0,5-0,6.

Покриття підлоги повинно бути матовим з коефіцієнтом відбиття 0,3-0,5, рівним, неслизьким, з антистатичними властивостями.

У цих приміщеннях повинно бути:

- опалення;
- система кондиціонування повітря або припливно-витяжна вентиляція.

У приміщеннях з ВДТ має здійснюватися щоденне вологе прибирання.

Приміщення для роботи з ПЕОМ повинні бути обладнані системами опалення, кондиціювання повітря або припливно-витяжною вентиляцією.

Приміщення з ПК повинні мати природне і штучне освітлення, яке відповідало б вимогам ДБН В.2.5-28:2018 «Природне і штучне освітлення», ДСАНПІН 3.3.2.007-98 «Гігієнічні вимоги до організації роботи з візуальними дисплейними терміналами електронно-обчислювальних машин».

Приміщення для роботи з ПЕОМ повинні мати природне й штучне освітлення. Віконні прорізи повинні бути орієнтовані на північ або на північний схід, забезпечувати коефіцієнт природної освітленості (К.П.О.) не менш 1,5% і мати жалюзі або штори.

Віконні прорізи повинні мати регульовані пристрої для відкривання, а також жалюзі, завіски, зовнішні козирки тощо.

Приміщення з ПЕОМ повинні бути обладнані системою загального рівномірного освітлення.

Штучне освітлення має здійснюватись системою загального рівномірного освітлення, яка включає суцільні або такі, що перериваються лінії світильників, розташованих збоку робочих місць (переважно ліворуч), паралельно лінії зору користувачів ПК. Світильники повинні мати розсіювачі

світла та екрануючі сітки. У світильниках місцевого освітлення можна використовувати лампи накалювання.

Вимоги до освітлення приміщень та робочих місць під час роботи з ВДТ:

- освітленість на робочому місці повинна відповідати характеру зорової роботи, який визначається трьома параметрами: об'єктом розрізнення, найменшим розміром об'єкта, що розглядається на моніторі ПК, фоном, який характеризується коефіцієнтом відбиття; контрастом об'єкта і фону;
- необхідно забезпечити достатньо рівномірне розподілення яскравості на робочій поверхні монітора, а також в межах навколишнього простору;
- на робочій поверхні повинні бути відсутні різкі тіні;
- в полі зору не повинно бути відблисків (підвищеної яскравості поверхонь, які світяться та викликають осліплення);
- величина освітленості повинна бути постійною під час роботи;
- слід обирати оптимальну спрямованість світлового потоку і необхідний склад світла.

Заходи з електробезпеки в електроустановках напругою до 1000 В будівель і споруд (житлових, адміністративно-побутових, суспільних, цехових і т.п.) регламентуються державними будівельними нормами ДСТУ Б В.2.5-82:2016 «Електробезпека в будівлях і спорудах. Вимоги до захисних заходів від ураження електричним струмом». В цих будівельних нормах приведені вимоги до виконання заходів захисту людей і тварин від прямого дотику, а також у випадках дотику до відкритих струмопровідних частин електроустановок за наявності пошкодження в цих електроустановках (захист при непрямому дотику). На особливу увагу заслуговують заходи дотримання протипожежної безпеки в приміщеннях, у яких знаходяться ЕОМ. Так, у всьому офісі лінії електромережі повинні бути забезпечені від виникнення короткого замикання, а також від перепадів напруги, що може викликати збоїв в роботі електронно–обчислювальної техніки. Приміщення повинні бути оснащені системою автоматичної пожежної сигналізації та вогнегасниками. Під час монтажу та експлуатації ліній електромережі необхідно повністю

виключити можливість виникнення електричного джерела загоряння внаслідок короткого замикання та перевантаження проводів, не допускати застосування проводів з легкозаймистою ізоляцією і, за можливості, застосовувати негорючу ізоляцію.

Під час виконання роботи, був проведений аналіз діючих законодавчих актів та норм щодо облаштування робочих приміщень. Проведений опис параметрів, особливостей розміщення джерел світла та заходів протипожежної безпеки у приміщеннях, де проводиться робота над проектуванням для дослідження методів та засобів автоматичного розпізнавання мови на базі мікроконтролерної системи.

5.2. Підтримання безпеки життєдіяльності в системі "людина - життєве середовище"

У найзагальнішому розуміння термін «людина» вказує на належність до людського роду – вищої сходинки живої природи на нашій планеті. Вона є найрозумнішою істотою на Землі і тому несе відповідальність за свої вчинки та дії. Людина як частина природи є біологічним суб‘єктом. За своєю тілесною будовою й фізіологічними функціями людина належить до тваринного світу. Але людина –вища сходинка розвитку живої природи на нашій планеті і вона якісно відрізняється від тварин. «Homo sapiens» – «людина розумна». Людина як біологічний вид має:

- характерні тілесні ознаки (прямоходіння, руки пристосовані до праці тощо);
- високорозвинений мозок, здатний творити;
- свідомість як здатність до пізнання сутності зовнішнього світу, так і своєї особистої природи;
- мислення та мову, які з‘явились в результаті трудової суспільної діяльності.

Найхарактернішою ознакою людини є свідомість, а саме, з погляду здатності, розмірковувати над зовнішніми обставинами, над своїми зв‘язками з ними й із іншими людьми, усвідомлювати сенс власного буття у світі.

Якщо тварина живе в природі, то людина –в соціумі.

Соціум – це особливий спосіб життя людей.

Таким чином, ми підходимо до людини з трьома різними вимірюваннями її суті: біологічним, психічним і соціальним.

Під психічним вимірюванням уявляється внутрішній духовний світ людини –її волю, переживання, пам'ять, характер, темперамент тощо.

Соціальне і біологічне вимірювання існують у нерозривній єдності. Біологічне, природне можна спрощено назвати системою «що живе», а соціальне –«як живе». Але і «що живе» і «як живе» злилися в єдине ціле, в соціальну істоту на ім‘я Людина.

Отже людина – це особистість.

Особистість – це міра цілісності людини. Головною властивістю особистості є світогляд. Особливим компонентом особистості є її моральність.

5.2.1. Діяльність людини

Головною відмінністю людини від тварини є спосіб життя.

Тваринне життя здійснюється природним чином, тобто як існування, людське –суспільним, соціальним, як життєдіяльність.

Діяльність – це специфічний людський спосіб ставлення людини досвіту. Діяльність постає як спосіб перетворення природи на предмети споживання, творіння культури.

Діяльність – це активна взаємодія людини з навколишнім середовищем, завдяки чому вона досягає свідомо поставленої мети, яка виникла внаслідок потреби.

Потреба – це необхідність для людини того, що забезпечує її існування і самозабезпечення.

Потреби поділяються на групи: фізіологічні, соціальні, престижні, особистісні, духовні.

Однією із форм діяльності людини є праця.

Праця – це цілеспрямована діяльність людини, в процесі якої вона впливає на природу і використовує її з метою виробництва матеріальних благ, необхідних для задоволення своїх потреб.

Коли ми говоримо про життя, то необхідно розглянути і протилежне йому поняття –смерть. Смерть –це кінець біологічного існування живої істоти, припинення її життєдіяльності.

Розуміння сутності смерті, усвідомлення, що людське особисте життя є єдиним і має кінець, сприяє осмисленню цінності життя, дає можливість зрозуміти ціль життя і пізнати призначення людини.

5.2.2. Середовище життєдіяльності людини

Життєве середовище -другий елемент системи «людина —життєве середовище», її об‘єкт. Життєве середовище є частиною Всесвіту, де перебуває або може перебувати в даний час людина і функціонують системи її життєзабезпечення. У цьому розумінні воно не має постійних у часі і просторі кордонів, його межі визначаються передусім рівнем системи, тобто тим, що в даному разі розуміється під терміном «людина».

Ми вже знаємо, що життєве середовище людини складається із трьох компонентів: природного, соціального, техногенного, тому дамо характеристику кожного середовища окремо.

природне середовище (земний ґрунт, повітря, водоймища, рослини, тварини, сонце, місяць, планети тощо);

соціальне, соціально-політичне середовище (форми спільної діяльності людей, спосіб життя, взаємостосунки тощо);

техногенне середовище (житло, транспорт, знаряддя праці, промислові та енергетичні об‘єкти, зброя, домашні і свійські тварини, сільськогосподарські рослини тощо).

