

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ АЕРОНАВІГАЦІЇ,
ЕЛЕКТРОНІКИ ТА ТЕЛЕКОМУНІКАЦІЙ
КАФЕДРА ТЕЛЕКОМУНІКАЦІЙНИХ ТА РАДІОЕЛЕКТРОННИХ СИСТЕМ**

ДОПУСТИТИ ДО ЗАХИСТУ
Завідувач кафедри

Роман ОДАРЧЕНКО
“ _____ ” _____ 2022 р.

**КВАЛІФІКАЦІЙНА РОБОТА
(ПОЯСНЮВАЛЬНА ЗАПИСКА)**

ВИПУСКНИКА ОСВІТНЬОГО СТУПЕНЯ МАГІСТР

Тема: «Система ідентифікації людини за голосом на базі нейронної мережі»

Виконавець: _____ Данило ЛОМАКО
(підпис)

Керівник: _____ Володимир КЛИМЧУК
(підпис)

Консультанти з окремих розділів пояснювальної записки:

Консультант розділу «Охорона праці» _____ Батир ХАЛМУРАДОВ
(підпис)

Консультант розділу «Охорона навколишнього середовища»
_____ Євгеній БОВСУНОВСЬКИЙ
(підпис)

Нормоконтролер: _____ Денис БАХТІЯРОВ
(підпис)

Київ 2022

НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Факультет аеронавігації, електроніки та телекомунікацій

Кафедра телекомунікаційних та радіоелектронних систем

Спеціальність 172 «Телекомунікації та радіотехніка»

Освітньо-професійна програма «Телекомунікаційні системи та мережі»

ЗАТВЕРДЖУЮ

Завідувач кафедри

Роман ОДАРЧЕНКО

“ ” 2022 р.

ЗАВДАННЯ на виконання кваліфікаційної роботи

Ломако Данила Сергійовича

(прізвище, ім'я, по батькові випускника в родовому відмінку)

1. Тема кваліфікаційної роботи: «Система ідентифікації людини за голосом на базі нейронної мережі»

затверджена наказом ректора від «07» вересня 2022 р. №1321/ст

2. Термін виконання роботи: з 05.09.2022 р. по 30.11.2022 р.

3. Вихідні дані до роботи: мовний сигнал, нейронна мережа

4. Зміст пояснювальної записки: аналіз методів розпізнавання мови, способи побудови блоків виділення ознак мовних сигналів, метод виділення ділянок з мовою на основі розподілу локальних екстремумів для ідентифікації людини

5. Перелік обов'язкового графічного (ілюстративного) матеріалу: приклад ділянки дискретного мовного сигналу з шістьма екстремумами, різні типи сигналів і розподілу екстремумів для них, підсистема виділення мови, тестовий сигнал «вперед-назад-шурхіт сторінок-вперед», розмічений за допомогою традиційної VAD-функції, тестовий сигнал «вперед-назад-шурхіт сторінок-вперед», розмічений розробленою системою виділення мови, модель мовотворення "збудження -фільтр", знаходження мел-кепстральних коефіцієнтів

6. Календарний план-графік

№ пор.	Завдання	Термін виконання	Відмітка про виконання
1	Розробити деталізований зміст розділів кваліфікаційної роботи	05.09.2022- 06.09.2022	Виконано
2	Вступ	07.09.2022- 10.09.2022	Виконано
3	Аналіз методів розпізнавання мови	12.09.2022- 05.10.2022	Виконано
4	Способи побудови блоків виділення ознак мовних сигналів	06.10.2022- 15.10.2022	Виконано
5	Метод виділення ділянок з мовою на основі розподілу локальних екстремумів для ідентифікації людини	17.10.2022- 05.11.2022	Виконано
6	Охорона праці	07.11.2022- 12.11.2022	Виконано
7	Охорона навколишнього середовища	14.11.2022- 19.11.2022	Виконано
8	Усунення недоліків та захист кваліфікаційної роботи	21.11.2022- 30.11.2022	Виконано

7. Консультанти з окремих розділів

Розділ	Консультант (посада, П.І.Б.)	Дата, підпис	
		Завдання видав	Завдання прийняв
Охорона праці	к.м.н., проф. Батир ХАЛМУРАДОВ		
Охорона навколиш- нього середовища	к.т.н., доц. Євгеній БОВСУНОВСЬКИЙ		

8. Дата видачі завдання: “22” серпня 2022 р.

Керівник кваліфікаційної роботи _____
(підпис керівника)

Володимир КЛИМЧУК
(П.І.Б.)

Завдання прийняв до виконання _____
(підпис випускника)

Данило ЛОМАКО
(П.І.Б.)

РЕФЕРАТ

Кваліфікаційна робота «Система ідентифікації людини за голосом на базі нейронної мережі» містить 79 сторінок, 24 рисунки, 5 таблиць, 56 використаних джерел.

РОЗПІЗНАВАННЯ МОВИ, НЕЙРОННА МЕРЕЖА, КЕПСТРАЛЬНІ КОЕФІЦІЄНТИ, ЛОКАЛЬНІ ЕКСТРЕМУМИ.

Мета роботи полягає у розробці методів розпізнавання мови на базі нейронних мереж.

Об'єктом дослідження – є ідентифікація людини за голосом на базі нейронної мережі.

Предметом дослідження – є виділення ділянок з мовою на основі розподілу локальних екстремумів.

Наукова новизна. Набув подальшого розвитку метод виділення мови, заснований на аналізі розподілу локальних екстремумів вхідного сигналу, що дозволяє знаходити межі мовного сигналу в умовах присутності різних шумів.

Практичне значення отриманих результатів полягає у розпізнаванні символів (літер українського алфавіту) на основі методу порівняння з прототипом за допомогою нейронної мережі. Розроблено систему виділення мови, що включає блоки нормалізації, знаходження обгинальної та розподілу екстремумів, вирішальної логіки.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	7
ВСТУП	8
РОЗДІЛ 1. АНАЛІЗ МЕТОДІВ РОЗПІЗНАВАННЯ МОВИ	11
1.1. Мовний сигнал і його опис	11
1.2. Загальна структура системи автоматичного розпізнавання мови	14
1.3. Вибір типу середовища для реалізації блоку розпізнавання мови	15
РОЗДІЛ 2. СПОСОБИ ПОБУДОВИ БЛОКІВ ВИДІЛЕННЯ ОЗНАК МОВНИХ СИГНАЛІВ	19
2.1. Побудова блоку виділення ознак. Вибір методу виділення ознак	19
2.2. Способи побудови блоку розпізнавання. Вибір методу розпізнавання	22
2.3. Нормалізація вхідного сигналу	27
РОЗДІЛ 3. МЕТОД ВИДІЛЕННЯ ДІЛЯНОК З МОВОЮ НА ОСНОВІ РОЗПОДІЛУ ЛОКАЛЬНИХ ЕКСТРЕМУМІВ ДЛЯ ІДЕНТИФІКАЦІЇ ЛЮДИНИ	31
3.1. Знаходження значень короткочасної енергії і частоти проходів через нуль	31
3.2. Метод виділення ділянок з мовою на основі розподілу локальних екстремумів	33
3.3. Виділення ознак мовного сигналу	40
3.4. Векторне квантування	47
РОЗДІЛ 4. ОХОРОНА ПРАЦІ	52
РОЗДІЛ 5. ОХОРОНА НАВКОЛИШНЬОГО СЕРЕДОВИЩА	65
ВИСНОВКИ	73
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	74

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАЧЕНЬ

BSI – відношення сигнал/шум.

ДПФ – дискретне перетворення Фур'є.

ЗДПФ – зворотне Дискретне перетворення Фур'є.

НМ – нейронна мережа

SARF – система автоматичного розпізнавання мови.

СЦФ – системи цифрової фільтрації.

DTW – Dynamic Time Warping.

GMM – Gaussian Mixture Models.

LPC – Linear Predictive Codes.

MFCC – Mel-Frequency Cepstral Coefficients.

PLP – Perceptual Linear Prediction.

STE – Short-Time Energy.

VAD – Voice Activation Detection.

ZCR – Zero-Crossing Frequency.

ВСТУП

Актуальність теми. Питання людино-машинної взаємодії є одними з найважливіших під час створення нових комп'ютерів. Найбільш ефективними засобами взаємодії людини з машиною були б ті, які є природними для неї: через візуальні образи та мовлення. Створення мовних інтерфейсів могло б знайти застосування в системах різного призначення [1-40]: голосове управління для людей з обмеженими можливостями, надійне управління бойовими машинами, «розуміючими» тільки голос командира, автовідповідачі, обробні в автоматичному режимі сотні тисяч дзвінків на добу (наприклад, у системі продажу авіаквитків) тощо. При цьому, мовний інтерфейс повинен включати у собі два компоненти: систему автоматичного розпізнавання промови прийому мовного сигналу і перетворення їх у текст чи команду, і систему синтезу промови, виконує протилежну функцію – конвертацію повідомлення від машини на мову.

Однак, незважаючи на стрімко зростаючі обчислювальні потужності, створення систем розпізнавання мови залишається надзвичайно складною проблемою. Це зумовлюється як її міждисциплінарним характером (необхідно мати знання у філології, лінгвістиці, цифровій обробці сигналів, акустиці, статистиці, розпізнаванні образів і т.д.), і високої обчислювальної складності розроблених алгоритмів. Останнє накладає суттєві обмеження на системи автоматичного розпізнавання мови – обсяг оброблюваного словника, швидкість отримання відповіді та її точність. Не можна також не згадати про те, що можливості подальшого збільшення швидкодії ЕОМ за рахунок удосконалення інтегральної технології рано чи пізно будуть вичерпані, а зростаюча різниця між швидкодіями пам'яті та процесора лише посилює проблему.

Існують сфери застосування систем автоматичного розпізнавання мови, де описані проблеми виявляються особливо гостро через жорстко обмежені обчислювальні ресурси, наприклад, на мобільних пристроях [2]. Виробники мобільних телефонів і планшетів знайшли вихід у перенесенні ресурсомістких обчислень з пристроїв кори-

стувачів на сервери у хмарі, де фактично і проводиться розпізнавання. Додаток користувача тільки надсилає туди мовні запити і приймає відповіді, використовуючи підключення до інтернету. За цією схемою успішно працюють системи Siri від Apple і Google Voice Search від Google [3, 4]. Однак, для такої реалізації необхідні певні умови, наприклад, безперервний доступ до інтернету, які в ряді випадків недосяжні, і потрібно створити компактний та надійний самостійний пристрій, що експлуатує лише доступні «на місці» обчислювальні потужності. Описані проблеми виникають при створенні інтелектуальних механізмів як у військовій сфері, так і в цивільній. Прикладом таких пристроїв може бути робот REX, розроблений ізраїльським концерном Israel Aerospace Industries [5]. REX призначений для транспортування боєприпасів, продуктів харчування та іншої амуніції, що дозволяє розвантажити солдата. При цьому робот здатний слідувати за людиною, а управляється він повністю голосовими командами. Іншим прикладом активного використання технологій розпізнавання мовлення у бойових комплексах є впровадження модулів голосового управління (або прямого голосового введення – Direct Voice Control) у кокпіти сучасних винищувачів, таких як Eurofighter Typhoon 1 Dassault Rafale 2 JAS 39 Gripen [6]. Це дозволило значно розвантажити пілота для того, щоб він міг зосередитись тільки на виконанні завдання. У невійськовій сфері розпізнавання мови широко впроваджується в автомобілебудуванні (наприклад, BMW, Ford), коли частина функціоналу машини, для якого помилка розпізнавання не призведе до аварійних ситуацій (клімат-контроль, навігація, мультимедіа та ін.), контролюється за допомогою голосу. Як і у разі застосування голосового управління у військових літаках, ця технологія дала можливість зняти частину навантаження з водія, щоб він міг зосередити увагу лише на дорозі. Нарешті, слід зазначити актуальність реалізації мовного інтерфейсу людей з обмеженими фізичними можливостями, наприклад, в інвалідних кріслах [7].

Мета роботи полягає у розробці методів розпізнавання мови на базі нейронних мереж.

Для досягнення поставленої мети вирішуються такі наукові завдання.

- вибір методу виділення та попередньої обробки мови, вилучення ознак;
- програмна реалізація виділення мови та її попередньої обробки;

- вибір методу розпізнавання мови;

Об'єктом дослідження – є ідентифікація людини за голосом на базі нейронної мережі.

Предметом дослідження – є виділення ділянок з мовою на основі розподілу локальних екстремумів.

Наукова новизна. Набув подальшого розвитку метод виділення мови, заснований на аналізі розподілу локальних екстремумів вхідного сигналу, що дозволяє знаходити межі мовного сигналу в умовах присутності різних шумів.

Практичне значення отриманих результатів полягає у розпізнаванні символів (літер українського алфавіту) на основі методу порівняння з прототипом за допомогою нейронної мережі. Розроблено систему виділення мови, що включає блоки нормалізації, знаходження обгинальної та розподілу екстремумів, вирішальної логіки.

РОЗДІЛ 1

АНАЛІЗ МЕТОДІВ РОЗПІЗНАВАННЯ МОВИ

1.1. Мовний сигнал і його опис

Мова – це форма спілкування людей у вигляді мовних конструкцій, створюваних з урахуванням певних правил [22, 23]. Якщо як провідне середовище для передачі інформації (спілкування) використовується повітря, то виходить усне мовлення - звукове коливання, яке характеризується частотою та амплітудою. Мова є носієм інформації, використовуваний людиною для передачі повідомлень. За фізичною природою це акустичний сигнал, що безперервно змінюється у часі. Бажаючи підкреслити природу цього сигналу і відрізнити його від сигналів інших типів, у технічній літературі називають мовним сигналом. Далі терміни «мова», «мовний сигнал» і «Усна мова» будуть вживатися як синоніми, за винятком випадків, коли потрібно виділити значення окремого терміна.

Більшість сигналів (мовних зокрема) мають аналогову природу, для обробки їх на цифрових комп'ютерах вони перетворюються на дискретні сигнали у вигляді аналого-цифрового перетворення (АЦП). За допомогою цієї процедури отримують набір відліків $s[n]$ – знятих у моменти Δt n миттєвих значень безперервного сигналу, які вже позбавлені фізичної природи, а їх максимальне та мінімальне значення задається розрядністю АЦП. Наприклад, якщо розрядність АЦП дорівнює 2 байтам, то всі значення у відліках укладаються у проміжок 2^{16-1} , $2^{16-1} - 1$. При цьому найважливішим параметром перетворення є частота дискретизації, що визначає, скільки миттєвих значень безперервного сигналу (відліків) буде збережено за секунду. Частота дискретизації – величина, обернена до кроку дискретизації Δt . За теоремою Котельникова, з дискретного сигналу можна відновити без втрат такий аналоговий сигнал, верхня частота f_h спектра якого вдвічі менше частоти дискретизації f_s

$$f_s > 2 \cdot f_h, \quad (1.1)$$

Для опису та перетворення дискретних сигналів застосовні засоби цифрової обробки сигналів (ЦОС). Найважливішою процедурою ЦГЗ є дискретне перетворення Фур'є (ДПФ) [24]:

$$S[m] = \sum_{i=1}^N s[n] \cdot e^{\frac{j2\pi nm}{N}}, m = 1, \dots, N, \quad (1.2)$$

де N - кількість відліків, по яким будується ДПФ;

j - уявна одиниця.

ДПФ дозволяє перейти з часової області в частотну, тобто, розкласти $s[n]$ на набір гармонік та знайти залежність амплітуди (енергії) гармоніки від її частоти. на рис. 1.1 представлений ділянка мовного сигналу з голосним звуком "а" у часовій області. При цьому абстрагуватися від розрядності АЦП, відліки оцифрованого сигналу прийнято зображати у відносних величинах: або у частках від максимального значення (для 2 байт це $s[n] \cdot 2^{16-1}$), або децибелах. У цьому роботі використовується перший спосіб представлення. Для знаходження ДПФ було виділено ділянку розміром $N=1024$ відліку; результат представлений рис. 1.2. При цьому по горизонталі відкладається частота гармонік, а по вертикалі - $|S[m]|$ що є амплітудою гармоніки.

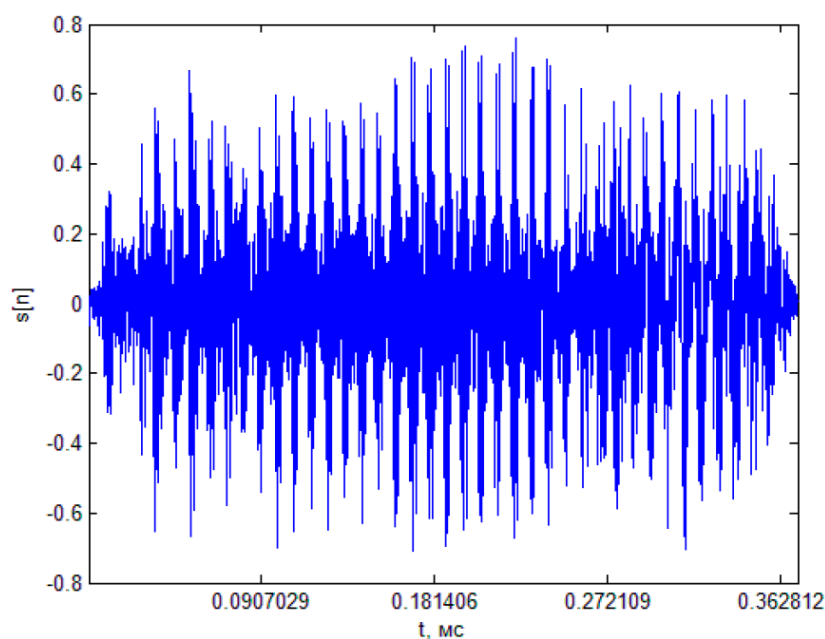


Рис. 1.1. Ділянка мови з голосним звуком "а"

Мова являється нестационарним сигналом, тобто. його характеристики змінюються у часі. Можна наочно зобразити ці зміни, побудувавши графіки модулів ДПФ для фрагментів (фреймів) мовного сигналу, що йдуть поспіль. Зображення, що вийшло, називається спектрограмою (рисунок 1.3.б). На рисунках 1.2 і 1.3 видно, що найбільше енергії несуть частоти до 8 КГц. Тому при оцифровці мовного сигналу типовим вибором частоти дискретизації є 16 КГц.