5.2.3. Природне середовище

Життєдіяльність людини неможлива без навколишнього природного середовища. Поняття «навколишнє середовище» включає все те, що оточує та

впливає на життя людини в процесі її діяльності. В більш широкому розумінні –це космічний простір, а в більш вузькому – біосфера (біоз – життя, сфера – діяльність). Біосфера – зовнішня оболонка нашої планети Земля.

5.2.4. Вплив діяльності людини на навколишнє середовище

Існування людства завжди базувалося на безперервній взаємодії, постійному обміні речовин та енергії з навколишнім середовищем.

У процесі життя і виробничої діяльності людина неминуче вносить в середовище, що її оточує, певні зміни. Вони стосуються як хімічного і біологічного складу середовища, так і його фізичних характеристик. Промислова, технічна та наступна науково-технічна революція настільки озброїла людину технікою і такою мірою підвищили її могутність, що вплив людини на середовище її проживання став незрівняним з тими природними процесами, які відбувались на Землі тисячоліттями. За цих умов природний стан навколишнього середовища виявився суттєво порушеним, і звичайними шляхами –з повітрям, питною водою та харчами –в організм людини почали надходити токсичні речовини, які утворюються під час спалення органічного палива –вугілля, торфу, нафти, газу, деревини, внаслідок виплавки чорних та кольорових металів, промислового синтезу в хімічному виробництві та інші.

Видобуваючи щорічно понад 10 млрд. тонн гірських порід з земних надр, ми порушуємо земну поверхню, послаблюємо міцність верхньої частини земної кори й до невпізнання змінюємо її вигляд кар‘єрами, горами відвальних порід, шлаконакопичувачами, полями зрошення, звалищами.

Виплавляючи понад 800 млн. тонн різних металів, людство викидає в повітря та гідросферу величезну кількість промислового бруду, різних відходів.

Величезної шкоди завдає всій живій природі хімізація с.г. За останні роки в грунті щорічно вноситься 300 млн. тонн мінеральних добрів і близько 4 млн. тонн пестицидів, а врожайність неухильно знижується, грунти виснажуються, на багато десятиріч втрачають родючість.

Крім того, отрутохімікатами в світі щорічно отруюються понад 2 млн. чоловік, гинуть комахи, гризуни, гриби, вони мутують.

Щорічно в водойми скидається близько 600 млрд. тонн різних промислових стоків.

Величезної шкоди природі, останнім часом, завдає автомобільний транспорт, шкідливі продукти вихлопних газів якого в більшості великих міст планети становить від 45 до 80% загального забруднення атмосфери.

Близько 4-5% території України втрачено після Чорнобильської аварії внаслідок радіоактивного забруднення земель. Крім того, 20% населення України проживає в зоні функціонування ліній електропередач (близько 2% території), тобто санітарно недопустимого електромагнітного забруднення.

Таким чином, можна зробити висновки, що головними факторами, які негативно впливають на природне навколишнє середовище і зумовлюють техногенне руйнування нашої планети, являються: індустріалізація, хімізація с.г., меліорація земель, розвиток транспорту та енергетики, збільшення видобутку корисних копалин, урбанізація.

В результаті негативної дії цих факторів навколишньому середовищу завдається велика шкода, а насамперед, самій людині.

Відбувається теплове забруднення внаслідок недосконалих методів утилізації тепла. Збільшується вміст в атмосфері CO2. Він в атмосфері діє як скло в парнику: пропускає сонячне світло і затримує тепло розігрітої Сонцем поверхні Землі. Це я вище називав «парниковий ефект». Для людства є два наслідки парникового ефекту: різко скорочуються врожаї зерна і піднімається рівень Світового океану на 2-3 м за рахунок танення полярних льодових шапок.

Окиси сірки і азоту, що потрапляють в атмосферу, випадають на землю кислотними дощами. Від них знижується врожайність с.г. культур на 3-8%; вимивається з грунту кальцій, калій і магній; гинуть ліса; отруюється вода озер, ставків, у яких гине риба і комахи; зникають птахи і тварини;

збільшується кількість гірських зсувів і селів; руйнуються пам‘ятки архітектури, житлові будинки; відбувається захворювання дихальних шляхів.

5.2.5. Пожежна небезпека

Пожежа – це стихійне поширення горіння, яке виявляється в нищівній дії вогню, що вийшов з-під контролю людини.

Горіння включає три необхідні інгредієнти: паливо, тепло і кисень-так званий пожежний трикутник. Пожежа виникає тільки тоді, коли вони всі наявні. Достатньо видалити один з них -і пожежа погасне.

Розрізняють два види горіння: гомогенне і гетерогенне. Зовнішніми ознаками гомогенного горіння є полум‘я, а гетерогенного -розжарювання.

У районі пожежі відбуваються суттєві зміни атмосфери: підвищується вміст токсичних газів, головним чином окису вуглецю, і зменшується кількість кисню.

Основними уражаючими факторами пожеж є : висока температура, задимлення великих районів, обмеження видимості, негативний вплив на психіку людей.

Для забезпечення успішної боротьби з пожежами в Україні створені воєнізовані й професійні пожежні частини Міністерства внутрішніх справ для охорони населених пунктів і об‘єктів промисловості, сільського і лісового господарства, відомчі професійні команди, протипожежні формування і добровільні пожежні дружини (команди) підприємств, організацій, установ, сільськогосподарських об‘єктів.

Основні способи гасіння пожежі: нахльостування або закидання ґрунтом кромки пожежі; улаштування загороджувальних і мінералізованих каналів і смуг; гасіння пожежі водою або розчинами вогнегасних хімікатів, відпал (пуск зустрічного вогню).

5.3. Виробничий і побутовий травматизм, наслідки і заходи щодо його попередження.

Досить високий рівень посідає побутовий травматизм, який прямо пов’язаний з порушенням чи недотриманням правил техніки безпеки при виконанні домашньої роботи або під час відпочинку.

Виробничою травмою, або нещасним випадком, вважають ушкодження раптового характеру, що сталися з робітниками у зв’язку з їх участю у виробничому процесі та супроводжувалися порушенням фізіологічної функції працівників. Виробничі травми можуть бути механічними, хімічними, термічними й електричними. Механічними ушкодженнями в умовах виробництва називають забиття, ураження м’яких тканин тіла, крововиливи, переломи кісток, розтягнення і розриви зв’язок і сухожилків, вивихи, потрапляння сторонніх предметів у різні органи тощо.

Токсичні речовини найчастіше уражають шкіру, слизові оболонки і очі. Вплив термічних агентів призводить до опіків. Електричний струм великої напруги зумовлює електричні травми.

Виробничі травми є наслідком порушення робітниками правил техніки безпеки і виробничої дисципліни, нераціональної організації робочого місця, несправності технологічного обладнання і порушення технології виробництва, а також нехтування засобами індивідуального захисту. У багатьох випадках травма пов’язана з неуважністю або недосвідченістю робітників. Серед причин травм на виробництві виділяють організаційно-технічні, тобто зумовлені недостатньою механізацією виробничих процесів і неправильною організацією праці, та санітарні, зокрема несприятливі мікрокліматичні умови, нераціональне освітлення, шум, вібрація, вплив токсичних речовин. Надмірна вологість повітря, що спричиняє туман, а відтак і погіршує видимість, підвищує небезпеку травматизму від рухомих частин машин.

Особливої уваги з боку медичного персоналу потребує очний травматизм, який характерний для металургійної і, особливо, машинобудівної

промисловості. Внаслідок поверхневих ушкоджень рогівка ока мутніє, через що знижується гострота зору. Гнійне запалення може призвести до сліпоти.

У металургійній та машинобудівній промисловості спостерігаються опіки. Вони характерні для таких професійних груп, як заливальники, вагранники, ливарники, прокатники, ковалі та ін. Захищають очі від розпечених часточок спеціальними екранами, а також за допомогою спеціального одягу. Важливим заходом запобігання хімічним опікам є раціоналізація транспортування кислот і основ, а також маніпуляцій, пов’язаних з їх використанням і виготовленням розчинів.

Електротравми бувають рідше, але за тяжкістю перебігу вони посідають одне з перших місць серед професійних захворювань. Виникають через відсутність заземлення, недотримання технічних правил, нехтування захисними пристосуваннями. Ступінь інтенсивності ураження людини електрострумом залежить від енергії, на яку він перетворюється. Клінічно ураження проявляється у вигляді тетанічного скорочення м’язів, яке утруднює дихання, і сильного болю. Після вимкнення струму, з’являються запаморочення, головний біль, блювота, психічні порушення, депресія, іноді – судоми епілептоїдного типу. Підвищується внутрішньо-черепний тиск, виникає кисневе голодування мозку, іноді – опіки шкіри.

Запобіжні заходи полягають у дотриманні техніки безпеки під час роботи електроустановок. Робітники повинні користуватися засобами індивідуального захисту та інструментами з ручками, що мають ізоляцію. Важливу роль у запобіганні травматизму відіграють організація обліку виробничого травматизму, а також реєстрація та аналіз його причин. До роботи людину можна допускати лише після ознайомлення з правилами техніки безпеки.

Для зниження травматизму потрібна правильна організація праці, обладнання робочого місця, справність інструментів. На новозбудованих і реконструйованих підприємствах з високим рівнем механізації виробничий травматизм значно нижчий. Стружкоподрібнювачі, механічні пристосування

для видалення стружки з верстатів, а також безпосередньо від різальних інструментів забезпечують профілактику порізів. У цехах хімічних виробництв переливання кислот має бути механізоване, шляхом перекачування кислоти за допомогою стиснутого повітря через переливні труби. Не можна захаращувати проходи і майданчики. Запобіганню травмам в умовах виробництва сприяє дотримання санітарних правил, особливо щодо раціонального освітлення. До індивідуальних засобів захисту від травм належать спеціальний одяг, взуття, головні убори. Захищають очі від ушкоджень спеціальні окуляри, маски і щитки.

Важлива роль у зниженні травматизму належить організації медико-санітарного обслуговування робітників. Проведення інженерно-лікарняних бригад, які проводять експертну оцінку робочих місць та планують і проводять заходи по зниженню виробничого травматизму.

5.4. Висновки до розділу 5

У даному розділі було проаналізовано вимоги з охорони праці та безпеки в надзвичайних ситуаціях. Проаналізовано основні законодавчі актом з охороні праці зокрема: НПАОП 0.00-7.15-18 «Вимоги щодо безпеки та захисту здоров’я працівників під час роботи з екранними пристроями» та ДСТУ Б В.2.5-82:2016 «Електробезпека в будівлях і спорудах. Вимоги до захисних заходів від ураження електричним струмом».

Було досліджено шляхи підтримання безпеки життєдіяльності в системі "людина - життєве середовище", в результаті чого було зроблено висновки, що головними факторами, які негативно впливають на природне навколишнє середовище і зумовлюють техногенне руйнування нашої планети, являються: індустріалізація, хімізація с.г., меліорація земель, розвиток транспорту та енергетики, збільшення видобутку корисних копалин, урбанізація.

Було досліджено виробничий і побутовий травматизм, наслідки і заходи щодо його попередження, в результаті чого було встановлено що для

зниження травматизму потрібна правильна організація праці, обладнання робочого місця, справність інструментів, що важлива роль у зниженні травматизму належить організації медико-санітарного обслуговування робітників, проведення інженерно-лікарняних бригад, які проводять експертну оцінку робочих місць та планують і проводять заходи по зниженню виробничого травматизму.

РОЗДІЛ 6
ЕКОЛОГІЯ

6.1. Методи узагальнення екологічної інформації

Для прийняття обґрунтованих управлінських рішень в галузі екології та охорони навколишнього середовища велике значення має створення бази даних екологічної інформації. Екологічна інформація носить синтетичний, аналітичний та оперативний характер про стан довкілля.

Синтетичний характер інформації є визначальним для аналізу глобального антропогенного впливу на великомасштабні екосистеми. Синтетичний характер інформації формується шляхом обліку обставин, що в першу чергу відносяться до охорони навколишнього середовища і раціонального використання природних ресурсів.

Аналітичний характер інформації диктується наявністю великого обсягу різнорідних і децентралізованих даних, які повинні бути приведені в порівняльний вигляд.

Оперативний характер інформації формується із задач оперативного впливу на локальні важелі деградації навколишнього середовища і виконує попереджувальні функції в різних напрямках підтримки рівноваги і охорони навколишнього середовища, відтворення її багатств і в першу чергу з позиції дбайливої та раціональної витрати матеріально- і паливносировинних ресурсів.

Під час збору й обробки інформації слід брати до уваги наступні аспекти:

- новизну та розміщення масивів екологічної статистики;
- інерційність носіїв інформації;
- вплив фонових та завадних факторів;
- багатоетапний збір статистичних даних і нормативних параметрів.

Вибір шляхів і методів одержання інформації, джерел визначаються конкретною ситуацією. У більшості випадків доцільним виявляється сполучення різних методів і джерел.

В процесі інформаційного пошуку, важливо дотримуватися наступних принципів:

- принцип адекватності джерела інформації;
- принцип обов'язкової оцінки джерела інформації;
- принцип надмірності і розумної достатності.

Інформацію про екологічний стан довкілля можна одержати з різних джерел:

- джерела правової інформації, що інформують про правову базу основ природокористування;
- джерела науково-теоретичної інформації, що відображають здобутки знань чи діяльності у формі карт, таблиць, описів чи фізичних теорій, тощо.
- джерела первинної інформації (статистична звітність);
- джерела вторинної інформації (екологічна експертиза та екологічні паспорти);
- додаткові джерела інформації – джерела, де інформація отримана з інвентаризації викидів, скидів забруднюючих речовин в об'єкти довкілля, кількісна оцінка викидів шкідливих речовин різними видами транспорту тощо.

Успіх всієї інформаційної роботи істотно залежить від достовірного балансу між принципами надмірності і розумної достатності.

Екологічне дослідження, як уже зазначалося, складається з наступних етапів: збір інформації, її обробка та узагальнення.

Кожен з названих етапів має особливе значення. Останній етап – це та стадія роботи з екологічною інформацією, коли зібраний і оброблений екологічний матеріал потребує узагальнення та наочного подання. Методами наочного подання фізичних величин, що використовують для більш раціонального та систематизованого викладення кількісної інформації у цифровому вигляді, є статистичні таблиці та графіки.

Табличний метод – метод раціонального та систематизованого викладення цифрової інформації. За допомогою таблиць можна досягти наступних цілей: систематизація цифрової інформації; полегшення і прискорення ефекту сприйняття; інтенсифікація пізнавального процесу; економія місця при викладенні інформації. В процесі обробки статистичних даних користуються допустимими, робочими таблицями. Їх слід відрізняти від допоміжних таблиць (таблиць коефіцієнтів). Статистичними таблицями вважаються тільки ті, що містять наслідки статистичного аналізу і процесів.

Графічний метод – метод умовного зображення кількісних і якісних властивостей дослідженого об’єкта довкілля у вигляді крапок, ліній, стовпчиків, кругів або різноманітних фігур. Мета графічного методу полягає у: популізації кількісної (цифрової) інформації; забезпеченні допустимої кількості інформації; узагальненні кількісної інформації.

Призначення графіків багатогранне: порівняння між собою різних величин; характеристика складу, структури і структурних зрушень сукупностей; з’ясування ступеня розповсюдження явищ і ознак у просторі; виявлення хронологічних явищ і ознак; дослідження темпів, тенденцій, закономірностей і перспектив розвитку явищ. За призначенням графіки поділяють на:

- аналітичні – дають можливість порівняння графічних образів;
- ілюстративні – допомагають порівнянням геометричних фігур показати динаміку зміни розмірів явищ;
- інформаційні – вміщують інформацію лише про об’єкти дослідження.

6.2. Отримання енергії за рахунок альтернативних джерел

Існуючі джерела енергії, визначені людством поділяють на два класи:

- Нетрадиційні (відновлювані) джерела енергії (НДЕ) – це джерела постійних або періодичних потоків енергії в навколишньому середовищі, які функціонують без участі людини (рис. 6.1).

- Невідновлювані природні запаси речовин і матеріалів, які використовує людина для отримання енергії (ядерне паливо, нафта, газ, вугілля). Вони вважаються скінченними джерелами енергії (СДЕ) (рис. 6.2).