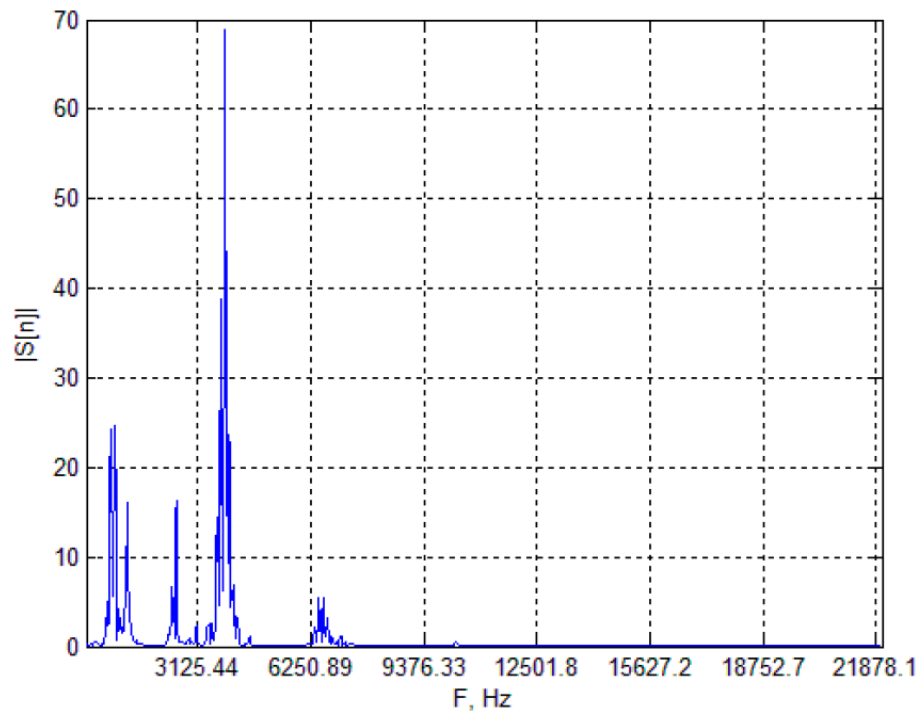


Рис. 1.2. ДПФ для ділянки мовного сигналу з голосним звуком "а"

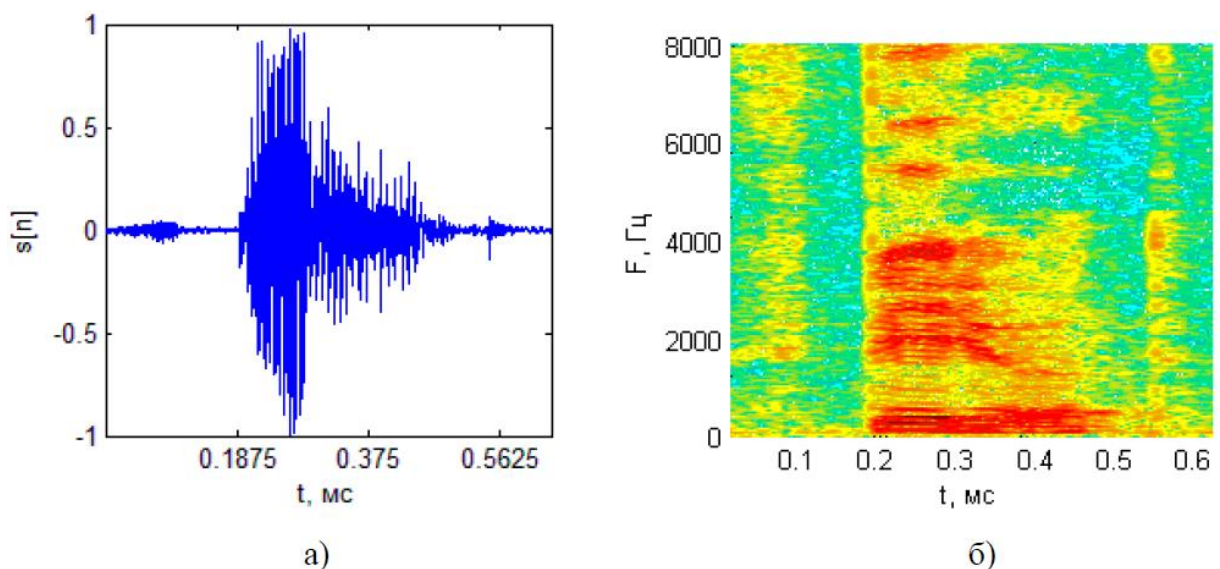


Рис. 1.3. Осцилограма слова "Вперед" і його спектрограма

Як слова в письмовій мові утворюються з кінцевого набору символів – алфавіту мови, так і усне мовлення при всій її варіативності включає обмежений набір звукових «літер». Мінімальною сенсорною одиницею мови є фонема [22, 23, 25]. В українській мові 38 основних фонем, з яких 6 голосних та 32 приголосних. Подальшої єдиної класифікації фонем, на жаль, не існує, тому на рис. 1.4 наведено один із поєднаних варіантів, що містить класи, що перетинаються, наприклад, дзвінки (voiced) і фрикативні, глухі (unvoiced) і вибухові і т.п.

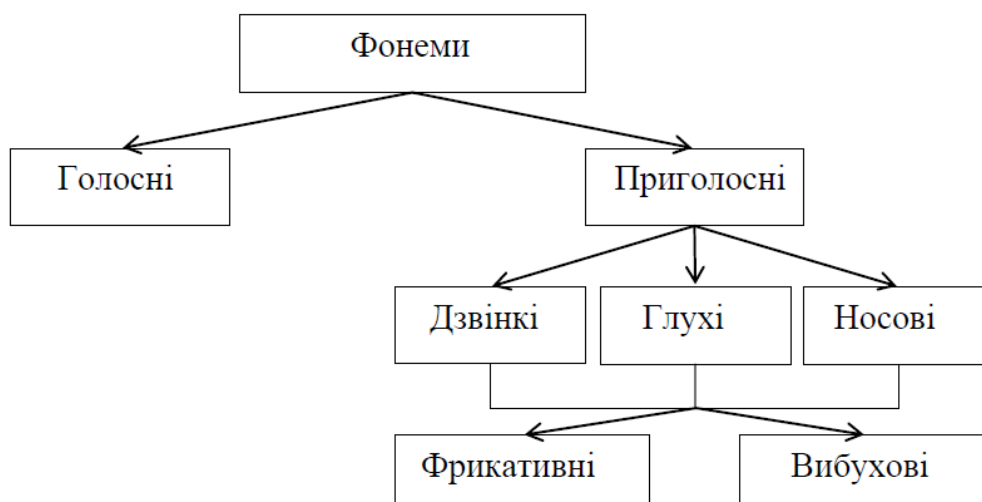


Рис. 1.4. Класифікація фонем української мови

1.2. Загальна структура системи автоматичного розпізнавання мови

У роботі системи автоматичного розпізнавання мови (САРМ) виділяють три етапи: виділення ознак, навчання та розпізнавання (рис. 1.5). На першому етапі з вихідного сигналу одержують вектор ознак – стислий опис мовного сигналу, в якому присутня тільки значуща для розпізнавання інформація. Для цього використовуються методи, що працюють як у частотній області (крейда-кепстральні коефіцієнти, коефіцієнти лінійного передбачення), так і в часовій (наприклад, на короткочасному значенні енергії), при цьому проблема подання мови не вирішена до кінця і дослідження ведуться в тому числі і авторами цієї роботи [27, 28]. Послідовність векторів ознак довжиною T називають акустичною або послідовністю, що спостерігається $\mathbf{O} = (o_1, o_2, \dots, o_T)$. За допомогою цієї послідовності людина передає ланцюжок слів $\mathbf{W} = (w_1$

, w_2, \dots, w_N). Завдання розпізнавання мови ставиться в такий спосіб: необхідно знайти ланцюжок слів W , який відповідає акустичній послідовності O [3, 4, 5].

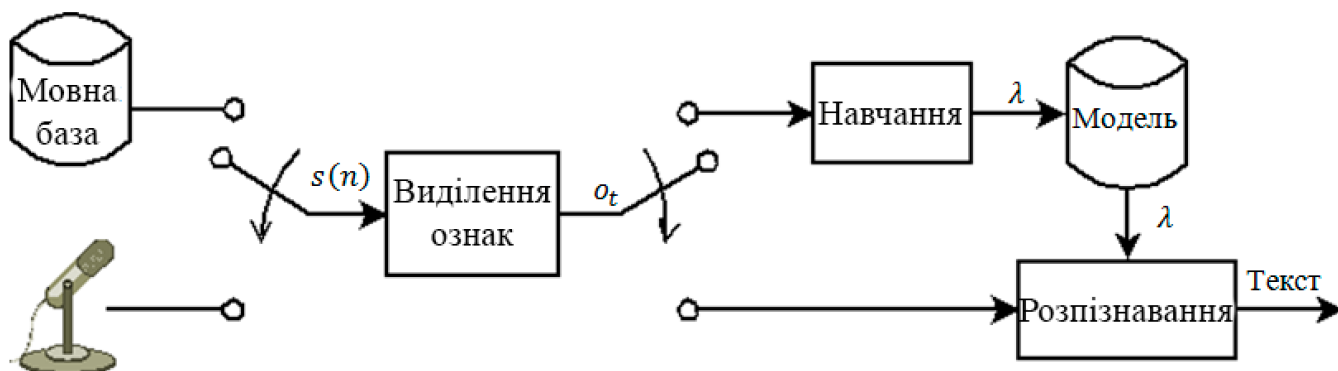


Рис. 1.5. Загальна схема системи автоматичного розпізнавання мови

Для вирішення цього завдання на етапі навчання складається модель λ яка здатна породжувати всі можливі послідовності O для всіх можливих W . Нехай функція $h(W, \lambda)$ повертає всі можливі O тільки для заданої W . Тоді розпізнаванням буде знаходження такого ланцюжка слів W^* , який, згідно моделі λ , породить акустичну послідовність, найбільш близьку до аналізованої O :

$$W^* = \text{ArgMin}_{W \in w} d(h(W, \lambda), O), \quad (1.3)$$

де $d O', O$ – дистанція між O' і O .

Таким чином, в ідеалі потрібно перевірити всі ланцюжки слів, що, природно, недосяжно на практиці. Для полегшення цього завдання вводять різні обмеження за допомогою граматики мови або вирішується вужче завдання, наприклад, розпізнавання лише ізольованих слів.

1.3. Вибір типу середовища для реалізації блоку розпізнавання мови

Комп'ютерна пам'ять виконує функції запису, зчитування та зберігання інформації в ЕОМ. За способом доступу до інформації, що зберігається, розрізняють пам'ять з адресним доступом і асоціативним. При адресному доступі необхідно вказати

місцезнаходження інформації в середовищі, в той час як при асоціативному доступі пошук інформації здійснюється за її змістом [4, 5, 6]. Запам'ятовуюче середовище, в якому доступ до інформації базується на механізмі асоціації, називається асоціативною середовищем. У ній, крім трьох основних функцій пам'яті, можливі також пошук і обробка інформації без вилучення її із накопичувача.

Асоціативний спосіб доступу забезпечує [15]:

- практично одночасний доступ до всієї інформації;
- відносну незалежність часу пошуку інформації від ємності пам'яті;
- внесення елементів обробки інформації безпосередньо у процес самого доступу;
- обробку інформації безпосередньо в середовищі її зберігання.

Першим і найпростішим асоціативним запам'ятовуючим пристроєм (АЗП) можна вважати однокоординатне АЗП, в якому асоціативний пошук був можливий тільки по одній координаті. Наступним кроком у розвитку асоціативних середовищ була розробка багатокоординатного АЗП, який здійснює пошук за кількома координатами.

Подальша розробка асоціативної пам'яті йшла шляхом ускладнення логіко-запам'ятного середовища: внесення локальних зв'язків, реалізації мікрооперацій обробки інформації в осередку середовища. Такими ознаками має асоціативне середовище з локальними зв'язками [14]. Було реалізовано різні алгоритми обробки зображень (Заповнення порожнин, фільтрація [17]), арифметичні операції цілими числами [18, 19].

Нарешті, останнім етапом у розвитку асоціативних середовищ стала відмова від ідеології однорідних клітинних автоматів, за якою всі осередки середовища виконували ту саму операцію за такт, що задається кодом мікрооперації. Крім цього, вирішили відмовитися від жорстких локальних зв'язків для всіх осередків, зв'язавши лише ті, які необхідні для роботи алгоритму обробки інформації, а в самих осередках з'явилася можливість реалізовувати різний функціонал. Таке середовище називається асоціативним осциляторним середовищем [37, 40].

Асоціативне осциляторне середовище (АОС) побудовано за принципом неоднорідних клітинних автоматів, де кожен осередок має свій закон функціонування. У термінах АОС осередки називаються клітинними ансамблями, а закон їх функціонування вибирається на етапі розробки алгоритму обробки інформації [27]. Основним поняттям АОС є спайк – інформаційний вплив, що передається між сусідніми клітинами локальних зв'язків. Спайки ніколи не стоять на місці, але весь час переміщуються між клітинними ансамблями середовища, які по-різному їх обробляють. Отже, АОС "живе" у часі, у ній постійно відбувається пульсація спайків, які з часом утворюють послідовності. Різні клітинні ансамблі середовища обробляють ці послідовності. Це дозволяє проводити потокову обробку даних.

Час для АОС дискретний, так що спайки переміщуються від клітини до клітини між тактами, і спостереження за тим, як АОС обробляє інформацію, здійснюється за тактами. Через кожен клітинний ансамбль середовища за проміжок N тактів проходить різна кількість спайків q . При цьому можна знайти інтенсивність потоку спайків P_N на виході із клітинного ансамблю:

$$P_N = \frac{q}{N}, \quad (1.4)$$

При досить великій кількості аналізованих тактів інтенсивність P_N наближається до ймовірності P проходження спайка через зв'язок у даному такті [57]:

$$P = \lim_{N \rightarrow \infty} P_N, \quad (1.5)$$

Таким чином, клітинні ансамблі можна описувати двояко: по-перше, з точки зору залежності значення на виході наступного такту від значень на вході в поточному такті; по-друге, через залежність інтенсивності потоку спайків на виході від інтенсивностей потоків спайки на входах.

Найпростішим клітинним ансамблем є провідник, який не змінює вхідну інтенсивність потоку спайків. Якщо побудувати замкнутий ланцюжок провідників довжиною q , то вийде замкнутий осцилятор. Цей клітинний ансамбль не впливає на оточуючі клітини, а число спайків, що курсує в його ланцюжку, називають зарядом осцилятора.

Для реалізації алгоритмів розпізнавання мови було обрано асоціативне осциляторне середовище нейронної мережі, тому що:

- в ньому можлива організація поточкових і конвеєрних обчислень;
- осередки середовища можна гнучко з'єднувати один з одним, не обмежуючись матричною структурою;
- за один такт кожна комірка обробляє інформацію (вхідні спайки) згідно з закладеним у неї законом функціонування.

ВИСНОВКИ ДО РОЗДІЛУ 1

Перший розділ був присвячений огляду завдання розпізнавання мови та шляхів його вирішення. У цьому розділі:

1. розглянуто об'єкт дослідження - мова, а також способи її подання у цифровій формі за допомогою осцилограм та спектрограм. Наведено класифікацію фонем української мови;
2. поставлено завдання розпізнавання мови та описана загальна структура системи автоматичного розпізнавання мовлення;
3. наведено класифікацію та опис різних підходів складання векторів ознак, що працюють як у частотній області (коефіцієнти лінійного передбачення, перцепційні коефіцієнти лінійного передбачення, крейдяні коефіцієнти), так і в часовій (частота проходів через нуль, короткочасна енергія сигналу). Серед проаналізованих методів знаходження векторів ознак для використання в даній роботі було обрано метод крейдяних коефіцієнтів.

РОЗДІЛ 2

СПОСОБИ ПОБУДОВИ БЛОКІВ ВИДІЛЕННЯ ОЗНАК МОВНИХ СИГНАЛІВ

2.1. Побудова блоку виділення ознак. Вибір методу виділення ознак

Завдання блоку виділення ознак – скласти ланцюжок векторів ознак $\mathbf{O} = (o_1, o_2, \dots, o_T)$ вихідного сигналу. Як було зазначено вище, мова – нестационарний сигнал. Проте, через властивості інертності мовного тракту в межах досить короткого проміжку часу від 10 до 40 мс його характеристики не змінюються, тобто, його можна вважати стаціонарним [9, 10, 24, 25]. Тому блок виділення ознак сканує вхідний сигнал короткочасним ковзним вікном, в межах якого і складається один вектор ознак o_t (Рис. 2.1). Ці вікна можуть перетинатися.

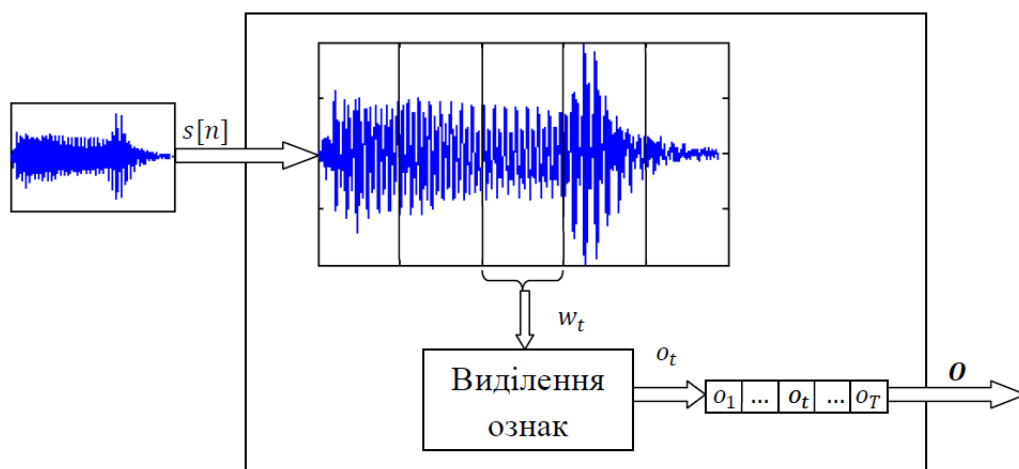


Рис. 2.1. Виділення ознак. Розбиття вхідного сигналу на ланцюжок вікон

Дослідження показали, що найкраще мова є ознаками, отриманими в частотній області [31, 32]. До таких ознак відносяться коефіцієнти лінійного передбачення (Linear Predictive Codes – LPC), перцепційні коефіцієнти лінійного передбачення (Perceptual Linear Prediction – PLP), мел-коефіцієнти (Mel-Frequency Cepstral Coefficients – MFCC) [21, 22, 23, 26]. Ці три ознаки ґрунтуються на акустичній моделі мовлення, згідно з якою мовний сигнал можна подати у вигляді

сигналу на виході лінійної системи з параметрами, що змінюються в часі, збуджуваної квазіперіодичними імпульсами (при виголошенні вокалізованого звуку) або випадковим шумом (при невокалізованих звуках) [32].

3. Оцінка параметрів цієї лінійної системи є завдання методів знаходження ознак у частотній області. Основною ідеєю методів лінійного передбачення є можливість апроксимації поточного відліку мовного сигналу за допомогою лінійної комбінації попередніх відліків [32]:

$$s(n) = \sum_{k=1}^{\rho} \alpha_k s(n - k), \quad (2.1)$$

де $\{\alpha_k\}$ - Коефіцієнти лінійного передбачення. Похибка передбачення визначається як:

$$e(n) = s(n) - \sum_{k=1}^{\rho} \alpha_k s(n - k), \quad (2.2)$$

Помилка $e(n)$ є сигналом на виході системи з передатною функцією $A(z)$:

$$A(z) = 1 - \sum_{k=1}^{\rho} \alpha_k z^{-k}, \quad (2.3)$$

Передатна функція лінійної системи, в яку входять збудження та мовний сигнал, описується так:

$$H(z) = \frac{G}{1 - \sum_{k=1}^{\rho} \alpha_k z^{-k}}, \quad (2.4)$$

де G – коефіцієнт посилення збудження. Таким чином, визначення параметрів $\{\alpha_k\}$ дозволить оцінити спектральні властивості мовного сигналу. Складання вектора ознак з коефіцієнтів лінійного передбачення є метод LPC.

PLP відрізняється від описаного вище підходу тим, що намагається врахувати особливості сприйняття різних частот людиною [33]. Встановлено, що роздільна здатність людського вуха нерівномірно розподілено за спектром: в області низьких частот воно вище, ніж в області високих. Цей ефект описується за допомогою психоакустичної барк-шкали. Тому перед знаходженням коефіцієнтів лінійного передбачення мовний сигнал пропускається через фільтри, смуги пропускання яких змінюються відповідно до барк-шкали.

Методом MFCC прагнуть відокремити сигнал збудження від параметрів мовного тракту, використовуючи гомоморфні перетворення [32, 33, 34]. Для цього за допомогою ДПФ переходять у частотну область, де обчислюють логарифм спектра вхідного сигналу, а потім виконують ЗДПФ або дискретне косинусне перетворення. Як і в PLP, для моделювання сприйняття мови людиною перед знаходженням логарифму на спектр вхідного сигналу накладають набір мел-фільтрів, смуги пропускання яких вибираються за мел-шкалою, знижуючи дозвіл у бік високих частот (аналог барк-шкали).