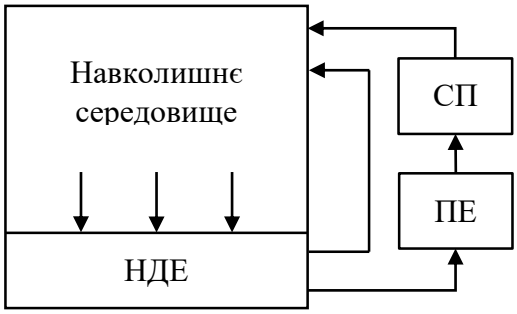


Рис. 6.1. Потік енергії від НДЕ

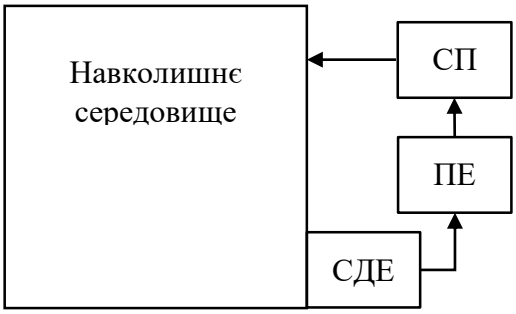


Рис. 6.2. Потік енергії від СДЕ

На рис. 6.1 і рис. 6.2 прийняті такі позначення: СП – споживачі енергії; ПЕ – перетворювачі енергії.

За класифікацією у [95] визначено п’ять джерел енергії:

- сонячне випромінювання;
- рух і притягання Сонця, Землі, Місяця;
- теплова енергія ядра Землі;
- ядерні реакції;
- хімічні реакції.

Перші три джерела – це є НДЕ, наступні джерела є СДЕ. На рис. 6.3 наведено існуючі НДЕ з можливими видами трансформації їх енергії в інші форми, з метою забезпечити зростаючі потреби людства у споживаній енергії.

6.2.1. Задачі використання НДЕ

Вивчаючи такі НДЕ як вітер, вода, сонячне випромінювання, біопаливо, тощо, необхідно дати відповідь на три основні питання:

- Який енергоресурс є у кожному НДЕ?
- З якою метою можна використати енергію НДЕ?
- Яка вартість енергії НДЕ порівняно з енергією інших джерел?

Для відповіді на третє питання треба усвідомити, що економічно обтрунтована експлуатація НДЕ може бути лише за виконання таких умов:

- чіткого виявлення і використання переваг зазначених вище НДЕ;
- максимально ефективного проведення процесу перетворення енергії НДЕ в енергетичних уставах ЕНУ на основі мінімізації втрат і максимізації економічних і соціальних потреб.

Рис. 6.3. НДЕ та їх застосування

1.3. Технічні проблеми застосування НДЕ

Проблеми використання та технічної реалізації систем з НДЕ можна розв'язати на основі таких дій:

- моніторингу навколишнього середовища:
- визначення характеристик споживачів енергії:

- узгодження джерел енергії зі споживачами (рис. 6.4):
- визначення методів керування енергосистемами з НДЕ.

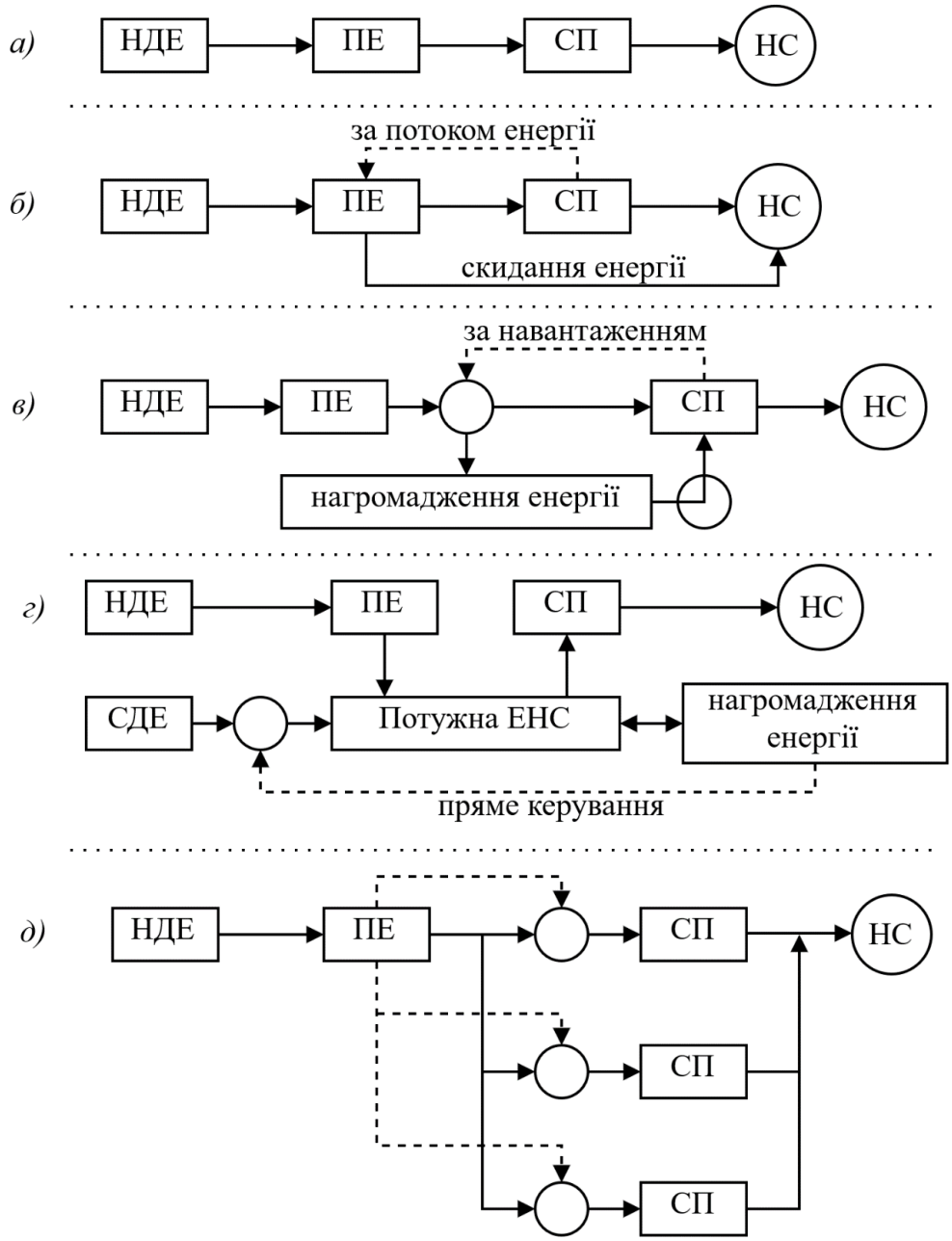


Рис. 6.4. Структури систем НДЕ – СП;

ПЕ – перетворювач енергії; СП – споживач енергії; НС – навколишнє середовище

Використовують три типи систем керування енергією від НДЕ:

- зі скиданням надлишку енергії;
- з нагромаджувачами (акумуляторами) енергії;
- з регулюванням навантаження відповідно до потреб споживачів.

ВИСНОВКИ

У дипломній роботі магістра досліджено існуючі методи та засоби розпізнавання мови та класифікацію систем розпізнавання мови. Проаналізовано існуючі реалізації систем розпізнавання мови на базі мікроконтролерних систем, у результаті встановлено, що існуючі реалізації можна покращити шляхом застосування поліпшених методів та засобів автоматичного розпізнавання мови, та застосування нової мікроконтролерної системи.

Визначено оптимальний метод автоматичного розпізнавання мови, який здійснюється за допомогою двох блоків: front-end та back-end процесорів. Front-end процесор займається зчитуванням звукового сигналу, його попередньою. фільтрацією, розбиттям на фрейми та накладанням вікна, визначенням початку та кінця мовлення та виділенням ознак за допомогою вейвлет-перетворення. Back-end процесор займається порівнянням представлення звукового сигналу у вигляді вектора ознак, з базою шаблонів відомих представлень мовленнєвих сигналів, використовуючи повнозв’язну штучну нейронну мережу.

За вибраними методами та засобами автоматичного розпізнавання мовлення, було змодельовано процес розпізнавання мовлення та здійснено його реалізацію на базі мікроконтролерної системи. Процес моделювання виконано засобами програмного пакету Matlab. У процесі моделювання було створено моделі попередньої фільтрації сигналу, розбиття сигналу на фрейми, виявлення голосової активації та виявлення ознак, за допомогою вейвлет-перетворення та класифікаційної штучної нейронної мережі. Отримані моделі було реалізовано на мікроконтролерній системі на базі плати розробки «STM32 Nucleo-64». Після чого було здійснено оцінку роботи отриманої системи та здійснено порівняння отриманих результатів з існуючими системами розпізнавання мовлення.

Обґрунтовано економічну ефективність проведення досліджень дипломної роботи магістра шляхом проведення відповідних розрахунків, що дало змогу встановити термін окупності на рівні 1,79 років при ціні 46749,71 грн.

Проаналізовано вимоги з охорони праці та безпеки в надзвичайних ситуаціях, що дало змогу визначити шляхи підтримання безпеки життєдіяльності в системі "людина-життєве середовище" та дослідити виробничий і побутовий травматизм, наслідки і заходи щодо його попередження.

Проаналізовано методи узагальнення екологічної інформації, та отримання енергії за рахунок альтернативних джерел.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Willis W. Speech recognition: instead of typing and clicking, talk and command. *THE Journal*. 1998. No. 1. P. 18-22.

2. Stone B. Look who’s talking now. *Newsweek*. 1999. P. 48-51.

3. Zue V., Cole R., Ward W. Speech Recognition.Survey of the State of the Art in Human Language Technology. Kauii. 1996.

4. Mengjie Z. Overview of speech recognition and related machine learning techniques URL: http://www.mcs.vuw.ac.nz/comp/Publications/archive/CS-TR-01/CSTR-01-15.pdf.

5. Ganapathiraju A., Hamaker J., Picone J. Syllable-Based Large Vocabulary Continuous Speech Recognition. *IEEE Transactions On Speech And Audio Processing*, 2001. Vol. 9, No. 4.

6. Furui S. 50 years of Progress in speech and Speaker Recognition Research. *ECTI Transactions on Computer and Information Technology*. 2005. Vol.1, No.2.

7. Davis K., Biddulph R., Balashek S. Automatic Recognition of spoken Digits. *J. Acoust. Soc. Am*. 1952. Vol. 24 No. 6. P. 637-642.

8. Forgie J., Forgie C. Results obtained from a vowel recognition computer program. *J.A.S.A.* 1959. Vol. 31 No. 11. P. 1480-1489.

9. Martin T., Nelson A., Zadell H. Speech recognition by feature abstraction techniques. Tech. Report AL-TDR-64-176, Air Force Avionics Lab, 1964.

10. Suzuki J., Nakata K. Recognition of Japanese Vowels Preliminary to the Recognition of Speech. *J. Radio Res. Lab*. 1961. Vol. 37, No. 8. P. 193-212.

11. Sakai T., Doshita S. The phonetic typewriter, information processing. *Proc. IFIP Congress*. 1962. P. 445-45.

12. Reddy D. An Approach to Computer speech Recognition by direct analysis of the speechwave. *Tech.Report No.C549. Computer Science Department of Stanford University*. 1996.

13. Velichko V., Zagoruyko N. Automatic Recognition of 200 words. *Int. J. Man-Machine Studies*. 1970. No. 2. P. 223.

14. Sakoe H., Chiba S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Trans.Acoustics, Speech, Signal Proc*. 1978. Vol. ASSP-26, No. 1. P. 43-49.

15. Itakura F. Minimum Prediction Residula Applied to Speech Recognition. *IEEE Trans. Acoustics, Speech, Signal Proc*. 1975. Vol. ASSP-23, No. 1. P. 67-72.

16. Tappert C., Chapman W., Dixon X., Rabino A. Automatic recognition of continuous speech utilizing dynamic segmentation, dual classification, sequential decoding and error recovery. *Rome Air Dev. Cen, Rome, NY, Tech. Report TR-71-146.* 1971.

17. Jelinek F., Bahl L., Mercer R. Design of a linguistic statistical decoder for the recognition of continuous speech. *IEEE Trans. Information Theory*. 1975. IT-21. P. 250-256.

18. Moore R. Twenty things we still don't know about speec. *Proc. CRIM/ FORWISS Workshop on Progress and Prospects of speech Research an Technology*. 1994.

19. Sakoe H. Two Level DP Matching A Dynamic Programming Based Pattern Matching Algorithm for Connected Word Recognition. *IEEE Trans. Acoustics, Speech, Signal Proc.* 1979. Vol. ASSP-27, No. 27. P. 588-595.

20. Bridle J., Brown M. Connected Word Recognition Using whole word templates. *Proc. Inst. Acoust. Autumn Conf*. 1979. P. 25-28.

21. Myers C., Rabiner L. A Level Building Dynaimc Time Warping Algorithm for Connected Word Recognition. *IEEE Trans. Acoustics, Speech Signal Proc*. 1981. Vol. ASSP-29. P. 284-297.

22. Lee C., Rabiner L. A Frame Synchronous Network Search Algorithm for Connected Word Recognition. *IEEE Trans. Acoustics, Speech, Signal Proc.* 1989. Vol. 37, No. 11. P. 1649-1658.

23. Furui S. Speaker independent isolated word recognition using dynamic features of speech spectrum. *IEEE Trans.Acoustics, Speech, Signal Processing*. 1986. Vol. ASSP-34. P. 52-59.

24. Katagiri S. Speech pattern recognition using neural networks, Pattern Recognition in Speech and Language Processing. *CRC Press*. 2003. P. 115-147.

25. Lippmann R. An introduction to computing with neural nets. *IEEE ASSP Mag*. 1987. Vol. 4, No. 2. P. 4-22.

26. Weibel A. Phoneme recognition using time-delay neural networks. *IEEE Trans. Acoustics, Speech, Signal Proc*. 1989. No. 37. P. 393-404.

27. Juang B., Furui S. Automatic speech recognition and understanding: A first step toward natural human machine communication. *Proc. IEEE*. 2000. Vol. 88, No. 8. P. 1142-1165.

28. Leggetter J., Woodland P. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech and Language*. 1995. No. 9. P. 171-185.

29. Varga A., Moore R. Hidden Markov model decomposition of speech and noise, *Proc. ICASSP*. 1990. P. 845-848.

30. Gales M., Young S. Parallel model combination for speech recognition in noise. Technical Report, CUED/F-INFENG/TR135. 1993.

31. Shinoda K., Lee C. A structural Bayes approach to speaker adaptation. *IEEE Trans. Speech and Audio Proc*. 2001. Vol. 9, No. 3. P. 276-287.

32. Afify M., Olivier Siohan O. Sequential Estimation With Optimal Forgetting for Robust SpeechRecognition. *IEEE Transactions On Speech And Audio Processing*. 2004. Vol. 12, No. 1. P. 19-26.

33. Afify M., Liu F., Jiang H. A New Verification-Based Fast-Match for Large Vocabulary Continuous Speech Recognition. *IEEE Transactions On Speech And Audio Processing*. 2005. Vol. 13, No. 4. P. 546-553.

34. Kinga S., Frankel J. Recognition, Speech production knowledge in automatic speech recognition. *Journal of Acoustic Society of America*. 2006.

35. Arminas V., Tamulevicius G., Navakauskas D., Ivanovas E. Acceleration of feature extraction for FPGA based speech recognition. *Proc. SPIE 7745, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments*. 2010. P. 511–516.

36. Tamulevičius G., Arminas V., Ivanovas E., Navakauskas D. Hardware accelerated FPGA implementation of Lithuanian isolated word recognition system, *Elektronika ir elektrotechnika (Electronics and Electrical Engineering)*. 2010. Vol. 99, No. 3. P. 57–62.

37. Chakravarty A. Speech recognition toolkit for the Arduino. 2013. URL: http: //arjo129.github.com/uSpeech/.

38. Natural Language Processor. *Sensory*. 2010. URL: www.sensoryinc.com/products/NLP-5x.html.

39. Choi J., You K., Sung W. An FPGA implementation of speech recognition with weighted finite state transducers. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*. 2010, P. 1602–1605.

40. Veitch R., Aubert L., Woods R., Fischaber S. Acceleration of HMM-based speech recognition system by parallel FPGA Gaussian calculation. *VI Southern Programmable Logic Conf*. 2010. P. 197–200.

41. Zhang G., Yin J., Liu Q., Yang Ch., A real-time speech recognition system based on the implementation of FPGA. *Proc. Cross Strait Quad-Regional Radio Science and Wireless Technology Conf*. 2011. P. 1375–1378.