Перевагою MFCC порівняно з LPC і PLP є простота реалізації за аналогічних показників якості розпізнавання. Швидкодія методу обумовлюється наявністю ефективною процедури знаходження ДПФ та ЗДПФ – швидкого перетворення Фур'є (ШПФ). Аналіз стану області розпізнавання мови показав, що сьогодні MFCC застосовується найбільш широко. Ці фактори дозволяють вибрати дрібні коефіцієнти для використання в даній роботі.

Існує також група ознак у часовій області. До них відносяться частота проходів через нуль (Zero-Crossing Frequency – ZCR) та короткочасна енергія сигналу (Short-Time Energy – STE) [32, 38]. Перший метод дозволяє грубо та швидко оцінити спектральні характеристики мовного сигналу. За допомогою підрахунку ZCR можна відрізнити вокалізовані звуки від невокалізованих, так як високі частоти наводять до біль-

шого числа переходів через нуль, а низькі – до малого. Короткочасна енергія дозволяє виявляти зміни гучності сигналу, що також може бути використане при класифікації звуків мови. Ознаки цієї групи, що використовується або як супутні описаним вище показники частотної області, або завдання виділення мови.

2.2. Способи побудови блоку розпізнавання. Вибір методу розпізнавання

Як було зазначено вище, на етапі розпізнавання робота ведеться із послідовністю векторів ознак довжиною T $O = (o_1, o_2, \dots, o_T)$, за допомогою якого передається ланцюжок слів $W = (w_1, w_2, \dots, w_N)$. Використовуючи загальну для розпізнавання образів термінологію, ланцюжок O називається образом – областю у просторі ознак [39, 40]. Для розпізнавання аналізованої O' необхідно за допомогою моделі мови λ , в якій встановлений зв'язок між усіма можливими O і всіма можливими W , знайти такий ланцюжок слів W^* , для якого λ породить послідовність ознак, найбільш близьку до O' .

Головна проблема розпізнавання мови полягає в тому, як скласти модель. Вирізняють дві процедури роботи САРМ:

1. навчання, коли налаштовуються параметри моделі на навчальній вибірці, що є безліччю пар O', W' . Чим ширша навчальна вибірка, тим адекватніше вийде модель;
2. розпізнавання, коли перевіряються всі ланцюжки слів W і вибирається той, чия акустична послідовність W, λ найближче до аналізованої.

Якби була можливість скласти таку модель, яка б на етапі навчання змогла б включити "знання" про всі можливі пари O', W' , то завдання розпізнавання мови було б, у принципі, вирішено. Однак, на практиці ми не маємо такого великого запасу навчальної вибірки, а моделі здатні вмістити в себе лише обмежений набір "відомостей". Крім цього питання про те, що вибрати як мовленнєвий образ, залишається відкритим – це можуть бути окремі фонемі, слова і т.п.

Аналіз літератури дозволив виділити три групи методів розпізнавання мови:

1. Методи, засновані на порівнянні за зразком.

Для кожного слова складається модель-еталон вимови \mathbf{O}' , щоб на етапі розпізнавання вибрати ту модель, еталон якої найближчий до акустичної послідовності \mathbf{O} . Головна проблема методів цієї групи у тому, що мовні образи сильно варіюються за тривалістю, отже необхідний спосіб порівнювати образи різної довжини. Єдиний представник групи – метод динамічного вирівнювання часу (Dynamic Time Warping - DTW) [29, 40]. У ньому проблема різниці довжини зразка і розглянутого образу вирішується наступним шляхом: складається матриця C розміром $M \times N$, де N

- довжина зразка, а M - довжина даної послідовності:

$$\begin{aligned} C_{1,1} &= D_{1,1} \\ C_{i,1} &= D_{i,1} + C_{i-1,1}, \quad i = 2, \dots, M \\ C_{1,j} &= D_{1,j} + C_{1,j-1}, \quad j = 2, \dots, N \\ C_{i,j} &= D_{i,j} + \min(C_{i-1,j}, C_{i-1,j-1}, C_{i,j-1}), \quad i = 2, \dots, M; \quad j = 2, \dots, N, \end{aligned} \quad (2.5)$$

де $D_{i,j}$ - відстань між i -м компонентом \mathbf{O}' і j -м компонентом \mathbf{O} , яка може обчислюватися різними способами, наприклад, як евклідова відстань або манхеттенська відстань:

$$D_{i,j} = \sqrt{o_i^2 + o_j^2}, \quad (2.6)$$

$$D_{i,j} = |o_i - o_j|, \quad (2.7)$$

Розглянутий метод, фактично, представляє собою Рішення завдання пошуку найкоротшого шляху на графі методом динамічного програмування, де початковий вузол «розташований» у лівому нижньому кутку сітки, а кінцевий – у верхньому правому з координатами (o'_N, o_M) (рис. 2.2).

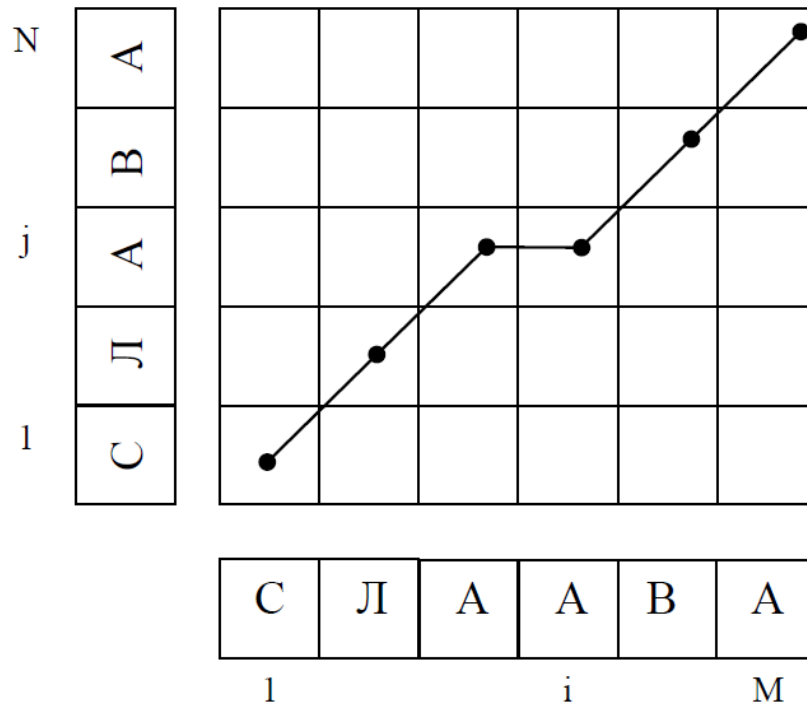


Рис. 2.2. Приклад порівняння образу з еталоном методом динамічного вирівнювання часу

Недоліком методу динамічного вирівнювання часу є труднощі, що виникають під час упорядкування еталона, викликані сильною варіативністю промови. Крім цього за допомогою DTW складно організувати розпізнавання злитого мовлення.

2. Методи, виконуючі побудову вирішальних функцій.

Суть методів цієї групи полягає у знаходженні такої функції, яка б за вхідним образом визначала його приналежність до того чи іншого класу. Для цього найчастіше використовуються штучні нейронні мережі (Artificial Neural Networks - ANN). Одношаровий перцептрон дозволяє побудувати розділяючі площини для лінійно-роздільних класів. Мінімальна обчислювальна одиниця перцептрону – j -й штучний нейрон, що визначається як лінійна функція з вагами $\mathbf{w}_j = (w_{0j}, w_{1j}, \dots, w_{Nj})$ від N аргументів, на які подається мовленнєвий образ $\mathbf{O} = (o_1, o_2, \dots, o_N)$ [26, 32] (Рис. 2.3):

$$y_j = w_{0j} + \sum_{i=1}^N w_{ij} o_i = \mathbf{w}_j \mathbf{x}^T = g_j(\mathbf{x}) \quad , \quad (2.8)$$

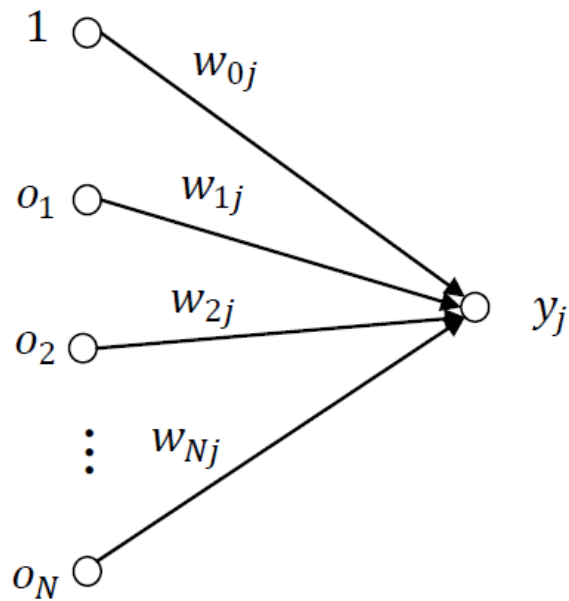


Рис. 2.3. Схема штучного нейрона одношарового перцептрону.

де $\mathbf{x} = (1, o_1, o_2, \dots, o_N)$. Кожен вихідний нейрон зіставляється одним із s класів $(\omega_1, \omega_2, \dots, \omega_s)$. У матричному вигляді це можна записати за допомогою матриці вагових коефіцієнтів перцептрона $\mathbf{W} = (\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_s^T)$ і вектора вихідних значень $\mathbf{y} = (y_1, y_2, \dots, y_s)^T$:

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) = \mathbf{W}^T \mathbf{x}, \quad (2.9)$$

Тоді для розпізнавання необхідно буде знайти той нейрон k , чиє вихідне значення буде максимальне:

$$k = \text{ArgMax}_{1 \leq i \leq s} g_i(\mathbf{x}), \quad (2.10)$$

У процесі навчання перцептрона на навчальній вибірці налаштовують матрицю вагових коефіцієнтів таким чином, щоб мінімізувати помилку його відповіді.

Багатошаровий перцептрон включає один і більше прихованих шарів і дозволяє будувати нелінійні розділяючі функції. Для навчання можна використовувати алгоритм зворотного поширення помилки [26, 32].

Широкий інтерес до нейронних мереж викликаний їхньою здатністю до виділення характерних рис образу та узагальнення [26]. Також плюсом можна вважати те, що штучний нейрон досить просто реалізувати апаратно, і, поєднавши нейрони в мережу потрібної конфігурації, можна побудувати нейрокомп'ютер. Мінусом ANN, також, як і DTW, є складність реалізації розпізнавання злитого мовлення. Для подолання цього недоліку було запропоновано такі архітектури ANN [33]: нейронні мережі з часовою затримкою (time-delay neural networks) та рекурентні нейронні мережі (recurrent neural networks). Однак, такі мережі не набули широкого поширення в галузі розпізнавання мови.

3. Приховані Марківські моделі.

Головною проблемою попередніх методів є їх обмежені можливості в обліку часу, наприклад, для організації розпізнавання злитого мовлення. У 1970-ті роки виникла ідея описати мовний сигнал як стохастичний процес, вбудувавши таким чином час [34, 35]. На сьогоднішній день найбільш поширеним підходом до розв'язання задачі розпізнавання мовлення є використання Прихованих Марківських Моделей (ПММ) [36, 37, 38, 39, 40]. Короткий історичний нарис про основні досягнення у всіх галузях АСР (розробка інфраструктури, представлення знань, моделі та алгоритми, алгоритми пошуку та необхідні метадані) представлений у роботі [31].

За допомогою ланцюжка станів ПММ моделюють фонему мови, які об'єднують у слова. Найбільш адекватною вважається модель фонему із трьох станів: початкового, середнього та кінцевого. Також зазвичай виділяють окремий стан під тишу та неінформативні звуки, наприклад, вдихи та видихи. При цьому вихідні можливості моделюються за допомогою моделей Гаусових сумішей (Gaussian Mixture Models – GMM).

Складають або окремі ПММ для кожного слова словника, що розпізнається, або одну велику ПММ, що об'єднує слова в пропозиції і більші структури. У першому випадку розпізнавання можна виконати за допомогою алгоритму прямого ходу (forward algorithm), знайшовши таку ПММ, яка здатна породити послідовність, що розглядається, з найбільшою ймовірністю [38]. У другому випадку, використовуючи алгоритм Вітербі, знаходять найбільш ймовірний ланцюжок станів,

через які повинна пройти ПММ для породження послідовності, що розглядається. Другий підхід використовується частіше, тому що з його допомогою можна розпізнавати зливу мову.

В даний час ведуться роботи з гібридних архітектур на основі ПММ, в яких моделювання вихідних ймовірностей здійснюється за допомогою глибоких нейронних мереж (deep neural networks) [22] та мереж глибокої довіри (deep belief networks) [23]. Хоча такі підходи суттєво ускладнюють аналіз, навчання та тестування розпізнавача, вони дозволяють досягти більш високої порівняно з GMM точності розпізнавання мови з великим словником (близько 65 000 слів).

Перевагою ПММ перед іншими методами є природне вбудовування часу в модель λ , що дозволяє врахувати варіативність вимов за довжиною та швидкістю, а також перейти до розпізнавання зливої мови. Крім того, розроблено ефективні алгоритми СММ, які мають потенціал для розпаралелювання, чим користуються фахівці при апаратній реалізації [11, 14].

2.3. Нормалізація вхідного сигналу

При записі мови на амплітуду звукового сигналу впливає низка чинників: гучність голосу диктора, його віддаленість від мікрофона тощо. Перелічені фактори призводять до великої варіативності гучності мовного сигналу [31]. Особливо сильне це явище помітне при використанні різномірної звукозаписної апаратури. Для усунення розкиду гучності застосовується процедура нормалізації амплітуди. З допомогою цього прийому амплітуда сигналу полягає в межах $\left[\frac{\Delta}{2}, -\frac{\Delta}{2}\right]$ (Рис. 2.4). Отримання відліків нормалізованого сигналу $s[n]$ виконується за наступною формулою:

$$\hat{s}[n] = \frac{\Delta}{\max_m |s[m]|} \cdot s[n], \quad (2.11)$$

де Δ – ширина смуги нормалізації, симетричної щодо осі абсцис (наприклад, на рис. 2.4. $\Delta=1$).

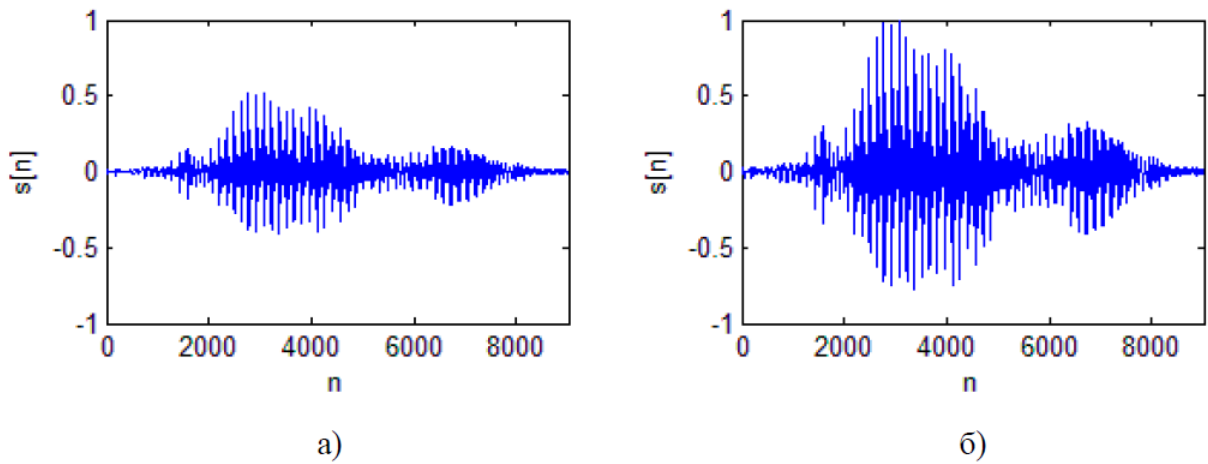


Рис. 2.4. Оцифрований мовленнєвий сигнал до (а) і після (б) нормалізації .

Для оцінки варіативності гучності розглянемо набір з прикладів вимови одного або декількох слів. Знайдемо середнє значення гучності M_q [32] для q -го прикладу довжиною в N відліків і середня M_Q для Q прикладів:

$$M(q) = \sum_{n=1}^N |s[n]|, q = 1, \dots, Q, \quad (2.12)$$

$$M_Q = \frac{1}{Q} \sum_{q=1}^Q M(q), \quad (2.13)$$

Після цього розрахуємо відносне відхилення $D(q)$ гучності кожного прикладу від середнього:

$$D(q) = \left| 1 - \frac{M(q)}{M_Q} \right|, \quad (2.14)$$

З формул (2.12), (2.13) видно, що результат впливає як абсолютне значення відліку, і кількість цих відліків у прикладі, тому необхідно оцінити варіативність гучності набору прикладів одного класу, довжина яких приблизно однакова, і варіювання

гучності бази в цілому. На рис. 2.5.а)-б) показано $D(q)$ для $Q = 100$ записаних прикладів слова «три» до та після нормалізації відповідно. Розкид гучності склав 28.5% для вихідних прикладів та 14.3% для нормалізованих. Графіки 2.5.в)-г) відображають $D(q)$ для мовленнєвої бази з $Q = 2000$ прикладів різних вимов. Розкид гучності склав 25.8 % для вихідної бази і 23.11 % нормалізованої.

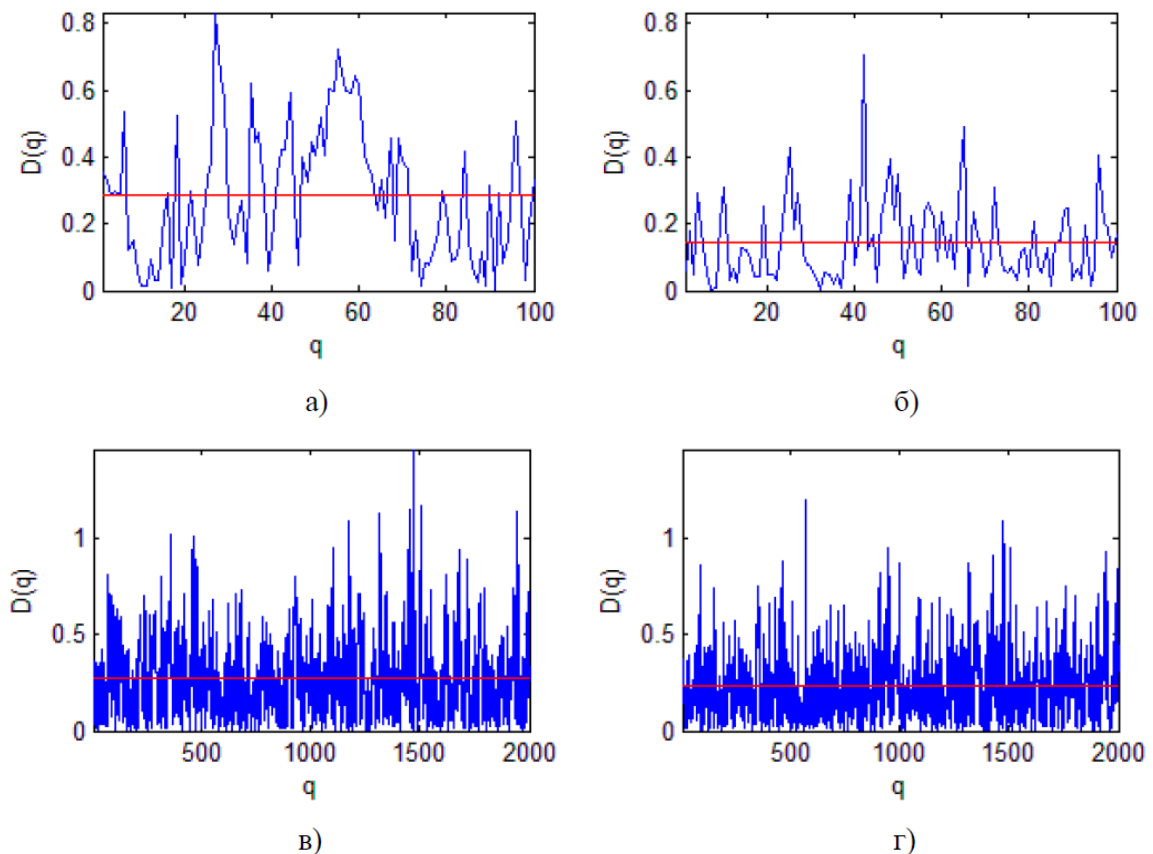


Рис. 2.5. Відхилення енергії прикладу від середнього значення енергії для (а, в) вихідних прикладів, (б, г) нормалізованих

Як видно з проведеного дослідження, застосування нормалізації завжди дозволяє зменшити розкид гучності для різних вимов. Дані результати були отримані для мовної бази, зібраної в однакових умовах на єдиній доступній апаратурі, тому нормалізація зіграла загалом незначну роль. Проте, нормалізація необхідна у роботі реальної системи, коли прийом мовного сигналу ведеться в різних умовах.