42. Pan S., Lai C., Tsai B. The implementation of speech recognition systems on FPGA-based embedded systems with SoC architecture. *Int. Journal of Innovative Computing, Information and Control*. 2011. Vol. 7, No. 11. P. 6161–6175.

43. Sart D., Mueen A., Najjar W., Niennattrakul V. Accelerating dynamic time warping subsequence search with GPUs and FPGAs. *IEEE 10th Int. Conf. on Data Mining*. 2010. P. 1375–1378.

44. Zhang Y., Glass A.J. Fast spoken query detection using lowerbound dynamic time warping on graphical processing units. *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*. 2012. P. 5173–5176.

45. Davel M., and Martirosian O. Pronunciation dictionary development in resource-scarce environments. *Interspeech*. P. 1-8.

46. Davel M., Barnard E. The efficient generation of pronunciation dictionaries: Human factors during bootstrapping. *Interspeech/ICSLP*. 2004.

47. Zen H., Senior S., Schuster M. Statistical parametric speech synthesis using deep neural networks. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 2013. P. 7962-7966.

48. Xin L., Senior A., Gruenstein A., Sorensen J. Accurate and Compact Large Vocabulary Speech Recognition on Mobile Devices. *INTERSPEECH*. 2013.

49. Xue J., Cui X., Daggett G., Marcheret E., Zhou B. Towards high performance LVCSR in speech-to-speech translation system on smart phones. *Proc. Interspeech*. 2012.

50. Prasad R. BBN Transtalk: Robust multilingual two-way speech-to-speech translation for mobile platforms. *Computer Speech and Language*. 2013. Vol. 27. P. 475–490.

51. Chakravarty A. Speech recognition toolkit for the Arduino. 2013. URL: http://arjo129.github.com/uSpeech/.

52. Shan J., GWu G., Hu Z., Tang X. Search by voice in Mandarin Chinese. *Proceedings of interspeech*. 2010. P. 354-357.

53.Hsuan Y., Jansche M. Moreno P. Deploying search by voice in Cantonese. *Proceedings of interspeech*. 2011.

54. Zhong Y., Raman T., Burkhardt C., Jeffrey P. JustSpeak: Enabling Universal Voice Control on Android. 2014.

55. Nakamura K. Speech recognition chip for monosyllables. *Asia and South Pacific Design Automation Conference*. 2001. P. 396–399.

56. Shi Y. Single-chip speech recognition system based on 8051 microcontroller core. *IEEE Transactions on Consumer Electronics*. 2001. Vol. 47, No. 1. P. 149–153.

57. Burchard B. A single chip phoneme based human machine speech recognition system for consumer applications. *IEEE Transactions on Consumer Electronics*. 2000. Vol. 46, No. 3. P. 914–919.

58. Karam Z., Campbell W. A new Kernel for SVM MIIR based Speaker recognition. Lexington: MIT Lincoln Laboratory.

59. Lori F. An Improved Endpoint Detector for Isolated Word Recognition. *IEEE transactions on acoustics, speech, and signal processing*. Vol. assp-29, No. 4.

60. Anand S., Singh V., Rajoriya D.K. Database Development and Analysis of Spoken Hindi Hybrid Words Using Endpoint Detection. *International Journal of Electronics and Computer Science Engineering*. 2012. Vol. 1.

61. Rabiner L.R., Juang B.H., Yegnanarayana B. Fundamentals of Speech Recognition. Pearson Publishers. 2010.

62. Garg A., Poonam N. Connected digits recognition using Distance calculation at each digit. *IJCEM International Journal of Computational Engineering & Management*. 2011. Vol. 14. P. 2230-2293.

63. Taabish G., Singh A., Vijay S. An Improved Endpoint Detection Algorithm using Bit Wise Approach for Isolated, Spoken Paired and Hindi Hybrid Paired Words. *International journal of computer applications*. 2014. Vol. 92, No. 15.

64. Rajoriya K., Anand S., та Maheshwari P. Hindi paired word recognition using probabilistic neural network. *International Journal Computational Intelligence studies*. 2010. Vol. 1.

65. Hisashi W. Normalization of Vowels by Vocal Tract Length and Its Applications to Vowel Identification. *IEEE Transactions on Acoustics, Speech and Signal Processing*. 1997. Vol. 25.

66. Nicolás M., John L. MFCC Compensation for improved recognition filtered and band limited speech. Boulder: University of Colorado at Boulder.

67. Samudravijay K. Speech and Speaker recognition report. URL: http://cs.jounsuu.fi/pages/tkinnu/reaserch/index.

68. Singer E. Torres-Carrasquillo P. Gleason T. Accoustic, phonetic and discriminative approach to automic Language Idantification.

69. Viet B., Laurent B., Tanja S. Acousticphonetic unit similarities for context dependant acoustic model portability. Pittsburgh: Mellon University.

70. Tran D., Fuzzy Approaches to Speech and Speaker Recognition. *A thesis submitted for the degree of Doctor of Philosophy of the university of Canberra*. 2000.

71. Itakura F. Minimum Prediction Residula Applied to Speech Recognition. *IEEE Trans. Acoustics*. 1975. Vol. ASSP-23, No. 1. P. 67-72.

72. Rabiner L., Juang B. Fundamentals of speech processing. New Jersey: Prentice Hall. 1993.

73. Baker J.K. Stochastic modeling for automatic speech recognition. Ed. New York: Academic. 1975.

74. Jelinek F. Continuous speech recognition by statistical methods. *Proc. IEEE*. 1976. Vol. 64, No. 4. P. 532-557.

75. Poor H. An Introduction to Signal Detection and Estimation. *Springer Texts in Electrical Engineering*. 1988.

76. Naik J. Speaker verification over long distance telephone lines. *Proc. ICASSP*. 1989. P. 524-527.

77. Sonkamble B., Doye D. Speech Recognition Using Vector Quantization through Modified K-means LBG Algorithm. *Computer Engineering and Intelligent Systems*. 2012. Vol. 3, No. 7.

78. Bahl L. A method of Construction of acoustic Markov Model for words. *IEEE Transaction on Audio, speech and Language Processing*. 1993.Vol.1.

79. McCullough W., Pitts W. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bull. Math Biophysics*. 1943. Vol. 5. P. 115-133.

80. Vapnik V. The Nature of Statistical Learning Theory. *Springer, New York*. 1995.

81. Furui S. Speaker dependent feature extraction, recognition and processing techniques. *Speech communication*. 1991. Vol. 10. P. 505- 520.

82. Svendsen T., Paliwal K., Harborg E., Husy P. O. *Proc. ICASSP'89*, 1989.

83. Katagiri S. Speech Pattern recognition using Neural Networks.

84. Rabiner L., Wilpon, J. Considerations in applying clustering techniques to speaker-independent word recognition. *Journal of Acoustic Society of America*. 1979. Vol. 66, No. 3. P. 663-673.

85. Wilpon J., DeMarco D., Mikkilineni R. Isolated word recognition over the DD telephone network -Results of two extensive field studies. *Proc. ICASSP*. 1988. P. 55-58.

86. Huang X., Deng L. An Overview of Modern Speech Recognition. Microsoft Corporation

87. Meyer Y. Ondelettes et fonction splines. Paris: Seminare EDP. Ecole Polytechnique. 1986.

88. Mallat S. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Part. Anal. and Mach. Intell*. 1989. Vol. 11, No. 7. P. 674-693.

89. Mallat S. Multifrequency channel decomposition of images and wavelet models. *IEEE Trans. Acoust., Speech, Signal Processing*. 1989. Vol. 37. P. 2091-2110.

90. Daubechies I. Orthonormal bases of compactly supported wavelets. *Comm. Pure and Appl. Math*. 1988. Vol. 41. P. 909-996.

91. Vetterli M., Herley C. Wavelets and filter banks: Theory and design. *IEEE Trans. Acoust., Speech, Signal Processing*. 1992. Vol. 40, No. 9. P. 2207-2232.

92. Daubechies I. Ten Lectures on Wavelets. SIAM. 1992.

93. Wasserman D.P. Neural computing: theory and practice. New York: Van Nostrand Reinhold Co. 1989.

94. Wang B., Zhang L., Liu J. Embedded speech recognition system on 8-bit MCU core. Beijing: Department of Electronic Engineering. 2004.

95. Шевченко Л.С. Основи економічної теорії. Харків: Право. 2008. С. 448.