ВИСНОВКИ ДО РОЗДІЛУ 2

1. наведено класифікацію методів розпізнавання мови та опис найбільш поширених з них – динамічне вирівнювання часу, побудова штучних нейронних мереж та апарат прихованих марківських моделей. На основі аналізу цих методів було обрано приховані марківські моделі в поєднанні з нейронною мережею;

2. розглянуто різні типи асоціативних середовищ у історичному порядку їх появи. Аналіз досягнень у цій галузі дозволив вибрати асоціативне осциляторне середовище як найбільш відповідне для використання в даній роботі.

3. на зібраній експериментальній мовній основі досліджено варіативність гучності мовного сигналу, що виникає через велику кількість факторів: чутливість апаратури, гучність голосу, що постійно змінюється, і т.д. Розглянуто процедуру нормалізації, що дозволяє знизити розкид гучності мови.

РОЗДІЛ 3

МЕТОД ВИДІЛЕННЯ ДІЛЯНОК З МОВОЮ НА ОСНОВІ РОЗПОДІЛУ ЛОКАЛЬНИХ ЕКСТРЕМУМІВ ДЛЯ ІДЕНТИФІКАЦІЇ ЛЮДИНИ

Вхідний сигнал, що надходить на системи автоматичного розпізнавання мови (САРМ), насичений різними сторонніми звуками: мікрофонним шумом, гучним диханням диктора і т.д. Для підвищення точності розпізнавання першим кроком роботи САРМ є виділення тих ділянок сигналу, на яких присутня мова [18, 22, 23, 34].

3.1. Знаходження значень короткочасної енергії і частоти проходів через нуль

Існуючі методи виділення мови із сигналу $s(n)$ працюють у часовій області. До них належать знаходження короткочасної енергії сигналу $E_s(m)$, короткочасної потужності сигналу $P_s(m)$ і частоти проходів через нуль $Z_s(m)$ [22, 28, 35]:

$$E_s(m) = \sum_{n=m-L+1}^m s^2(n), \quad (3.1)$$

$$P_s(m) = \frac{1}{L} \sum_{n=m-L+1}^m s^2(n), \quad (3.2)$$

$$Z_s(m) = \frac{1}{L} \sum_{n=m-L+1}^m \frac{|\operatorname{sgn}(s(n)) - \operatorname{sgn}(s(n-1))|}{2}, \quad (3.3)$$

$$\operatorname{sgn}(s(n)) = \begin{cases} +1, & s(n) \geq 0 \\ -1, & s(n) < 0, \end{cases} \quad (3.4)$$

Ці характеристики обчислюються для m -ї ділянки (блоку) довжини L сигналу $s(n)$. Видно, що потужність, фактично, еквівалентна енергії сигналу, віднесеної до довжини блоку. Вирішальне правило VAD (Voice Activation Detection) - це порогова функція, де враховують всі вищеописані характеристики [38]:

$$VAD(m) = \begin{cases} 1, & W_s(m) \geq t_w \\ 0, & W_s(m) < t_w, \end{cases} \quad (3.5)$$

$$W_s(m) = P_s(m) \cdot (1 - Z_s(m)) \cdot S_c, \quad (3.6)$$

$$t_w = \mu_w + \alpha \delta_w, \quad (3.7)$$

$$\alpha = 0.2 \cdot \delta_w^{-0.8}, \quad (3.8)$$

У функції $VAD(m)$ як поріг виступає t_w , що складається з μ_w - середнього значення шумових відліків, і δ_w - відхилення значень шумових відліків від середнього. При цьому коефіцієнти, що масштабують, підбираються експериментально, наприклад, $S_c = 1000$.

Недоліком цього є припущення у тому, що мова звучить голосніше, ніж шум, що, звісно завжди так. Якщо у вхідному сигналі зустрінуться ділянки, скажімо, з досить гучним шумом прогортання сторінок або іншим шурхіттям, значення $P_s(m)$ буде порівняно з тим, що прийняте для ділянок, що містять мову. Також недоліком вважаються необхідність ретельного експериментального підбору масштабуючих коефіцієнтів.

Для усунення зазначених недоліків було розроблено підхід з урахуванням аналізу розподілу локальних екстремумів.

3.2. Метод виділення ділянок з мовою на основі розподілу локальних екстремумів

Розглянемо ділянку мовного сигналу, зображену рис. 3.1. Відліки під номерами 2, 5, 6, 10, 11, 12 є екстремумами. Позначимо через M загальну кількість екстремумів на ділянці. Максимально можливе значення екстремуму e (взагалі кажучи, будь-якого відліку) визначається розрядністю відліку k :

$$e \in [-2^{k-1}, 2^{k-1} - 1], \quad (3.9)$$

Наприклад, в даній роботі всі мовні образи були оцифровані з $k=16$, а $2^{k-1} = 32768$. Позначимо через E_j - й інтервал значень екстремумів, так що ширина інтервалів дорівнює $L = e_{jmax} - e_{jmin}$:

$$E_j = [e_{jmin}, e_{jmax}], j = 0, \dots, \frac{2^k}{L}, \quad (3.10)$$

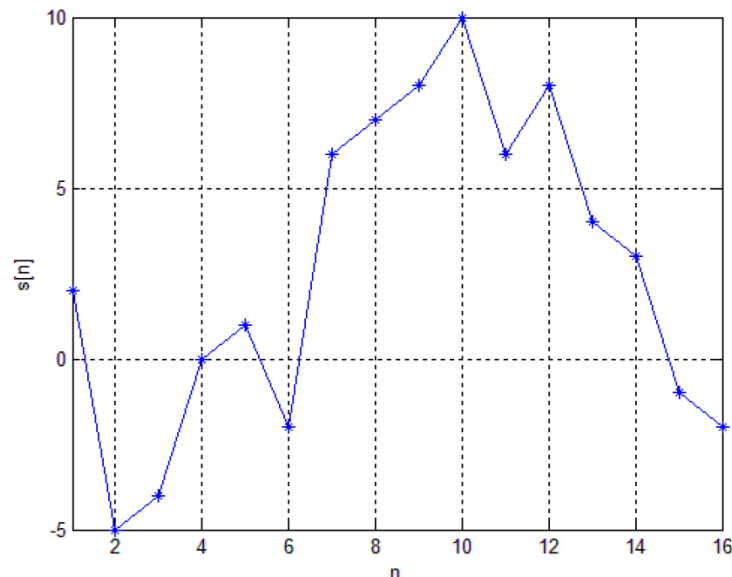


Рис. 3.1. Приклад ділянки дискретного мовного сигналу з шістьма екстремумами

Тепер введемо величину D_j , що позначає кількість екстремумів, що потрапили в інтервал E_j :

$$D_j = \sum_{i=1}^M \gamma(e_i, E_j), j = 0, \dots, \frac{2^k}{L}, \quad (3.11)$$

Функція γ визначає, чи потрапив екстремум e_i в інтервал E_j :

$$\gamma(e_i, E_j) = \begin{cases} 1, & e_i \in E_j \\ 0, & e_i \notin E_j \end{cases}, \quad (3.12)$$

Зрештою, позначимо P_j частку екстремумів, що потрапили в інтервал E_j , від загальної кількості M :

$$P_j = \frac{D_j}{M}, j = 0, \dots, \frac{2^k}{L}, \quad (3.13)$$

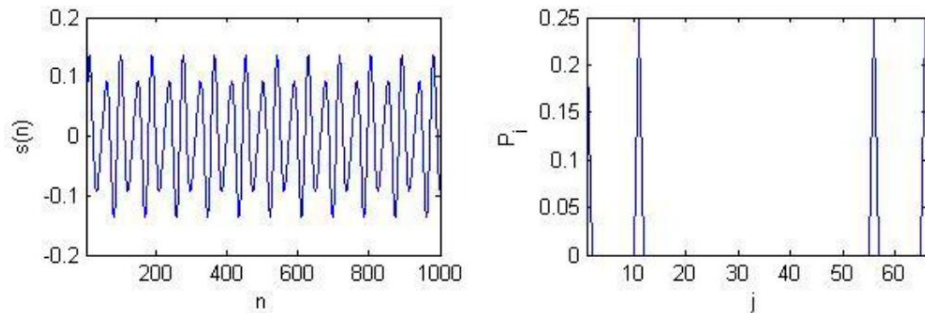
Таким чином, для кожної ділянки мовного сигналу можна побудувати вектор розподілу екстремумів $P = (P_0, \dots, P_j, \dots, P_{\frac{2^k}{L}})$, в якому P_j - це ймовірність того, що значення екстремуму потрапить в інтервал E_j , при цьому $\sum_j P_j = 1$. Отриманий вектор P_i є вектор ознак, що можна подавати на розпізнавач.

Для чистого тону розподіл екстремумів є яскраво вираженими симетричними піками (рис. 3.2.а). Для білого шуму розподіл має асиметричний та рівномірний характер без яскраво виражених піків (рис. 3.2.б). Додавання все більш високочастотних обертонів заповнює ненульовими значеннями інші інтервали розподілу, зберігаючи

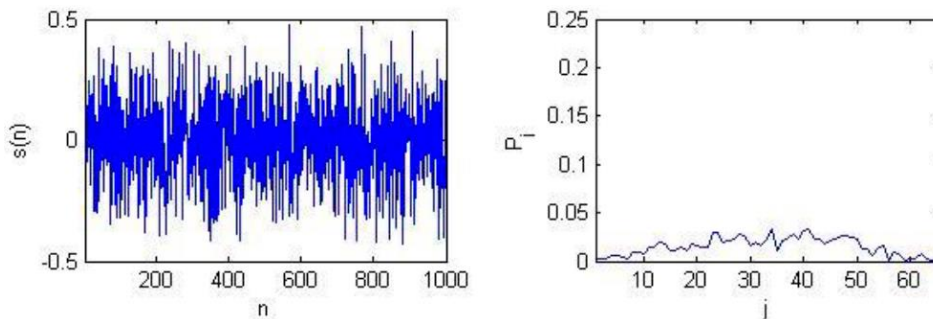
симетричність щодо $\frac{2^k}{L}/2$ (рис. 3.2.в). Для усунення цього ефекту, будується огинаюча сигналу, яка згладжує малоамплітудні коливання, усуваючи обертони [28, 36] (рис. 3.2.г). Для цього будуються дві огинаючі сигналу e^{max} і e^{min} , містять тільки максимуми та мінімуми відповідно. Їхня довжина дорівнює довжині вихідного сигналу; позначимо її N . Очевидно, $M < N$, так що між екстремумами ми знаходимо значення відліків за допомогою лінійної інтерполяції. Потім будується середній

сигнал s , у якого кожен відлік s_i лежить між максимумом і мінімум вихідного сиг-

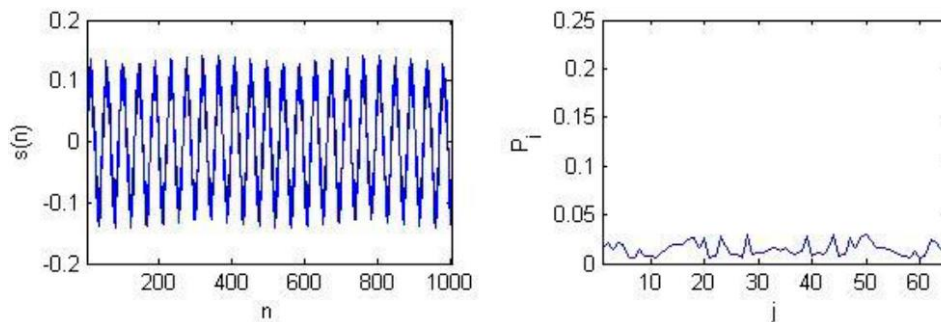
налу: $s_i = \frac{e_i^{max} + e_i^{min}}{2}, i = 0, \dots, N$.



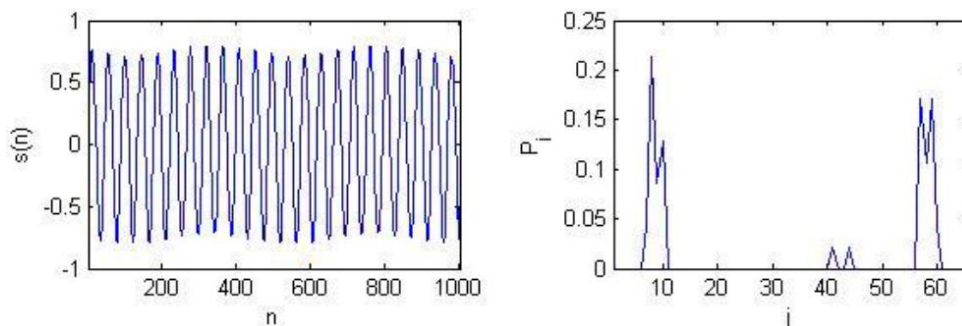
а) Тон 1 КГц і обертон 1,5 КГц та його розподіл екстремумів



б) Білий шум і його розподіл екстремумів



в) Тон 1 КГц і обертон 14 КГц і його розподіл екстремумів



г) Огинаюча сигналу з пункту в) та його розподіл екстремумів

Рис. 3.2. Різні типи сигналів і розподілу екстремумів для них

Використовуючи ці особливості розподілів екстремумів, набула подальшого розвитку підсистема виділення мовного сигналу із вхідного сигналу. Вона включає наступні компоненти (рис. 3.3):

1. Блок нормалізації - знятий з мікрофона сигнал нормалізується по амплітуді вікном, що ковзає, шириною 10 мс. Ця процедура необхідна для усунення варіативності гучності сигналу, яка залежить від індивідуальних характеристик голосу та апаратури. Для того щоб всі ділянки були посилені до однакового значення гучності, нормалізація проводиться в межах вікна.
2. Блок побудови огинаючої, що згладжує малоамплітудні коливання.
3. Блок нормалізації – через побудову огинаючої, необхідна повторна нормалізація сигналу.
4. Блок знаходження розподілу екстремумів для m -ї ділянки.
5. Вирішальна логіка. Приймається рішення, чи містить m ділянка сигналу мову або шум. Як вирішальну логіку було обрано **тришарову повнозв'язну нейронну мережу (НМ)** прямого поширення. При цьому, перед використанням мережі її потрібно навчити, за що відповідає блок навчання.

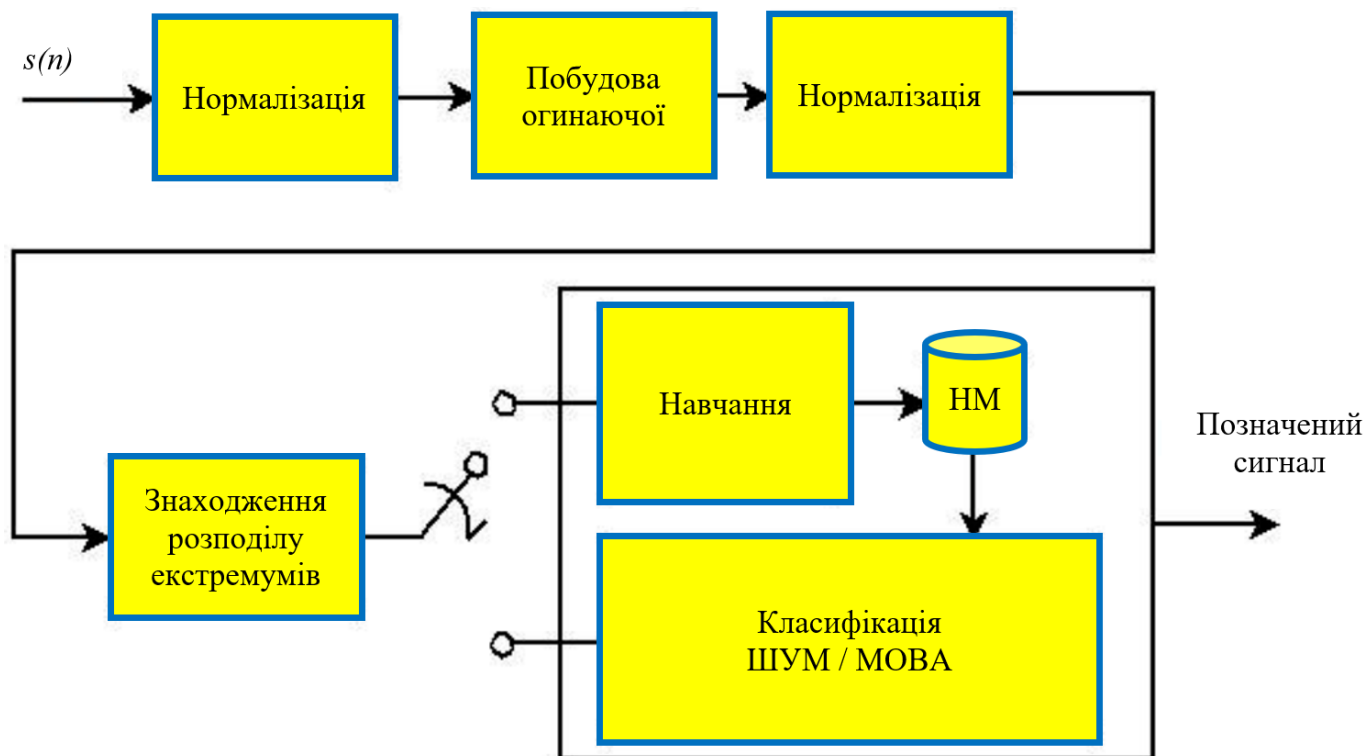


Рис. 3.3. Розроблена підсистема виділення мови

Передумови використання багатошарової нейронної мережі були такими:

1) Заміна одновимірної величини $W_s(m)$ на вектор P змусила уникнути порогової функції і застосувати роздільника у вибраному просторі ознак. CP – чудово опрацьований інструмент для вирішення завдань класифікації.

2) Для надійного поділу образів потрібно можливість будувати нелінійні розділяючі функції - завдання, з якою справляється багатошарова CP [32].

3) CP існує як у програмному, так і в апаратному втіленні. Хоча в даній роботі була використана програмна модель, можливість ефективної апаратної реалізації – внесок у майбутнє системи, що розробляється.

Для навчання системи бажано використовувати звуки мови, із якими системі доведеться зіткнутися при класифікації, у своїй виключивши всі фрикативні приголосні, так як їх дуже важко відрізнити від білого шуму.

Для перевірки запропонованої методики була навчена та протестована система виділення мовного сигналу, здатна стійко виділяти слова «Вперед» та «Назад» із вхідного звукового потоку. Такий приклад немає сенсу, так як виділені слова могли б надходити на розпізнавач голосових команд управління. Результати порівняли з традиційною VAD- функцією.

Експериментально була підібрана довжина вектора $P = (P_0, \dots, P_j, \dots, P_{\frac{2k}{L}})$, що дорівнює 66, так, що збільшення розмірності не призводить до покращення результатів. Таким чином, вхідний шар нейронної мережі включав шістдесят шість вузлів, а вихідний – два.

Спочатку мережа навчалася на звукових фрагментах двох типів – з промовою і шумом. Навчальний фрагмент, містить мова, включав всі звуки слів "Вперед" і "Назад", за винятком фрикативних приголосних (наприклад, /f/). Навчальний фрагмент із шумом був наповнений різними неінформативними звуками: мікрофонним шумом, шарудінням паперу, клацаннями кнопок комп'ютерної миші. Обидва фрагменти мали однакову тривалість і сканувалися ковзним вікном шириною 10 мс, для якого знаходився вектор P – це і був навчальний образ для нейронної мережі. Разом навчальна вибірка містила 1856 образів, 928 кожного типу.

Для тестування був записаний звуковий фрагмент, що містить команди "Вперед" і "Назад", а також шуми, гучність яких не поступається гучності мови. Тестовий сигнал було розмічено за допомогою VAD-функції (рис. 3.4) і навченої описаним вище способом системою (Рис. 3.5). У обох випадках вхідний сигнал сканувався вікном 10 мс, яке позначалося або як "мова" або як "шум". На рис. 3.4 пунктиром зображено VAD-функцію, а на рис. 3.5 – межі областей «мова» та «шум».

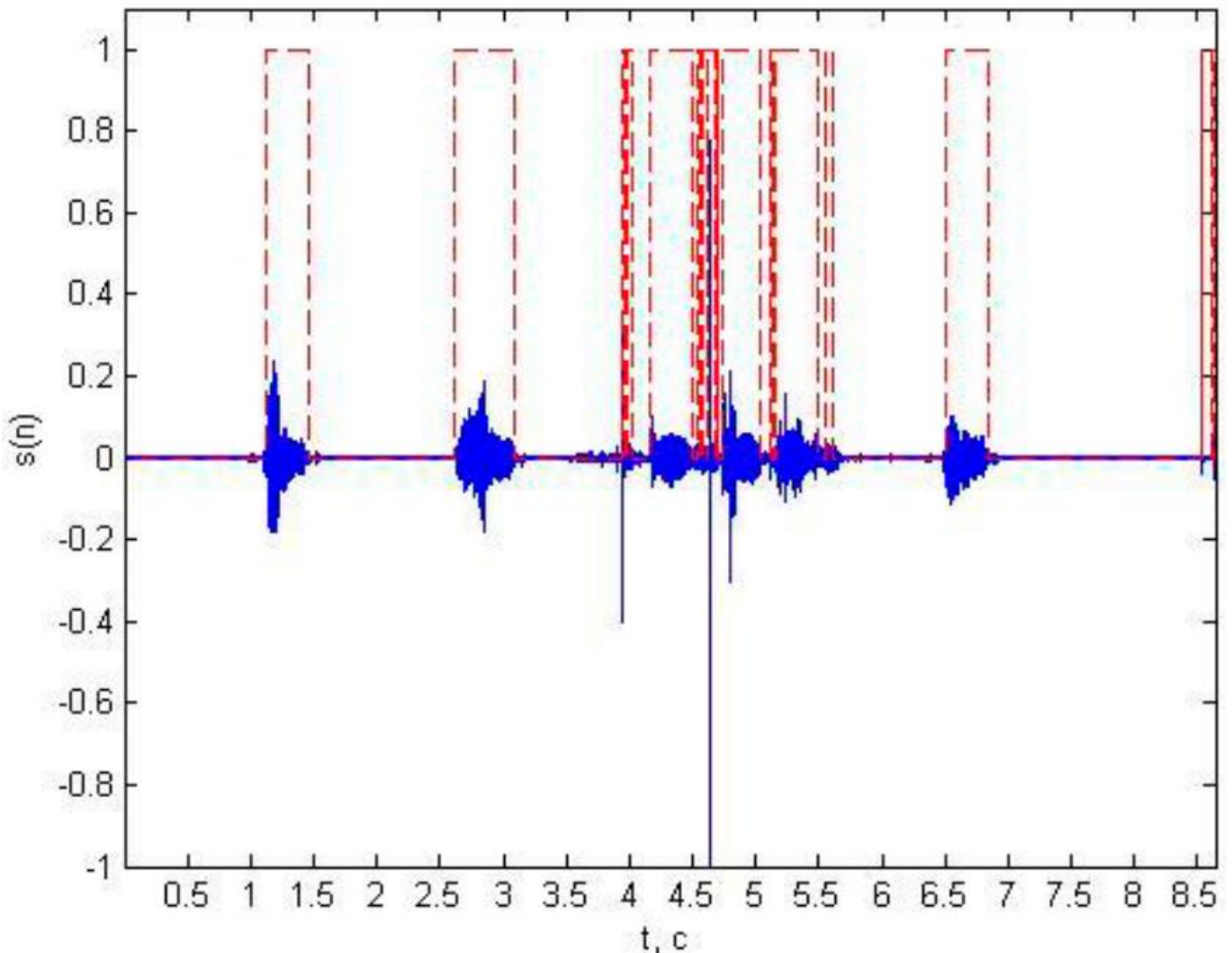


Рис. 3.4. Тестовий сигнал «Вперед-Назад-шурхіт сторінок-Вперед», розмічений за допомогою традиційної VAD-функції

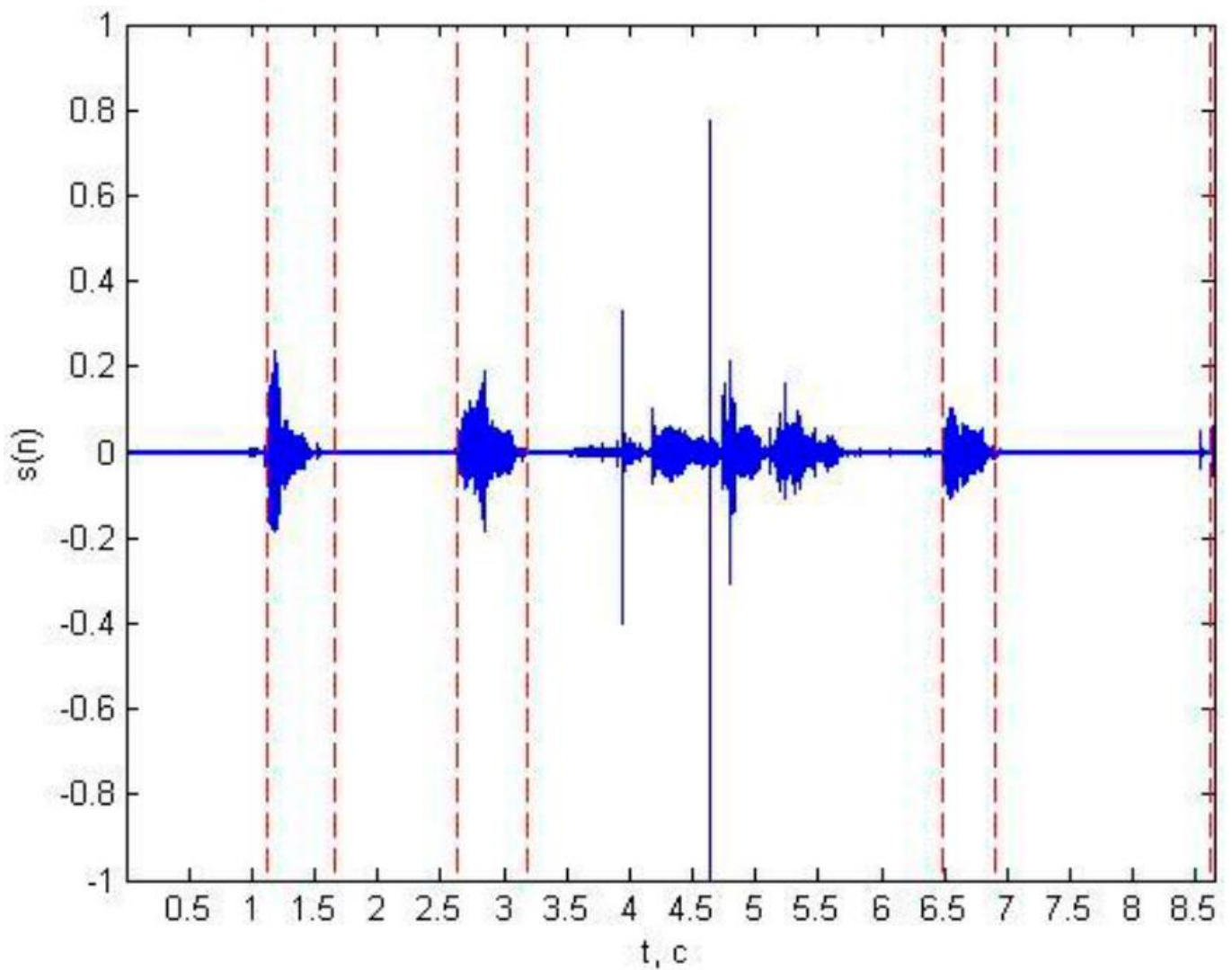


Рис. 3.5. Тестовий сигнал «Вперед-Назад-шурхіт сторінок-Вперед», розмічений розробленою системою виділення мови

Результат оцінювався як відсоток чітко помічених фрагментів по 10 мс від загального числа (табл. 3.1). Експерименти показали, що запропонована система, будучи навченою відрізняти шуми різної природи від мови, добре працює за їх наявності. Однак, обидва підходи упускають фрикативні звуки (на $t=1$ с. втрачено звук /f/).

Результати тестування методу екстремумів і порівняння з VAD-функцією

	VAD- функція		Метод екстремумів	
	Число фрагментів	Частка від спільного числа фрагментів, %	Число фрагментів	Частка від спільного числа фрагментів, %
Усього оброблено	864	100	864	100
«Мова»	117	13.54	117	13.54
"Шум"	747	86.46	747	86.46
Усього помічено «мова»	236	27.31	152	17.59
Усього помічено «шум»	628	72.69	712	82.41
Правильно позначено «мова»	117	49.57	140	92.11
Правильно позначено «шум»	623	99.20	705	99.02
Усього вірно позначено	740	85.65	845	97.80

3.3. Виділення ознак мовного сигналу

Акустична модель мовотворення. Мова – складний сигнал, в якому присутня незначна для розпізнавання інформація про вік, поле, настрій диктора, вона може бути спотворена фоновим шумом та акцентом. Тому існує завдання виділення таких ознак мовного сигналу, якими можна було б відрізнити різні фонемі мови. Склад ознак визначається моделлю мовотворення, що включає сигнал збудження $u(n)$ і фільтр мовного тракту, імпульсна характеристика $h(n)$ якого змінюється в часі (Рис. 3.6). Вихідний сигнал $s(n)$, репрезентуючий собою згортку $u(n)$ і $h(n)$, і є мова [30, 32].

Для формування мовлення людина використовує складний мовний апарат, що включає легені, бронхи, трахею, голосові зв'язки, носові проходи і т.д. Спочатку в легенях формується сигнал збудження, що проходить через інші ділянки мовного тракту, які його всіляко перетворюють. У цьому чільну роль цих перетвореннях грає стан і форма голосових зв'язок.

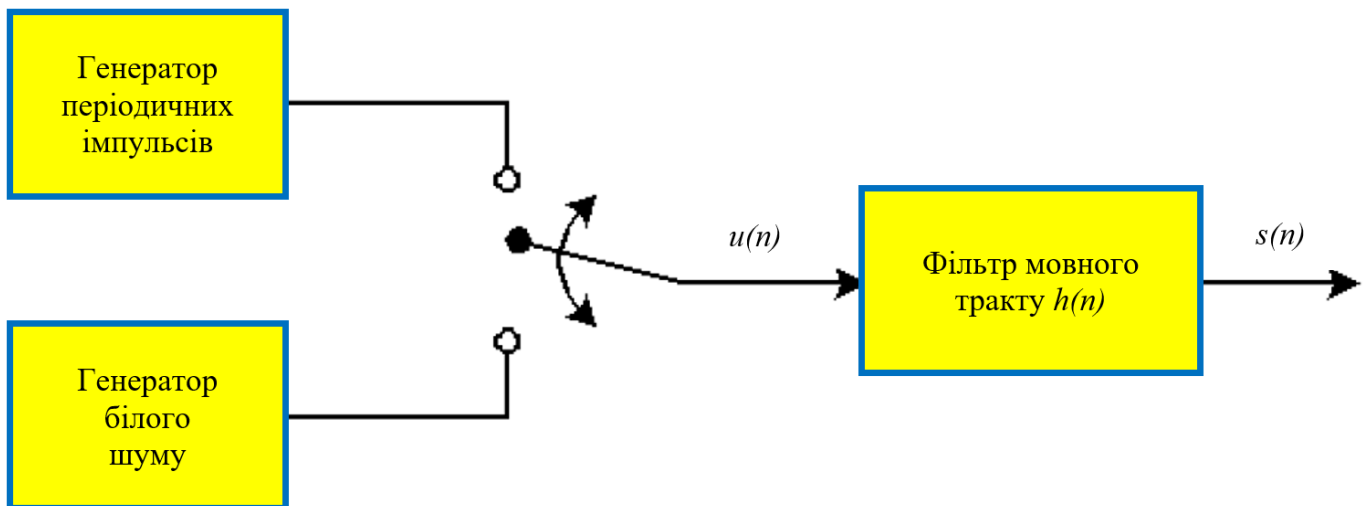


Рис. 3.6. Модель мовотворення "збудження -фільтр"

При промові голосних звуків сигнал збудження набуває квазіперіодичної форми, а у разі глухих звуків – форму білого шуму. У цьому конкретну фонему визначає зміна мовного тракту, а не форма сигналу збудження.

Як було зазначено вище, параметри фільтра змінюються у часі, проте ця зміна не може відбуватися миттєво. Дослідження показали, що в межах 10-40 мс мовний тракт не змінює своєї конфігурації, отже в межах цього короткого проміжку часу мовний сигнал вважатимуться стаціонарним, тобто. його спектральні властивості не змінюються.

Мел-кепстральні коефіцієнти. У цій роботі використовується найбільш популярний на сьогоднішній день метод виділення ознак - перебування мел-кепстральних коефіцієнтів (Mel Frequency Cepstral Coefficients – MFCC). Він дозволяє відокремити сигнал збудження від фільтра мовного тракту за допомогою обчислення кепстра - зворотного дискретного перетворення Фур'є (ЗДПФ) від логарифму дискретного перетворення Фур'є (ДПФ) вхідного сигналу. Ідея, що стоїть за застосуванням кепстру, наступна: згортці сигналу $u(n)$ та імпульсної характеристики фільтра $h(n)$ у часовій області відповідає добуток їх ДПФ у частотній:

$$\mathfrak{Z}(u(n) * h(n)) = \mathfrak{Z}(u(n)) \cdot \mathfrak{Z}(h(n))$$

де $\mathfrak{Z}(x)$ – ДПФ x ,

3.14

Логарифмування ДПФ відноситься до класу гомоморфних перетворень, оскільки дозволяє замінити їх добуток на суму [26]:

$$\log(\mathfrak{Z}(u(n)) \cdot \mathfrak{Z}(h(n))) = \log \mathfrak{Z}(u(n)) + \log \mathfrak{Z}(h(n)), \quad (3.15)$$

Знаходження MFCC здійснюється у кілька етапів (рис. 3.7) [29, 33, 37]:

1. Нормалізація вихідного сигналу, що дозволяє вирівняти його амплітуду та посилення високих частот. Форманти низьких частот мають велику амплітуду порівняно з формантами вищих частот, хоча останні також несуть важливу для розпізнавання інформацію. Тому до вхідного сигналу застосовують КІХ-фільтр:

$$s'(n) = s(n) - 0.95 \cdot s(n - 1), \quad (3.16)$$

2. Виділення короткочасної ділянки сигналу (~32 мс.) та накладання віконної функції w_k для мінімізації витоку діапазону. Сегмент, що вийшов, називають фреймом довжиною K відліків та подальшу роботу ведуть у межах кадру:

$$x_t(k) = s'(k + t \cdot K) \cdot w(k), 0 \leq k \leq K - 1, \quad (3.17)$$

Таким чином, вектор ознак одержують для кожного $0 \leq t \leq T$ кадру вхідного сигналу $s'(n)$. Як віконну функцію зазвичай використовують функцію Хеммінга:

$$w(k) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi k}{K - 1}\right), \quad (3.18)$$

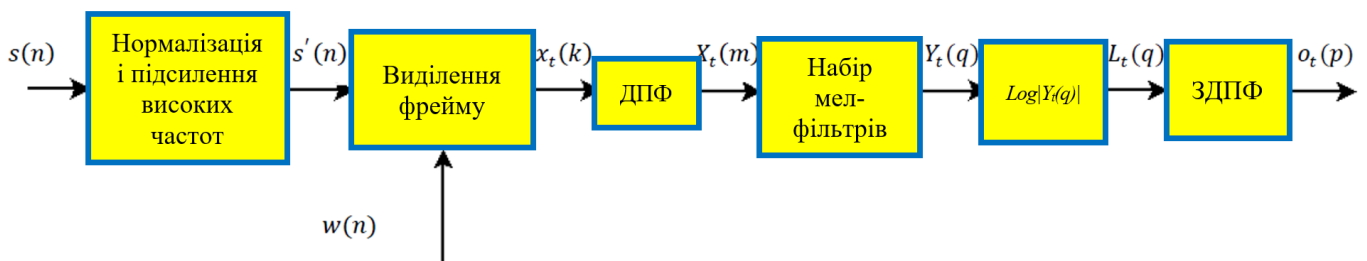


Рис. 3.7. Знаходження мел-кепстральних коефіцієнтів

3. Обчислення ДПФ для фрейму:

$$X_t(m) = \sum_{k=0}^{K-1} x_t(k) \cdot e^{-j2\pi mk/K}, \quad (3.19)$$

Для обчислення ДПФ використовується алгоритм Швидкого Перетворення Фур'є (ШПФ).