96. Твайделл Д., Уєйр А. Возобновляемые источникик энергии. Пер. с англ. Москва: Энергоатомиздат. 1990. С. 390.

Додаток А

Тези конференцій

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Тернопільський національний технічний університет імені Івана Пулюя (Україна)
Національна академія наук України
Університет імені П'єра і Марії Кюрі (Франція)
Маріборский університет (Словения)
Технічний університет у Кошице (Словаччина)
Вільнюський технічний університет ім. Гедимінаса (Литва)
Шяуляйська державна колегія (Литва)
Жешувський політехнічний університет ім. Лукасевича (Польща)
Білоруський національний технічний університет (Республіка Білорусь)
Міжнародний університет цивільної авіації (Марокко)
Національний університет біоресурсів і природокористування України (Україна)
Наукове товариство ім. Шевченка
ГО «Асоціація випускників Тернопільського національного технічного університету імені Івана Пулюя»

АКТУАЛЬНІ ЗАДАЧІ СУЧАСНИХ ТЕХНОЛОГІЙ

Збірник

тез доповідей

Том II

VIII Міжнародної науково-технічної конференції молодих учених та студентів

27-28 листопада 2019 року

УКРАЇНА
ТЕРНОПІЛЬ – 2019

ПЕРЕДАВАННЯ ЕКГ В СИСТЕМАХ ТЕЛЕМОНІТОРИНГУ

76. **Є.В. Тиш, Я.І. Юськів**
МОДЕЛЬ ВИЯВЛЕННЯ ВПЛИВУ ДЕФЕКТІВ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ НА НАДІЙНІСТЬ КОМП'ЮТЕРНИХ СИСТЕМ 103

77. **Р.Б. Трембач, Є. В. Кучірка**
ДОСЛІДЖЕННЯ СИСТЕМИ АВТОМАТИЧНОГО УПРАВЛІННЯ ЗА ДОПОМОГОЮ MATLAB 105

78. **Р.М. Фудаль, М.І. Яворська**
МОДЕЛЮВАННЯ РЕЖИМІВ РОБОТИ ПРИЛАДУ ДЛЯ КОНТРОЛЮ ЯКОСТІ ДОРІЖОК ПІДШИПНИКІВ КОЧЕННЯ 107

79. **Л.В. Хвостівська, І.Ю. Дедів, Д.В. Ісаєнко**
ГЕНЕРУВАННЯ РАДІОСИГНАЛІВ ДЛЯ ТЕСТУВАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ КОМП'ЮТЕРНИХ РАДІОСИСТЕМ 108

80. **М.О. Хвостівський, І.М.Паньків, Я.С.Моргулець**
СИНФАЗНИЙ МЕТОД ОЦІНЮВАННЯ ПСИХОЕМОЦІЙНОГО СТАНУ ЛЮДИНИ ЗА ЕЛЕКТРОЕНЦЕФАЛОСИГНАЛАМИ 110

81. **О. О.Цебрик**
МЕТОДИ ТА ЗАСОБИ ПОБУДОВИ СПЕЦІАЛІЗОВАНИХ КОМП'ЮТЕРНИХ СИСТЕМ ДЛЯ ОЦІНЮВАННЯ ЯКОСТІ БЕНЗИНУ 112

82. **О.П. Ясній, Б.І. Цюприк**
ПРОТОКОЛ MQTTB IoT 113

83. **В. А. Часник, Н. С. Луцик**
МЕТОДИ ТА ЗАСОБИ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВИ НА БАЗІ МІКРОКОНТРОЛЕРНОЇ СИСТЕМИ 114

84. **Н.Я. Черкас, Ю.Ю. Замосьний, А.А. Липак**
ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ЗНАХОДЖЕННЯ ЕФЕКТИВНОГО РОЗМІРУ ПОВІТРЯНОГО ЗАЗОРУ ЗА ДАНИМИ СЕНСОРІВ 115

85. **І.В. Чихіра, С.Ю. Мокрійчук, Т.І. Афтанащук**
АВТОМАТИЗОВАНА СИСТЕМА БЛОКУ КЕРУВАННЯ КОНТАКТОРОМ ДЛЯ МЕРЕЖ ПОСТІЙНОГО ТА ЗМІННОГО СТРУМУ 117

86. **І.В. Чихіра, О.О. Ремінник, Т.Б. Смачило**
АВТОМАТИЗОВАНА ІНФОРМАЦІЙНА СИСТЕМА ДЛЯ ОБЛІКУ ДІЯЛЬНОСТІ АВТОРЕМОНТНОГО ПІДПРИЄМСТВА 118

87. **Г.В. Шимчук, Р.М. Небесний**
СТРАТЕГІЯ ПАРАЛЕЛІЗМУ ДЛЯ АЛГОРИТМУ MRRR 119

88. **В.Р. Шищак, Р.М. Карабін, В.П. Кубашок, О.В. Тотосько**
РОЗРОБКА ВИСОКОШВИДКІСНОЇ МАГІСТРАЛЬНОЇ ЛІНІЇ ПЕРЕДАЧІ 121

УДК 004.5
В. А. Часник, Н. С. Луцик, докт. філос.
Тернопільський національний технічний університет імені Івана Пулюя, Україна

МЕТОДИ ТА ЗАСОБИ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВИ НА БАЗІ МІКРОКОНТРОЛЕРНОЇ СИСТЕМИ

V. A. Chasnyk, N. S. Lutsyk Ph. D.
METHODS AND MEANS OF SPEECH RECOGNITION ON MICROCONTROLLER SYSTEM

Сучасні комп'ютерні системи обробки мови досягнули великої складності та високої точності [1] внаслідок використання складних алгоритмів обробки сигналів та появи потужних обчислювальних засобів. Тим не менш, ріст продуктивності вбудованих процесорів (які також називаються мікроконтролерами) дозволив реалізувати розпізнавання мови в споживчій електроніці, де вартість має найбільше значення.

Застосування дешевих мікроконтролерів для автоматичного розпізнавання мови накладає обмеження на складність системи внаслідок малого об'єму оперативної та постійної пам'яті і невеликої обчислювальної потужності. Тому вибір параметрів та методів її реалізації є досить складною задачею.

Нещодавно опубліковано декілька робіт у цьому напрямі, де запропоновано застосовувати спеціалізовані інтегральні схеми (ASIC) [2, 3] або окремі спеціалізовані процесори [4, 5]. Недоліком вказаних систем є висока вартість та складність розробки. У роботі [6] скористалися недорогим стандартним мікроконтролером, однак достатню точність розпізнавання досягнуто лише для дикторів, на мовленні яких систему навчено.

На основі проведеного аналізу існуючих методів та засобів автоматичного розпізнавання мови поставлено за мету створити недорогу систему розпізнавання мови на базі мікроконтролера STM32F446. Користуючись стандартними інструменти розробки і застосувавши новий метод попередньої обробки голосового сигналу для мікроконтролера та навчання класифікаційної нейронної мережі, необхідно запрограмувати мікроконтролер початкового рівня для автоматичного розпізнавання мови з достатньою точністю.

Література

1. Mirghafori N. Multi-bandspeechrecognition: A summaryofrecentworkat ICSI / Mirghafori // TechnicalReport TR97-051. ICSI / Mirghafori. – Berkeley: ICSI, 1997

2. Sensory | EmbeddedVoice&Biometric AI [Електронний ресурс] – Режим доступу до ресурсу: http://www.sensoryinc.com/.

3. Nakamura K. Speechrecognitionchipformonosyllables / Nakamura // ProceedingsoftheAsiaandSouthPacificDesignAutomationConference / Nakamura., 2001. – С. 396–399.

4. Shi Y. Y. Single-chipspeechrecognitionsystembasedon 8051 microcontrollercore / Y. Y. Shi, J. Liu, R. S. Liu // IEEE TransactionsonConsumerElectronics / Y. Y. Shi, J. Liu, R. S. Liu., 2001. – (том 47; вип. 1). – С. 149–153.

5. Burchard B. A singlechipphonemebasedhumanmachinespeechrecognitionsystemforconsumerapplications / B. Burchard, R. Romer // IEEE TransactionsonConsumerElectronics / B. Burchard, R. Romer., 2000. – (том 46; вип. 3). – С. 914–919.