4. Накладення набору Q мел-фільтрів на кадр у частотній області, так що на виході кожного q -го фільтра знімають енергію $Y_t(q)$ у смузі пропускання цього фільтра. Таким чином моделюють сприйняття мови людиною: роздільна здатність слуху зростає при русі по спектру від низьких частот до високих. Центральні частоти F_q мел-фільтрів вибираються за так званою мел-шкалою, яка залежить від звичайної за логарифмічним законом (рис. 3.8):

$$F_{mel} = 2095 \cdot \log_{10} \left(1 + \frac{F_{Hz}}{700} \right), \quad (3.20)$$

Мел-фільтри розташовують в діапазоні частот від 0 до $F_{max} = 8$ кГц. Таким чином, якщо знайти відповідну F_{max} частоту на мел-шкалі F_{max} , потім поділити на Q рівних відрізків проміжок від 0 до F_{max} , і для F_{max}^{mel} кожного значення знайти відповідну їй частоту F_{Hz} , то отримаємо Q значень частот, відстань між якими зростає за логарифмічним законом. Перехід від мел-шкали до традиційної здійснюється за такою формулою:

$$F_{Hz} = 700 \cdot \left(10^{\frac{F_{mel}}{2595}} - 1 \right), \quad (3.21)$$

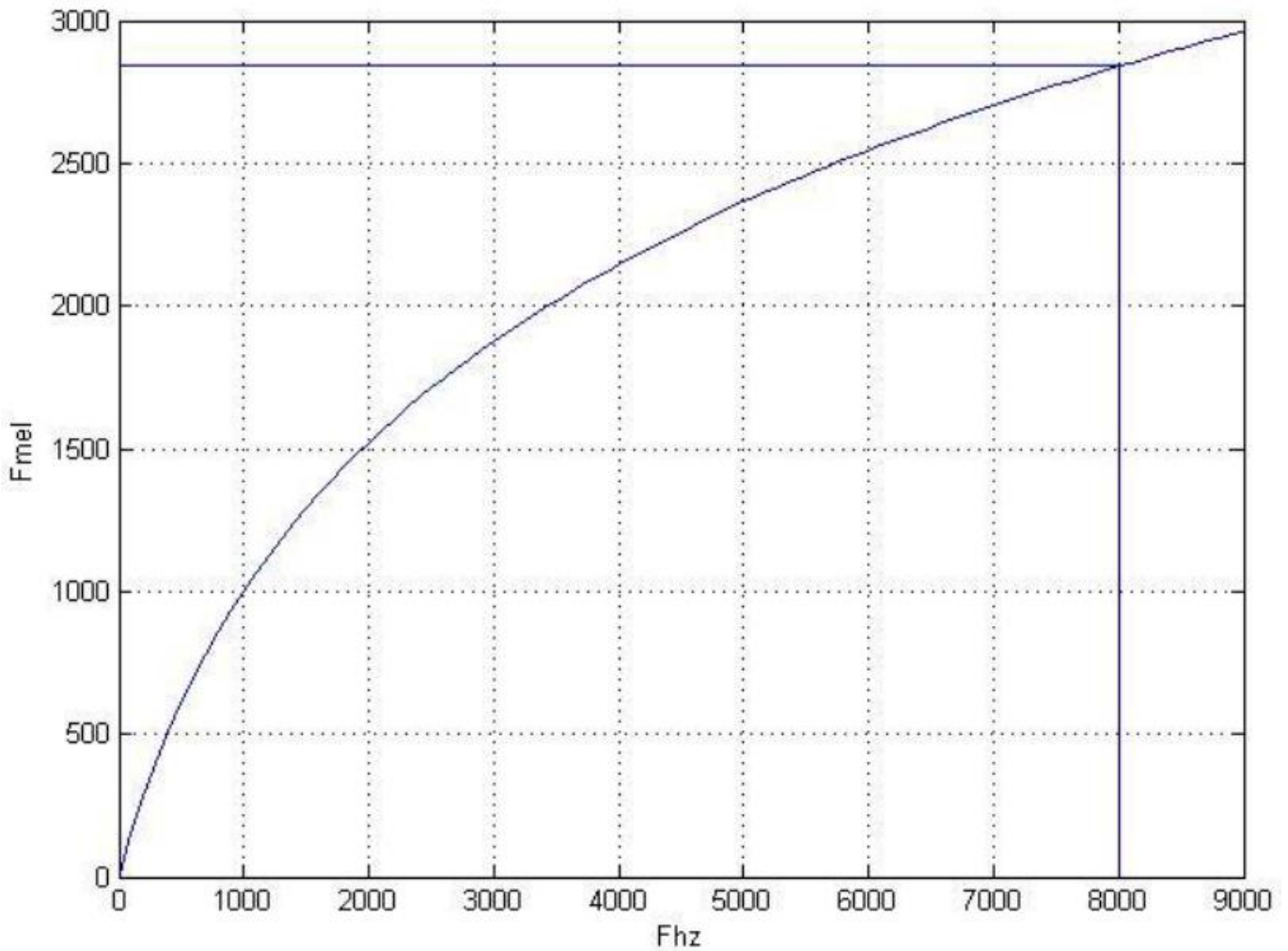


Рис. 3.8. Відповідність частот звичайної шкали частотам мел-шкали

Після знаходження центральних частот потрібно побудувати амплітудно-частотні характеристики (АЧХ) фільтрів, які мають трикутну форму (рис. 3.9):

$$U_q(k) = \begin{cases} 1 - \frac{|k - F_q|}{\Delta_q}, & |k - F_q| < \Delta_q \\ 0, & |k - F_q| \geq \Delta_q \end{cases}, \quad (3.22)$$

Смуги пропускання фільтрів теж ростуть за логарифмічним законом, причому смуга пропускання наступного фільтра до середини "наповзає" на смугу пропускання кожного попереднього.

В результаті накладання мел-фільтрів отримують Q значень енергії кадру:

$$Y_t(q) = \sum_{k=0}^{K-1} |X_t(k)| \cdot U_q(k) \quad (3.23)$$

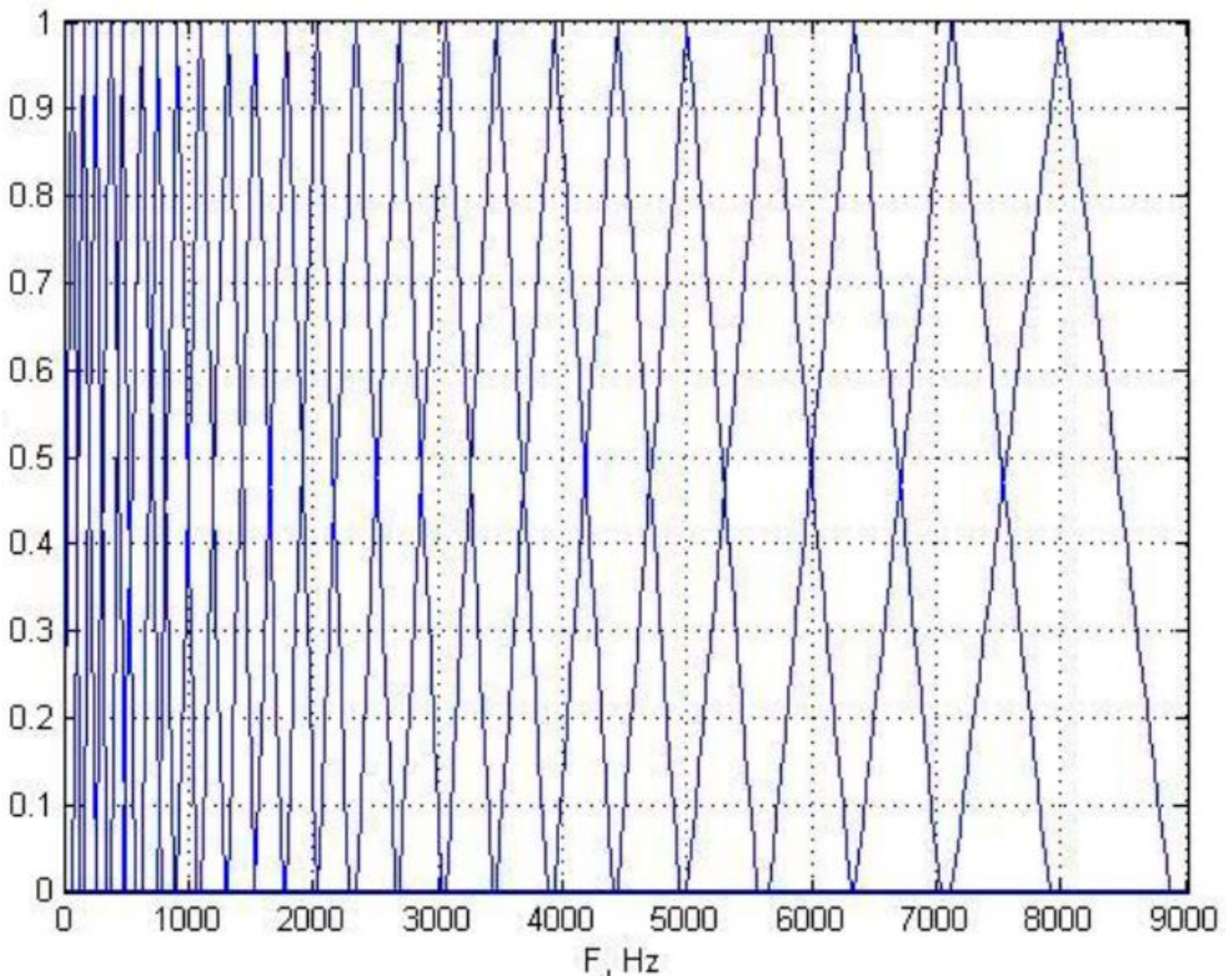


Рис. 3.9. АЧХ Q=24 мел-фільтрів

5. Логарифмування $Y_t(q)$. На цьому етапі виконується гомоморфне перетворення. Для того, щоб відокремити сигнал збудження від характеристики фільтра, обчислюється логарифм модуля $Y_t(q)$:

$$L_t(q) = \log|Y_t(q)| \quad (3.23)$$

6. Обчислення ЗДПФ. На останньому етапі отримують мел коефіцієнти шляхом обчислення ЗДПФ для $L_t(q)$. При тому, що $L_t(q)$ дійсний та симетричний, ЗДПФ буде еквівалентно Дискретному Косинусному Перетворенню:

$$o_t(p) = \sum_{q=1}^Q L_t(q) \cdot \cos\left(p \cdot \left(q - \frac{1}{2}\right) \cdot \frac{\pi}{Q}\right), p = 0, \dots, P, \quad (3.23)$$

В результаті одержують P коефіцієнтів, при цьому, хоча P може дорівнювати Q , зазвичай беруть лише половину значень MFCC. Це тим, що кепстр сигналу збудження, зазвичай "правіше" кепстра мовного тракту.

Таким чином, на розпізнавач надходить p -мірний вектор o_t , що містить мел-кепстральні коефіцієнти для t -го кадру. На рис. 3.10 зображені MFCC кадру тривалістю 32 мс для п'яти прикладів голосного звуку /a/.

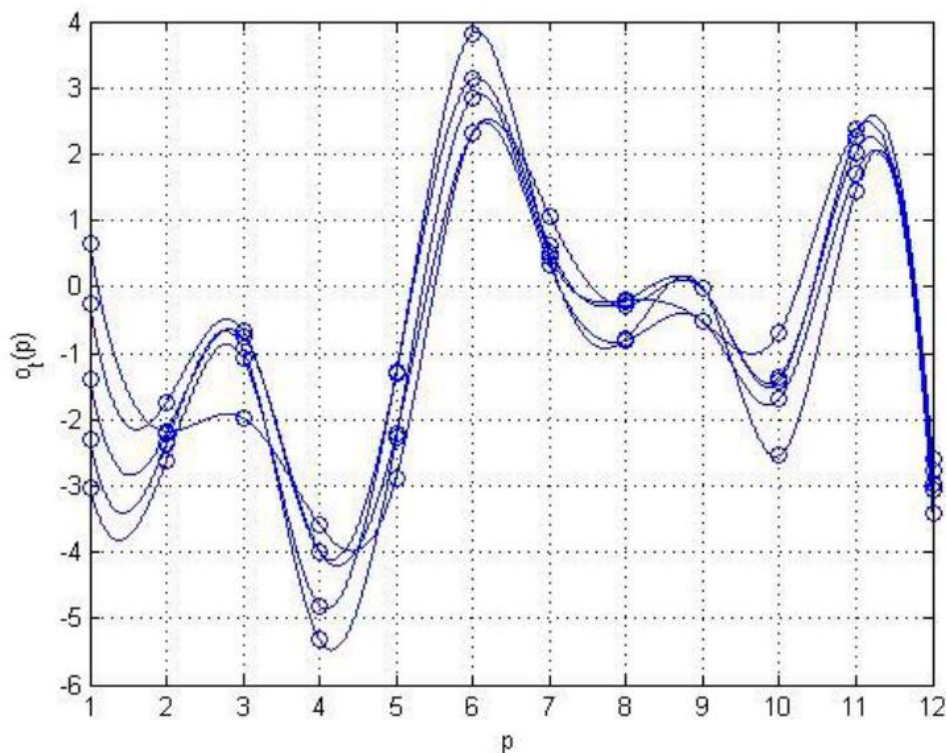


Рис. 3.10. MFCC одного кадру для п'яти прикладів голосного звуку /a/

Видно, що значення MFCC для однієї ділянки вимови близькі до різних прикладів вимови.

3.4. Векторне квантування

Для подальшого стиснення мовного сигналу, а також спрощення розпізнавача застосовують процедуру векторного квантування (Vector Quantization – VQ), суть якої полягає в тому, щоб на етапі навчання (кластеризації) складається словник з $M=2^b$ - мірних векторів (кодових слів – еталонів), а потім на етапі класифікації аналізований вектор замінюється індексом найближчого до нього кодового слова [26, 30, 48]. Таким чином досягаються такі цілі:

- стиск мовного сигналу – замість послідовності p -мірних векторів отримують послідовність тієї ж довжини, але з b -бітних чисел;
- перехід від безперервного вектора ознак до дискретного – кожен з мел коефіцієнтів є безперервною величиною, тоді як індекс еталона - ціле число в діапазоні $[1, 2^b]$. Це дозволяє побудувати простіший розпізнавач, наприклад, дискретні приховані марківські моделі.

Як було згадано вище, векторне квантування складається з двох етапів:

1. навчання або кластеризація, що виконується методом k -середніх. На рис. 3.11 зображено приклад кластеризації для двовимірних векторів V_j .

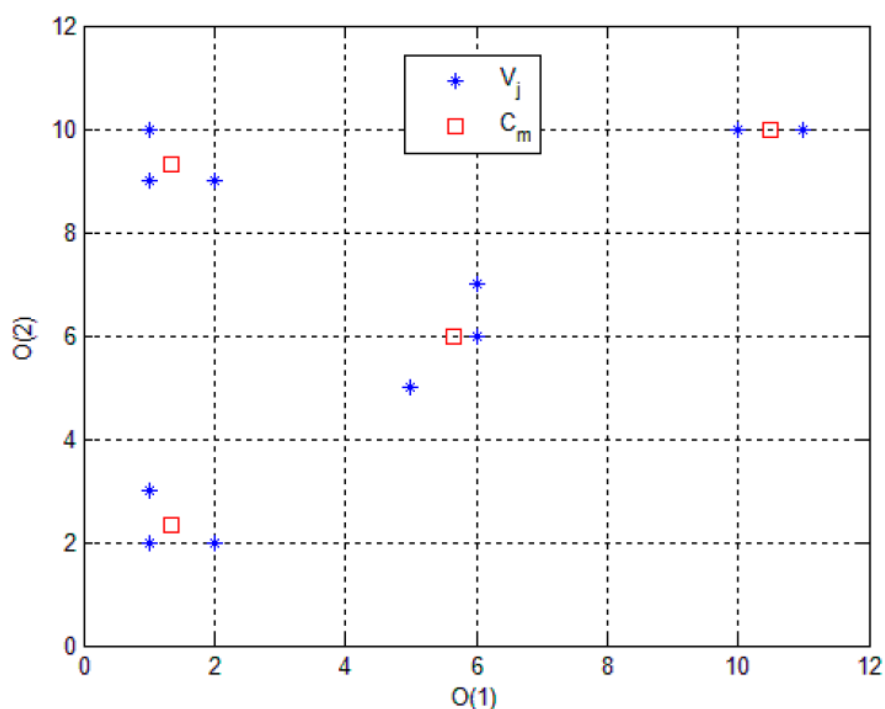


Рис. 3.11. Результат кластеризації для двовимірних векторів V_j

2. класифікація, коли на вхід квантувача надходить невідомий - мірний вектор, а на виході отримують b -бітний індекс найближчого до нього кодового слова.

Кластеризація включає наступні кроки (Рис. 3.12):

1. ініціалізація – випадковим чином вибрати M векторів з навчального набору V_1, V_2, \dots, V_L як початкова безліч кодових слів C_1, C_2, \dots, C_M ;

2. пошук найближчого кодового слова - для кожного навчального вектора V_j знайти таке слово C^* з кодової книги, що $dV_j, C^* = \min m dV_j, C_m$, і прив'язати V_j до C^* ;

3. оновлення мел книги - розрахувати координати кожною центроїди C_m відповідно до набору навчальних образів, асоційованих з нею на кроці 2.

4. ітерація – якщо зміна середньої дистанції навчальних векторів D щодо середньої дистанції D' , отриманої на попередній ітерації, більше заданого порога, то перехід до кроку 2, інакше – кінець.

На малюнку 3.13.а) зображено приклад послідовності індексів для однієї вимови слова «раз», а на малюнку 3.13.б) десять послідовностей (прикладів) слова «раз» накладено на один графік. Видно, хоча індекси різні, вони групуються в області. Природно, застосування векторного квантування вносить помилку уявлення вхідного сигналу, що може вплинути на точності розпізнавання. Експериментально встановлено, що словника $M=256$ цілком достатньо для того, щоб не вносити помилки у розпізнавання [48].

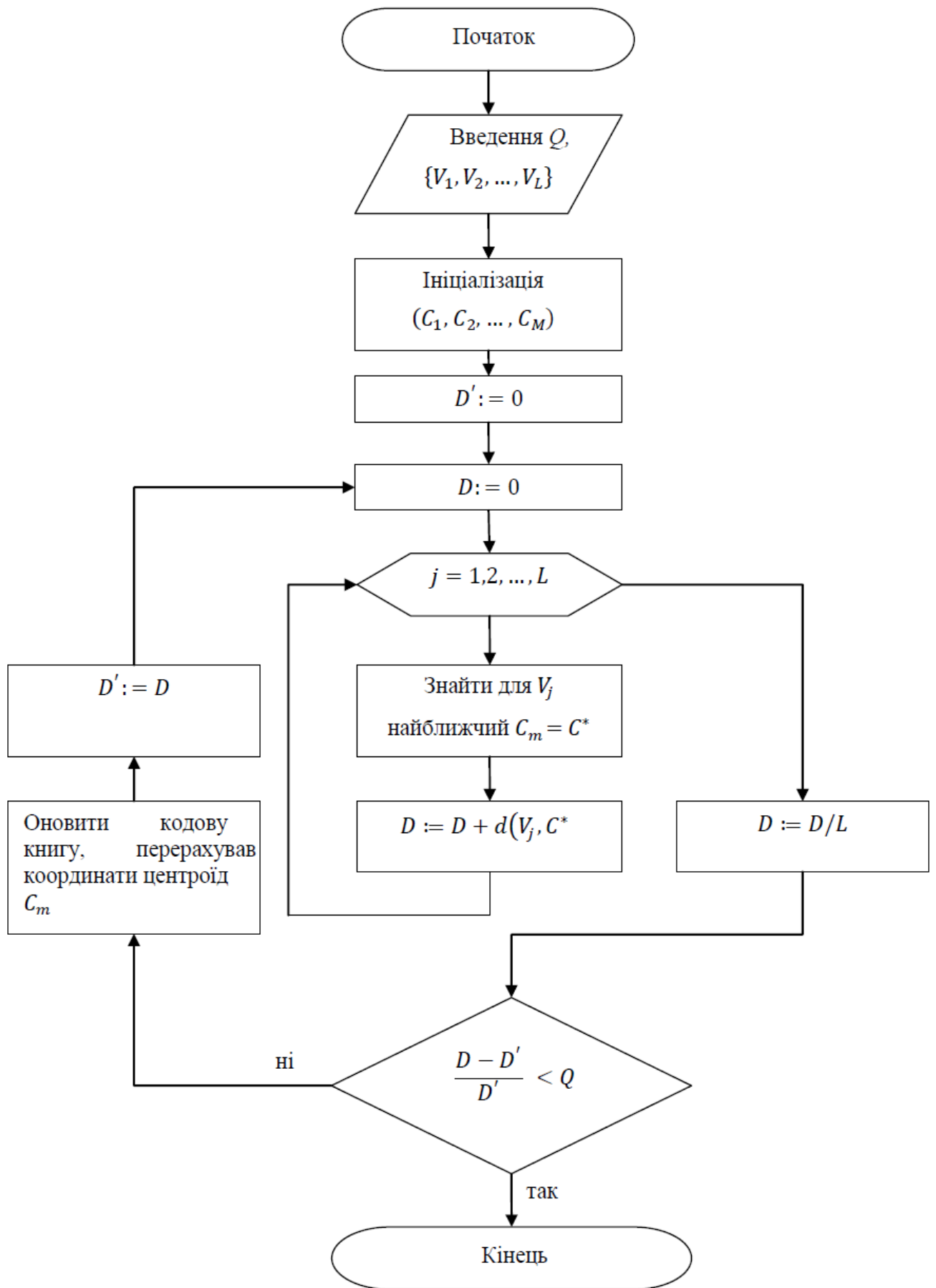
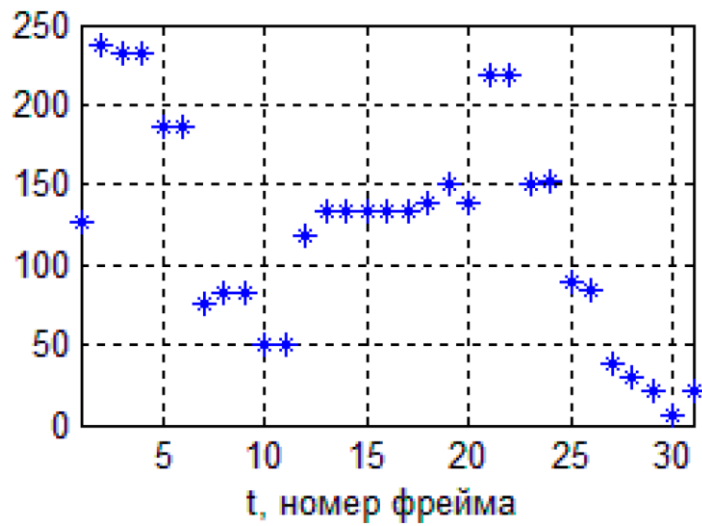
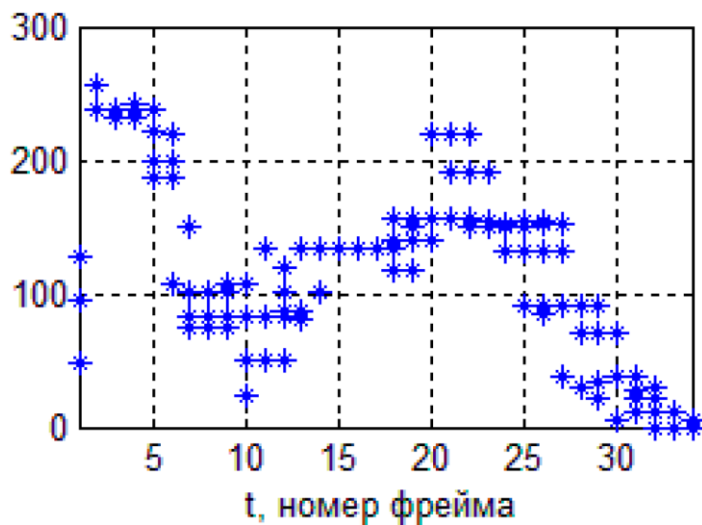


Рис. 3.12. Блок-схема алгоритму навчання у векторному квантуванні



а)



б)

Рис. 3.13. Послідовності індексів для одного (а) і п'яти (б) прикладів вимовлення слова "раз" (а)

ВИСНОВКИ ДО РОЗДІЛУ 3

1. набув подальшого розвитку метод виділення мови, заснований на аналізі розподілу локальних екстремумів вхідного сигналу, що дозволяє знаходити межі мовного сигналу в умовах присутності різних шумів. Розроблено систему виділення мови, що включає блоки нормалізації, знаходження обгинальної та розподілу екстремумів, вирішальної логіки;

2. зроблено порівняння існуючого методу виділення мови та методу екстремумів, яке показало, що останній дає на 12% кращу точність;
3. описано процедуру виділення ознак - перебування вектора мел-кепстральних коефіцієнтів.
4. розглянуто векторне квантування, що дозволяє представити p - мірний вектор ознак за допомогою b -бітного індексу кодового слова. Цим досягається високий стиск мовного сигналу і можливість побудови простішого розпізнавача.

РОЗДІЛ 4

ОХОРОНА ПРАЦІ

4.1. Аналіз небезпечних і шкідливих факторів, що впливають на інженера

Відділ проектування знаходиться на другому поверсі п'ятиповерхового будинку. Приміщення має розміри: довжина 8 м, ширина 4 м, висота 4. Загальна площа - 32 м², загальний об'єм – 128 м³. У відділі знаходиться 5 робочих місць інженерів-проектувальників, оснащені комп'ютерами.

Робоча площа одного співробітника становить:

$$S_{\text{роб}} = \frac{S_{\text{заг.пл}}}{N} = \frac{32}{5} = 6,4 \text{ м}^2$$

Робочий об'єм одного співробітника:

$$V_{\text{роб}} = \frac{V_{\text{заг.об}}}{N} = \frac{128}{5} = 25,6 \text{ м}^3$$

N - кількість співробітників у відділі

$S_{\text{заг.пл}}$ – загальна площа;

$V_{\text{заг.об}}$ – загальний об'єм.

Відповідно до [41] площа на одне робоче місце має становити не менше ніж 6 м², а об'єм не менше ніж 20 м³. Робоче місце інженера-проектувальника відповідає вимогам.

В проектному відділі інженера-проектувальника знаходяться: комп'ютери, принтер. У даному приміщенні температура повітря у теплий період року становить 30°C, використовується природне та штучне освітлення. Штучне освітлення виконано у вигляді переривчастих ліній світлодіодних світильників. Рівень шуму в приміщенні становить 54 дБ, а згідно з Державними санітарними нормами [42] не повинен перевищувати 50 дБ.

Робоче місце розташоване так, щоб природне світло падало з лівої сторони, при цьому відстань зі світлом до робочого місця - 1 м. Висота робочої поверхні столу над підлогою 750 мм, глибина столу – 800 мм, ширина столу 1300мм. Робочий стіл має простір для ніг висотою 650 мм та шириною 600 мм.

Перелік шкідливих та небезпечних виробничих чинників.

Створення сприятливих умов праці, в роботі інженера-проектувальника, має велике значення як для полегшення, так і для підвищення продуктивності праці.

Відповідно до [43] шкідливими виробничими факторами є:

1. Підвищена температура робочого приміщення
2. Недостатня освітленість робочої поверхні
3. Виробничий шум
4. Електромагнітні випромінювання радіочастотного діапазону
5. Іонізуючі випромінювання

Відповідно до [44] робота інженера-проектувальника у приміщенні з енерговитратами 90-120 ккал/год. відносяться до категорії легких фізичних робіт Ia (роботи, що виконуються сидячи і не потребують фізичного напруження).

Таблиця 4.1

Оптимальні величини температури

Період року	Категорія робіт	Температура повітря, °С
Холодний період року	Легка Ia	22-24
Теплий період року		23-25

Допустимі величини температури на постійних робочих місцях:

Період року	Категорія робіт	Температура повітря, °С	
		Верхня межа	Нижня межа
Холодний період року	Легка Ia	25	21
Теплий період року		28	22

Допустимі величини температури на постійних робочих місцях:

Період року	Категорія робіт	Температура повітря, °С	
		Верхня межа	Нижня межа
Холодний період року	Легка Іа	25	21
Теплий період року		28	22

У проектному відділі температура повітря становить 30°C в теплий період року, що перевищує допустиму на 2 °С. Забезпечили температуру приміщення 23 °С, за допомогою механічної вентиляції з вентилятором VORTICE VARIO, повітрообмін якого становить 680 м³ /год.

Недостатня освітленість. В приміщенні встановлені персональні комп'ютери, присутнє природне та штучне освітлення . За вимогами [45], величина коефіцієнта природної освітленості повинна бути не менше 1.5%. В проектному відділі порушенні вимоги, освітленість робочої поверхні складає 370 лк , а коефіцієнт освітленості складає 1.2%. Природне світло проникає у приміщення через бічні світло прорізи. Вікна мають жалюзі. Штучне освітлення виконано у вигляді переривчастих ліній світлодіодних світильників, розташованих паралельно лінії зору інженера-проектувальника. Для місцевого освітлення використовувати галогенні лампи розжарювання

Виробничий шум. Шум на робочому місці створюється: комп'ютером та периферійним пристроєм. Допустимі рівні звукового тиску на робочому місці повинні відповідати вимогам [46]:

Таблиця 4.2

Санітарні норми виробничого шуму, ультразвуку та інфразвуку

Вид трудової діяльності, робоче місце	Рівні шуму та еквівалентні рівні шуму, ДБА, дБАекв
Конструювання та проектування.	50

Реальний рівень шуму в проектному відділі становить 54 дБ, що перевищує допустимий рівень.

Для зменшення рівня шуму рекомендується використовувати місцеву та загальну звукоізоляцію, шумопоглинаючі екрани, поглинаючі фільтри.

4.2. Організаційні та конструктивно-технологічні заходи для зниження впливу шкідливих виробничих факторів

Нормалізація повітря робочої зони. Для створення й автоматичної підтримки в ІТ відділі незалежно від зовнішніх умов оптимальних значень температури, вологості, чистоти і швидкості руху повітря, у холодний час року використовується водяне опалення, у теплий час року застосовується кондиціонування повітря [47].

Виробниче освітлення. Під час аналізу освітлення на робочому місці програміста було встановлено, що воно не відповідає встановленим нормам, тому для покращення умов праці рекомендуємо збільшити рівень загальної освітленості приміщення шляхом встановлення 5 додаткових світильників, щоб загальна кількість лам відповідала розрахованому вище значенню, а саме 36 світлодіодних ламп. Також для підтримки запроєктованого освітлення у чистому виді необхідно скласти графік, де передбачити очищення віконних блоків і світильників не менше 2 разів на рік [48].

Електробезпека. Електробезпечність у приміщенні ІТ відділу пропоную забезпечити наступними технічними способами і засобами захисту:

- для зменшення накопичення статичної електрики застосовувати зволожувачі і нейтралізатори, антистатичне покриття підлоги;

- забезпечити приєднання металевих корпусів устаткування до жили, що заземлює. Заземлення корпусу ПК забезпечити підведенням жили, що заземлює, до розеток. Опір заземлення 4 Ом, згідно (ПУЕ) для електроустановок з напругою до 1000 В. А також організаційними заходами:

- своєчасне проведення інструктажів з техніки безпеки [49].

Ергономіка та організація робочого місця. Після проведення аналізу робочого місця програміста в ІТ Відділі було з'ясовано, що воно відповідає встановленим вимогам.

Виходячи з результатів аналізу важкості та напруженості праці пропоную скоротити час роботи за комп'ютером, робити перерви сумарний час яких повинен скласти 50 хвилин при 8-ми годинному робочому дні [50].

4.2.1. Розрахунок повітрообміну за надлишком тепла у проектному відділі

Приміщення має розміри 4×8×4, яке розміщується на другому поверсі п'ятиповерхового будинку з південного боку. Площа вікон $F = 2,88 \text{ м}^2$. На вікнах розміщені жалюзі. У приміщенні 5 інженерів-проектувальників, розташовано $N_{\text{пк}} = 5$ персональних комп'ютерів та принтер. Для штучного освітлення використовується 4 офісних світлодіодних світильника потужністю 125 Вт.

1. Розраховуємо загальну кількість тепла:

$$Q_{\text{над}} = Q_{\text{осв}} + Q_{\text{облад}} + Q_{\text{ін-пр.}} + Q_{\text{рад}}, \text{ Вт} \quad (4.1)$$

$Q_{\text{над}}$ – загальна кількість тепла

$Q_{\text{осв}}$ - кількість тепла від джерел штучного освітлення

$Q_{\text{облад}}$ - кількість тепла від обладнання

$Q_{\text{ін-пр.}}$ - кількість тепла від інженерів-проектувальників

$Q_{\text{рад}}$ - кількість тепла від сонячної радіації

2. Розраховуємо кількість тепла від джерел штучного освітлення:

$$Q_{\text{осв}} = N \cdot \eta, \quad (4.2)$$

де N - сумарна потужність джерел освітлення, Вт; η - коефіцієнт теплових витрат ($\eta = 0,55$ – для світлодіодних ламп).

$$Q_{\text{осв.}} = 125 \cdot 4 \cdot 0,55 = 275 \text{ Вт}$$

2. Розраховуємо кількість тепла при роботі обладнання: 5 комп'ютерів і принтера (в режимі друку):

$$Q_{облад} = n \cdot P_{комп.} + P_{пр.}, \quad (4.3)$$

де n – кількість комп'ютерів (обладнання);

$P_{комп}$ – встановлена потужність комп'ютерів, $P_{комп} = 400$ Вт

$P_{пр.}$ – потужність принтера в режимі друку, $P_{пр.} = 465$ Вт

$$Q_{облад} = 5 \cdot 400 + 465 = 2.5 \text{ кВт}$$

3. Розраховуємо кількість тепла від інженерів-проектувальників:

$$Q_{ін-пр.} = n \cdot q, \text{ Вт} \quad (4.4)$$

n – кількість інженерів-проектувальників

q – кількість тепла, що виділяється одним інженером-проектувальником

Кількість тепла, що виділяється одним інженером-проектувальником, який виконує легку фізичну роботу дорівнює 99 Вт.

$$Q_{ін-пр} = 5 \cdot 99 = 495 \text{ Вт}$$

4. Розраховуємо кількість тепла від сонячної радіації:

$$Q_{рад} = m \cdot S \cdot k \cdot q_{скл} \quad (4.5)$$

де m – число вікон; $S_{вікна}$ – площа одного вікна, $S_{вікна} = 2,88 \text{ м}^2$;

k – коефіцієнт, віконного переплетення: $k = 0,6$ матові;

$q_{скл.}$ – надходження тепла через 1 м^2 вікна при різній орієнтації вікон: $q_{скл.} = 150$ – південь;

$$Q_{рад} = 1 \cdot 2,88 \cdot 0,6 \cdot 150 = 259,2 \text{ Вт}$$

5. Загальна кількість тепла в проектному відділі:

$$Q_{над} = Q_{осв} + Q_{облад} + Q_{ін-пр.} + Q_{рад} = 275 + 2500 + 495 + 259,2 = 3,529 \text{ кВт}$$

6. Потрібний повітрообмін за надлишком тепла:

$$L = \frac{Q}{c \cdot \rho \cdot (t_{вид} - t_{зовн})}, \text{ м}^3/\text{год} \quad (4.6)$$

Q - кількість тепла, яке виділяється в приміщення за годину, Дж:

$$Q = 3600 \cdot Q_{над} = 3600 \cdot 3529 = 12704 \text{ Вт} = 5328 \text{ кДж};$$

c – теплоємність повітря, Дж/кг (в інтервалі температур від 0°C до 100°C приймається рівною $1,01 \cdot 10^3$ Дж/кг);

ρ – густина повітря, кг/м³ (дорівнює $\rho_{внт} = 1,2$ кг/м³);

$t_{вид}$ – температура повітря, що видаляється, $t_{вид} = 30^\circ\text{C}$

$t_{зовн.}$ - температура повітря, що подається до робочої зони, $t_{зовн.} = 23^\circ\text{C}$

$$L = \frac{5328}{1,01 \cdot 10^3 \cdot 1,2 \cdot (30 - 23)} = 628 \text{ м}^3/\text{год}$$

Оскільки, в проектному відділі підвищена температура повітря на 2 °C від допустимого значення 28°C, встановили механічну вентиляцію з вентилятором VORTICE VARIO , яка забезпечила надходження до приміщення температури повітря 23 °C, дане значення є оптимальним.

4.3. Пожежна безпека

Відповідно до [51] дане приміщення відноситься до категорії В по вибухово-пожежній та пожежній небезпеці із-за використання у ньому твердих горючих матеріалів з температурою спалаху понад 61°C.

Проектний відділ оснащено:

- Двама безпроводними датчиками детектування диму SD-02 (оповіщає при задимленні приміщення; площа обслуговування: до 20 м²);
- двома порошковими вогнегасниками ВП-5 (для приміщення категорії В за відсутності горючих газів і рідин, площею до 50 м² і масою вогнегасної речовини – 5 кг, мінімальна кількість порошкових вогнегасників 2).
- LifeSOS LS-30LR бездротова пожежно-охоронна система (при детектуванні вторгнення, датчики передають на центральний блок сигнал тривоги по радіоканалу без проводів. Централь приймає сигнал від датчиків, включає сирену, відправляє інформацію на пульт централізованого нагляду, дзвонить на зазначені телефонні номери та відправляє SMS повідомлення з повідомленнями про тривогу.)

Для попередження виникнення пожеж проводяться організаційно-технічні заходи пожежної безпеки, які включають:

- включення питань пожежної безпеки у всі інструкції по техніці безпеки;
- виконання встановленого режиму експлуатації електричних мереж та обладнання;
- заборона куріння в недозволеному місці;
- видання необхідних інструктажів, планів евакуації

План евакуації складається з графічної і текстової частин. Графічна частина являє собою схематичний план поверху (рис. 5.1), в якому зеленими суцільними стрілками вказують шляхи евакуації, що ведуть до основних евакуаційних виходів, а пунктирними зеленими стрілками - до аварійних виходів. Двері на шляху евакуації відчиняються назовні у напрямку виходу з будівлі. На плані евакуації умовними знаками показано розміщення вогнегасників, пожежних гідрантів, телефонів, аптечок медичної допомоги, електрощитів, датчиків диму, системи охоронно-пожежної сигналізації.



Умовні позначення											
	- телефон		- пожежний гідрант		- аптечка		- евакуаційний вихід		напрямок руху до виходу		- датчики диму
	- Вогнегасник		- електрошитога		- місце для куріння		- запасний вихід		місце інженера-проектувальника		- охоронно-пожежна система
									шлях до евакуаційного виходу		
									шлях до запасного виходу		

Рис 4.1. План евакуації 2 поверх

4.4. Інструкція з охорони праці при роботі з персональним комп'ютером

Вимоги безпеки перед початком роботи.

- Перед початком роботи працівник повинен зовнішнім оглядом перевірити цілісність корпусів системного блоку, відео монітора, принтера, клавіатури.
- Перевірити цілісність кабелів живлення, місць їх підключення (розеток електромережі, продовжувачів електромережі, розгалужувальних коробок, штепсельних вилок).

- Підготувати своє робоче місце, прибравши речі, які можуть заважати при виконанні роботи.

- Ввімкнути живлення ПК.

- У випадку, якщо після ввімкнення ПК не проходить загрузка або комп'ютер не виходить на робочий режим, працівник повинен повідомити керівника чи спеціаліста відділу інформаційних технологій.

- При виявленні ушкодження або яких-небудь інших недоліків повідомити безпосереднього керівника. Не приступати до роботи без його вказівки.

Вимоги безпеки під час роботи

- Необхідно стійко розташувати всі складові пристрої на столі, в тому числі і клавіатуру. Разом з тим повинна бути передбачена можливість переміщення клавіатури. Її розташування і кут нахилу повинні відповідати побажанням користувача ПК. Якщо в конструкції клавіатури не передбачений простір для опору долонь, то її слід розташувати на відстані не менше 100 мм від краю столу в оптимальній зоні моніторного поля. При роботі на клавіатурі слід сидіти прямо, не напружуватись.

- Для зменшення несприятливого впливу на користувача пристроїв типу "миша" (вимушена поза, необхідність постійного контролю за якістю дій) слід забезпечити вільною більшу площу поверхні столу для переміщення "миші" і зручного упору ліктьового суглоба.

- Не припустимі сторонні розмови, роздратовуючи шуми тощо.

- Періодично при вимкнутому ПК слід видаляти злегка зволоженою мильним розчином хлопко-паперовою салфеткою пил з поверхонь апаратури. Екран і захисний екран протирають ватою, зволоженою спиртом.

- Не дозволяється використовувати рідинні або аерозольні засоби чистки поверхонь ПК.

Забороняється:

- самостійно ремонтувати апаратуру, в яких кінескоп та інші елементи можуть знаходитись під високою напругою (до 25 кВ0.)

- класти будь-які речі на апаратуру ПК, бутерброди та напої на клавіатуру або поруч з нею. Це може вивести її з ладу;

- затуляти вентиляційні отвори в апаратурі, це може призвести до її перегріву і виходу з ладу.

- Для зменшення негативного впливу на стан здоров'я працівників різних факторів ризику, пов'язаних з роботою на ПК, передбачаються додаткові регламентовані перерви для відпочинку користувачів ПК:

- через кожний час безперервної роботи – 10 хвилин;

- через кожні 2 години – 15 хвилин.