6. Nicolas J. Microcontrollerimplementationof a voicecommandrecognitionsystemforhuman-machineinterfaceinembeddedsystems / Nicolas // Proceedingsof 10th IEEE InternationalConferenceonEmerging Technologies andFactoryAutomation / Nicolas., 2006.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ

М А Т Е Р І А Л И

VII НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

«ІНФОРМАЦІЙНІ МОДЕЛІ, СИСТЕМИ ТА ТЕХНОЛОГІЇ »

11–12 грудня 2019 року

ТЕРНОПІЛЬ
2019

В. Лукашук
ЗАСОБИ ДИСТАНЦІЙНОГО КОНТРОЛЮ ПАРАМЕТРІВ ВАНТАЖУ В ЛОГІСТИЧНИХ СИСТЕМАХ 128
А. Мельничук, М. Хвостівський, І. Горбовий
ЗАБЕЗПЕЧЕННЯ КОМП'ЮТЕРНИХ ДІАГНОСТИЧНИХ СИСТЕМ 129
К. Моха, М. Хвостівський, А. Кравчук
КОМП'ЮТЕРНІ СИСТЕМИ ГЕНЕРУВАННЯ ТЕСТОВИХ СИГНАЛІВ КРОВОНОСНИХ СУДИН ТА СІТКІВКИ ОКА ЛЮДИНИ 130
В. Нестор, В. Яцишин
ПРОЦЕДУРА КЛАСИФІКАЦІЇ АТРИБУТІВ ЗА ХАРАКТЕРИСТИКАМИ ЯКОСТІ КОМП'ЮТЕРНИХ СИСТЕМ 131
А. Паламар
ПРОГРАМНО-АПАРАТНИЙ КОМПЛЕКС ДЛЯ ДИСТАНЦІЙНОГО МОНІТОРИНГУ СТАНУ ДЖЕРЕЛ БЕЗПЕРЕБІЙНОГО ЕЛЕКТРОЖИВЛЕННЯ 132
Н. Паляниця, В. Дорофей
РОЗРОБКА ПРОГРАМНОГО ПАКЕТУ ДЛЯ РОЗМІЧУВАННЯ МЕДИЧНИХ ЗОБРАЖЕНЬ У МАШИННОМУ НАВЧАННІ 133
Л. Пуляк, С. Лупенко
МЕТОДИ ОПРАЦЮВАННЯ МЕДИЧНИХ ЗОБРАЖЕНЬ В КОМП'ЮТЕРНИХ СИСТЕМАХ 135
Б. Равчак
ХАРАКТЕРИСТИКА МЕТОДОЛОГІЇ JAMSTACK 136
Є. Сов'як, Є. Тиш
МЕТОДИ ТА ЗАСОБИ ПОПЕРЕДНЬОГО ОПРАЦЮВАННЯ ЕКГ ДЛЯ СИСТЕМИ ТЕЛЕМОНІТОРИНГУ 137
В. Стеблик, У. Поливана
МЕРЕЖЕВИЙ МОНІТОРИНГ ЯК ЗАСІБ АНАЛІЗУ ІНФОРМАЦІЙНИХ ПРОЦЕСІВ У ЛОКАЛЬНІЙ І ГЛОБАЛЬНІЙ МЕРЕЖІ 138
Є. Тиш, О. Зима
ВИБІР КРИТЕРІЇВ ЕФЕКТИВНОСТІ БЕЗПРОВІДНИХ ТЕЛЕМЕТРИЧНИХ МЕРЕЖ 139
С. Туркот
НЕЙРОННІ МЕРЕЖІ В СИСТЕМАХ БІОМЕТРИЧНОЇ АУНТЕНТИФІКАЦІЇ 140
О. Цебрик
МЕТОДИ ТА ЗАСОБИ ПОБУДОВИ СПЕЦІАЛІЗОВАНИХ КОМП'ЮТЕРНИХ СИСТЕМ ДЛЯ ОЦІНЮВАННЯ ЯКОСТІ БЕНЗИНУ 141
Б. Цюприк, О. Ясній
БЕЗПЕКА МЕРЕЖІ ІНТЕРНЕТУ РЕЧЕЙ 142
В. Часник, Н. Луцик
ПРОЦЕС АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВИ НА БАЗІ МІКРОКОНТРОЛЕРНОЇ СИСТЕМИ 143
Я. Чирський, В. Яцишин
АНАЛІЗ МОДЕЛІ ЗРУЧНОСТІ ВИКОРИСТАННІ ДЛЯ ОЦІНЮВАННЯ ЯКОСТІ ЛЮДИНО-МАШИННОЇ ВЗАЄМОДІЇ 144
Х. Юркевич, А. Луцків, Н. Попович
АНАЛІЗ ЕФЕКТИВНОГО ОПРАЦЮВАННЯ ВЕЛИКИХ ТЕКСТОВИХ ДАНИХ ЗАСОБАМИ ХМАРНИХ СЕРВІСІВ 145
Я. Юськів, Є. Тиш
БАЗА ДАНИХ ПІДТРИМКИ ПРОЦЕСУ ОЦІНЮВАННЯ ВПЛИВУ ДЕФЕКТІВ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ НА НАДІЙНІСТЬ КОМП'ЮТЕРНИХ СИСТЕМ 146

УДК 004.5
В. Часник, Н. Луцик
(Тернопільський національний технічний університет імені Івана Пулюя)

ПРОЦЕС АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВИ НА БАЗІ МІКРОКОНТРОЛЕРНОЇ СИСТЕМИ

UDC 004.5
V. Chasnyk, N. Lutsyk
(Ternopil Ivan Puluj National Technical University, Ukraine)

PROCESS OF AUTOMATIC SPEECH RECOGNITION ON MICROCONTROLLER SYSTEM

На основі аналізу існуючих методів та засобів автоматичного розпізнавання мови створено недорогу систему розпізнавання мови на базі мікроконтролера STM32F446. Для проектування даної системи використано стандартні інструменти розробки і застосовано новий метод попередньої обробки голосового сигналу та навчання штучної нейронної мережі для мікроконтролера.

Розробка системи автоматичного розпізнавання мовлення складається з двох основних частин: розробка front-end процесора та back-end процесора (рис. 1).

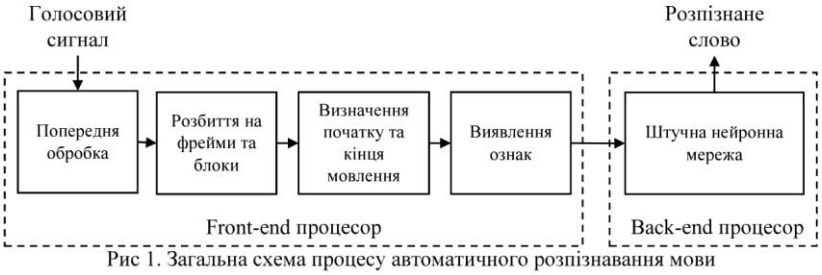


Рис 1. Загальна схема процесу автоматичного розпізнавання мови

Front-end процесор виконує наступні задачі: зчитування звукового сигналу, його попередня фільтрація, розбиття на фрейми, застосування віконної функції та збиття у блоки, попередня обробка мовленнєвого сигналу та виділення ознак. Back-end процесор займається порівнянням звукового сигналу, представленого front-end процесором, з базою шаблонів відомих мовленнєвих сигналів.

Фільтрація звукового сигналу здійснюється за допомогою КІХ-фільтру низьких частот. Відфільтрований сигнал розбивається на фрейми довжиною 5 мс, до яких застосовується віконна функція Хеммінга. Отримані фрейми об'єднуються у блоки таким чином, щоб сусідні блоки мали перетин в декілька фреймів.

Для кожного блоку сигналу обчислюються потужність та кількість нульових моментів, за якими визначається момент початку та кінця мовлення. Після чого сигнал в кожному блоці нормалізується.

Виділення ознак в системі автоматичного розпізнавання мовлення здійснено за допомогою дискретного вейвлет-перетворення. За материнський вейвлет вибрано вейвлет Добеші четвертого порядку з розкладом вхідного сигналу на 10 рівнів.

Для класифікації вхідного голосового сигналу використано багатошарову повнозв'язну нейронну мережу з трьома прихованими шарами. Для її навчання застосовано алгоритм зворотного поширення помилки.

Програмування мікроконтролерної системи для автоматичного розпізнавання мови було здійснено на базі мікроконтролера STM32F446.

За отриманими результатами, система розпізнавання мови показала достатню точність та швидкість розпізнавання, при цьому вартість такої системи залишається досить низькою.