- При можливості слід чергувати зміну діяльності з іншою, не пов'язаною з роботою на ПК.

- З метою зменшення негативного впливу монотонності доцільно застосовувати чергування операцій введення тексту і введення даних (зміна змісту і темпу роботи) і т.п.

- При роботі на лазерних принтерах:

- Розташовувати принтер необхідно поряд з системним блоком так, щоб з'єднувальні шнури не були натягнуті. Забороняється ставити принтер на системний блок.

- Перш, ніж програмувати роботу принтера, впевніться, що він знаходиться в режимі зв'язку з системним блоком.

- Для досягнення високоякісного, чистого, з високою роздільною здатністю зображення щоб не зіпсувати апарат, потрібно використовувати папір, марка якого вказана в інструкції до принтера (найчастіше папір вагою 60-135 г/м², типу Canon або Хerox 4024).

- Обрізання країв паперу повинно бути виконаним гострим лезом ножа, без заусенців – це зменшить вірогідність загинання паперу.

- При виконанні роботи (більше 20 хвилин), коли втручання користувача в роботу програми не потрібне, бажано вимикати живлення відео монітора.

- Для підтримки загального тону м'язів, профілактики кістково-м'язових порушень, зорового дискомфорту та інших несприятливих суб'єктивних почуттів під час регламентованих перерв необхідно виконувати комплекси рекомендованих вправ для очей, для хребта, для рук.

- Кількість мікро пауз до 1-2 хвилин слід визначити індивідуально. Форма та зміст перерв можуть бути різними виконання допоміжних робіт, не пов'язаних з роботою ПК, приймання їжі, виконання рекомендованих вправ.

- Виконання фізичних вправ протягом дня рекомендується індивідуально, залежно від почуття втоми. Гімнастика повинна біти на корекцію вимушеної пози покращення кровообігу, часткову компенсацію, дефіциту рухової активності.

- Про виявлені несправності (іскріння, пробоїв, запаху гару, ознак горіння тощо) негайно припинити роботу, відключити все обладнання від електромережі і терміново повідомити безпосереднього керівника або спеціаліста по ремонту ПК.

Вимоги безпеки при закінчення роботи на ПК.

- Закінчити і зберегти в пам'яті ПК файли, які знаходились у роботі. Виконати всі дії для коректного завершення роботи в оперативній системі.

- Вимкнути принтер та інші периферійні пристрої, вимкнути системний блок. При наявності пристрою безперебійного живлення (ПБЖ) вимкнути його живлення.

- Вимкнути ПК кнопкою «POWER» (ЖИВЛЕННЯ) та вийняти штепсельну вилку кабелю живлення з розетки

- Накрити клавіатуру кришкою для попередження попадання в неї пилу.

- Навести порядок на робочому місці.

Вимоги безпеки в аварійних ситуаціях.

- Якщо після ввімкнення ПК відчувається запах горілого або при доторканні до металевих частин ПК відчувається дія електричного струму, потрібно негайно відключити ПК від електромережі та повідомити про це своєму керівникові.

- У випадку виникнення пожежі негайно розпочати гасіння наявними засобами пожежогасіння і повідомити за телефоном 101 (міська пожежна охорона) та начальнику ДПД підприємства. Пам'ятайте, що загашувати електроустановки слід вуглекислотними вогнегасниками, сухим піском, щоб уникнути ураження електричним струмом.

У разі виникнення інших аварійних ситуацій слід припинити роботу і повідомити про це керівника робіт.

ВИСНОВКИ ДО РОЗДІЛУ 4

На підставі виконаного розрахунку повітрообміну за надлишком тепла, значення якого 628 м³/год, встановили механічну вентиляцію з вентилятором VORTICE VARIO, оскільки використання природної вентиляції є малоефективним. Механічна вентиляція здатна забезпечити виведення з проектного відділу температури 30°C і підтримувати температуру повітря допустимого та навіть оптимального значення.

РОЗДІЛ 5

ОХОРОНА НАВКОЛИШНЬОГО СЕРЕДОВИЩА

5.1. Аналіз впливу техногенних чинників

Широке використання електричного та електронного обладнання дозволило не тільки підвищити якість життя людей, але й призвело до негативних наслідків для навколишнього середовища та здоров'я людини. Можна виділити основні шкідливі та небезпечні чинники, які впливають на навколишнє середовище [52]:

- шумове забруднення;
- вібраційне забруднення;
- електромагнітне забруднення
- теплове забруднення
- радіаційне забруднення

Шумове забруднення. У сучасному світі в умовах науково-технічного прогресу шум став однією з форм фізичного (хвильового) забруднення природного середовища. Шумом прийнято вважати усі неприємні та небажані звуки або їх сукупність, які заважають нормально працювати, сприймати потрібну звукову інформацію та відпочивати.

Адаптація до нього практично неможлива. Фоновий рівень шуму навколишнього середовища становить 30-60 децибел. До цього природного фону за сучасних умов додаються виробничі й транспортні шуми, рівень яких нерідко перевищує 100 децибел. Джерелами шуму є: промислові об'єкти, транспорт, гучномовні пристрої, телевізори, радіоприймачі, музичні інструменти, юрби людей тощо. Шум у виробничих умовах негативно впливає на працівника: послаблює увагу, посилює розвиток втоми, сповільнює реакцію на небезпеку. Внаслідок цього знижується працездатність та підвищується ймовірність нещасних випадків. Допустимі рівні звукового тиску в октавних смугах частот на робочих місцях у виробничих приміщеннях наведені в таблиці 5.1 [52]:

Допустимі рівні звукового тиску в октавних смугах частот

Рівні звукового тиску в дБ, в октавних смугах частот, Гц								
31,5	63	125	250	500	1000	2000	4000	8000
107	95	87	82	78	75	73	71	69

Встановлено, що рослини під впливом шуму знижують енергію до зростання, у них спостерігається надмірне (навіть повне, що призводить до загибелі) виділення вологи через листя, можливі порушення у клітинах. Гинуть листя і квіти рослин, які розташовані близько до джерела інтенсивного шуму (звуку). Відсутність шуму особливо необхідний для тварин, які обмінюються звуковою інформацією, а також аналізуючи звуки навколишнього середовища з метою покращання отримання інформації, в тому числі сигналів тривоги. Аналогічно діє шум на тварин. Від шуму реактивного літака гинуть личинки бджіл, самі

вони втрачають здатність орієнтуватися, у пташиних гніздах дає тріщини шкаралупа яєць. Від коливань повітря, які утворюються звуками переносної радіоапаратури, не можуть піднятися у повітря жуки, джмелі та інші комахи.

Вібраційне забруднення. Вібрація – це механічні коливання твердого тіла. Вібрацію поділяють на природну та штучну. Джерелами природної вібрації є землетруси, що викликаються природними чинниками. Джерелами штучної вібрації є промисловість, транспорт. Тривалі вібрації завдають великої шкоди здоров'ю людини – від сильної втоми до змін багатьох функцій організму: порушення серцевої діяльності, нервової системи, спазмів судин, деформації м'язів, струсу головного мозку тощо. Особливо небезпечна вібрація з частотою, яка є резонансною з частотою коливання окремих органів чи частин тіла людини, що може призвести до їх пошкодження. Тривала дія вібрації може спричинити професійне захворювання – вібраційну хворобу [52].

Електромагнітне забруднення. У процесі еволюції біосфера постійно знаходилася і знаходиться під впливом електромагнітного поля (ЕМП) природного похо-

дження (природний фон): електричного й магнітного поля Землі, космічного електромагнітного випромінювання, насамперед того, що генерується Сонцем. У період науково-технічного прогресу людство створювало і дедалі ширше використовувало штучні (антропогенні) джерела ЕМП. У наш час ЕМП антропогенного походження значно перевищують природний фон і є тим несприятливим чинником, вплив якого на людину та довкілля рік за роком зростає. Ступінь впливу ЕМП на організм людини залежить від діапазону частот, інтенсивності та тривалості дії, характеру випромінювання (неперервного чи модульованого), режиму опромінювання, розміру поверхні тіла, що зазнає опромінювання, індивідуальних особливостей організму. Електромагнітні поля можуть викликати біологічні та функціональні порушення у функціонуванні організму. Функціональні ефекти проявляються у передчасній втомлюваності, частих болях голови, погіршенні сну, порушенні функцій серцево-судинної та центральної нервової систем. Тривалий та інтенсивний вплив ЕМП призводить до стійких порушень та захворювань. Біологічні негативні ефекти впливу ЕМП проявляються у тепловій та нетепловій діях. Теплова дія призводить до підвищення температури тіла та місцевого вибіркового нагрівання органів і тканин організму внаслідок переходу електромагнітної енергії в теплову. Таке нагрівання особливо небезпечне для органів із слабкою терморегуляцією (головний мозок, очі, нирки, шлунок тощо). Наприклад, випромінювання сантиметрового діапазону призводить до появи катаракти, тобто до поступової втрати зору [52].

Теплове забруднення. Теплове забруднення – це результат розсіювання в навколишнє середовище теплоти, яка виділяється у багаточисельних теплових процесах, насамперед пов'язаних зі згоранням палива. Під час згорання палива щорічно витрачається до 23% кисню, що утворюється в процесі фотосинтезу на Землі за рік. За підрахунками під час спалювання вугілля в навколишнє середовище викидається радіоактивних компонентів більше, ніж за той самий час на всіх атомних електростанціях у разі безаварійної роботи. Теплове забруднення гідросфери відбувається переважно внаслідок скидання у водойми підігрітих вод від ТЕС, АЕС та інших енергетичних об'єктів. Тепла вода змінює термічні та біологічні режими водойм і шкідливо впливає на їхніх мешканців [52].

5.2. Вплив приймальних пристроїв на навколишнє середовище

Абонентський приймач – телевізійний приймач (приставка), пристрій, що приймає сигнал цифрового телебачення, декодує його і перетворює в аналоговий сигнал для виведення через роз'єми RCA або SCART або перетворює в цифровий сигнал для виведення через роз'єм HDMI , і передає його далі на телевізор.

Перехід до цифрового телебачення призвів до зростання виробництва цифрових абонентських приймачів, що в свою чергу може негативно впливати на навколишнє середовище. Приймач продукує слабкі електричні і магнітні змінні поля в широкому діапазоні частот. Проте проблема впливу електромагнітних випромінювань, що продукуються заслуговує на особливу увагу. Наукові дослідження показали, що ЕМВ мають у своєму складі чинник, котрий впливає на користувачів при наявності сучасних екранів від ЕМВ. Вчені України ідентифікували цей чинник як торсіонові поля, котрі супроводжують будь-яке електромагнітне випромінювання та являються його інформаційною компонентою [53]. Робоча група Всесвітньої організації охорони здоров'я з гігієнічних аспектів користування моніторами та радіо терміналами виявили порушення стану здоров'я при користуванні пристроями, які мають електромагнітне випромінювання, найсерйозніші з яких:

- погіршення зору;
- порушення імунної системи;
- порушення психоемоційної сфери (стресовий синдром, агресивність)

Для забезпечення безпеки здоров'я користувачів в Україні діють Державні санітарні норми і правила при роботі з джерелами електромагнітних полів «ДСанПіН 3.3.6.096-2002». Значення ГДР напруженості електричної ($E_{гд}$) і магнітної ($H_{гд}$) складових залежно від тривалості їх дії наведені в таблиці 5.2.

Таблиця 5.2.

Значення ГДР напруженості електричної ($E_{гд}$) і магнітної ($H_{гд}$) складових

Час перебування персоналу, год	$E_{гд}$, В/м					$H_{гд}$, А/м			
	1-10 кГц	10-60 кГц	0,063-30 МГц	3-30 МГц	30-300 МГц	1-10 кГц	10-60 кГц	0,063-3 МГц	30-50 МГц
8	120	70	50	30	10	9	7	5	0,3
7	130	75	53	32	11	9,8	7,5	5,3	0,32
6	140	82	58	34	12	10,6	8,1	5,8	0,34
5	155	90	63	37	13	11,6	8,8	6,3	0,38
4	175	110	71	42	14	13	9,9	7,1	0,42
3	200	115	82	48	16	15	11,4	8,2	0,49
2	250	140	100	59	20	18,4	14	10	0,6
1	350	200	141	84	28	26	19,7	14,2	0,85
0,5	500	280	200	118	40	37,6	27,9	20	1,2

У результаті дії на організм людини електромагнітних випромінювань в діапазоні 30 кГц - 300 МГц (НЧ) спостерігається: загальна слабкість, підвищена втома, сонливість, порушення сну, головний біль та біль в ділянці серця. З'являється роздратованість, втрачається увага, сповільнюються рухово-мовні реакції. Виникає ряд симптомів, які свідчать про порушення роботи окремих органів - шлунку, печінки, підшлункової залози.

Для того, щоб зменшити рівень електромагнітного випромінювання потрібно обмежити безперервний час роботи абонентського приймача [54-55].

В Україні норми електромагнітної безпеки регламентуються Державними санітарними нормами і правилами захисту населення від впливу електромагнітного випромінювання, згідно з якими допустимі рівні інтенсивності електромагнітного випромінювання для цивільного населення становлять $2,5 \text{ мкВт/см}^2$.

Абонентський приймач під час роботи створює шум, рівень якого становить 54 дБ. Допустимий рівень звукового тиску повинний відповідати «ДСН 3.3.6.037-99 Санітарні норми виробничого шуму, ультразвуку та інфразвуку», а саме 50 дБ.

Велика кількість звукових сигналів, що поступають до кори головного мозку, викликають переживання, страх, передчасну втому. Дія шуму на людину виражається в широкому діапазоні - від суб'єктивного роздратування до об'єктивних змін в ЦНС, органах слуху, серцево-судинних та ендокринній системах, травному акті та інших органів і систем. Першим показником шкідливої дії шуму є скарги на роздратованість, переживання, порушення сну [56].

4.3. Засоби для захисту від електромагнітного випромінювання та шуму, проблема електронних відходів

Захист від електромагнітного випромінювання. Для зменшення впливу ЕМП на персонал та населення, яке знаходиться у зоні дії радіоелектронних засобів, потрібно вжити ряд захисних заходів. До їх числа можуть входити організаційні, інженерно-технічні та лікарсько-профілактичні.

До заходів щодо зменшення впливу на працівників ЕМП належать: організаційні, інженерно-технічні та лікарсько-профілактичні.

Організаційні заходи здійснюють органи санітарного нагляду. Вони проводять санітарний нагляд за об'єктами, в яких використовуються джерела електромагнітних випромінювань.

Інженерно-технічні заходи передбачають таке розташування джерел ЕМП, яке б зводило до мінімуму їх вплив на працюючих, використання в умовах виробництва дистанційного керування апаратурою, що є джерелом випромінювання, екранування джерел випромінювання, застосування засобів індивідуального захисту (халатів, комбінезонів із металізованої тканини, з виводом на заземлюючий пристрій). Для захисту очей доцільно використовувати захисні окуляри ЗП5-90. Скло окулярів вкрито напівпровідниковим оловом, що послаблює інтенсивність електромагнітної енергії при світлопропусканні не нижче 75%.

Взагалі, засоби індивідуального захисту необхідно використовувати лише тоді, коли інші захисні засоби неможливі чи недостатньо ефективні: при проходженні через зони опромінення підвищеної інтенсивності, при ремонтних і налагоджувальних роботах в аварійних ситуаціях, під час короткочасного контролю та при зміні інтенсивності опромінення. Такі засоби незручні в експлуатації, обмежують можливість виконання трудових операцій, погіршують гігієнічні умови.

У радіочастотному діапазоні засоби індивідуального захисту працюють за принципом екранування людини з використанням відбиття і поглинання ЕМП. Для захисту тіла використовується одяг з металізованих тканин і рідіопоглинаючих матеріалів. Металізовану тканину роблять із бавовняних ниток з розміщеним всередині них тонким проводом, або з бавовняних чи капронових ниток, спірально обвитих металевим дротом. Така тканина, наче металева сітка, при відстані між нитками до 0,5 мм значно послаблює дію випромінювання. При зшиванні деталей захисного одягу треба забезпечити контакт ізольованих проводів. Тому електрогерметизацію швів здійснюють електропровідними масами чи клеями, які забезпечують гальванічний контакт або збільшують ємнісний зв'язок безконтактних проводів.

Лікарсько-профілактичні заходи передбачають проведення систематичних медичних оглядів працівників, які перебувають у зоні дії ЕМП, обмеження в часі перебування людей в зоні підвищеної інтенсивності електромагнітних випромінювань, видачу працюючим безкоштовного лікарсько-профілактичного харчування, перерви санітарно-оздоровчого характеру.

Захист від шуму. Для зменшення і ліквідації шуму застосовується цілий комплекс заходів, що називається шумозахистом. Це застосування звукопоглинаючих матеріалів, раціональне розміщення будівельних об'єктів, створення вздовж вулиць екранів у вигляді земляних валів, стін різних конструкцій, шумовідбиваючих, як правило не житлових будівель - магазинів, складів, гаражів.

Проблема електронних відходів. Згідно Закону України «Про відходи» з метою запобігання або зменшення обсягів утворення відходів потрібно здійснювати системи збирання та утилізації електричного та електронного обладнання [55-56]. Вирі-

шення проблеми електронних відходів в Україні мав би забезпечити «Технічний регламент з поводження з відходами електричного та електронного обладнання», розробка якого в Україні здійснюється з 2008 року. Згідно з проектами цих законодавчих актів імпортери і виробники можуть як самостійно утилізувати електровідходи, так і підписувати договори на виконання робіт з організації збирання, заготівлі та утилізації відповідних видів техніки з уповноваженими підприємствами. Розроблено також проект Постанови Кабінету Міністрів України «Про затвердження Технічного регламенту з поводження з відходами електронного та електричного устаткування». Цим регламентом передбачається створення пунктів збору відходів електронного та електричного обладнання, які повинні розташовуватися у місцях, зручних для користувачів, та забезпечувати безоплатність послуг, що надаються цими пунктами для користувачів. Наразі обговорюється ще один варіант вирішення проблеми, а саме проект внесення змін до Податкового Кодексу, в якому передбачає централізоване стягнення коштів з імпортерів та виробників різних споживчих товарів з метою забезпечення за рахунок цих коштів належної організації збирання, заготівлі та утилізації відходів від зазначених товарів.

Однак, загалом проблему електронних відходів в Україні необхідно вирішити як в організаційно-правовому аспекті – створення фондів виробників, підтримка держави підприємств з утилізації відходів, так і в соціально-інформаційному: українців треба переконати в тому, що виносити на звичайний смітник поламаний електронний пристрій – не можна.

ВИСНОВОК ДО РОЗДІЛУ 5

Для мінімізації ризику виникнення захворювань, ефективними є інженерно-технічні заходи, які зменшують дію шкідливих чинників. Також були розглянута проблема електронних відходів, одним зі шляхів вирішення якої є створення пунктів збору відходів електронного та електричного обладнання.

ВИСНОВКИ

У кваліфікаційній роботі вирішено науково-технічне завдання дослідження та розробки методів та засобів розпізнавання мови за допомогою нейронних мереж.

Основні результати роботи укладаються в наступному:

1. Проведено аналіз задачі подання та розпізнавання мови, виділено основні компоненти систем автоматичного розпізнавання мовлення.
2. Розглянуто методи попередньої обробки та виділення ознак мовного сигналу, серед яких обраний підхід, що ґрунтується на знаходженні мел коефіцієнтів.
3. Розглянуто методи розпізнавання мови та обрано апарат прихованих Марківських моделей.
4. Набув подальшого розвитку метод виділення ділянок з мовленням у сигналі, заснований на аналізі розподілу його локальних екстремумів. Створено його програмну реалізацію.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. R. Cowie, E. Douglas-Cowie and N. Tsapatsoulis, "Emotion recognition in human-computer interaction", *IEEE Signal Processing Magazine*, vol. 18, no. 1, pp. 32-80, 2021.
2. B. Reeves and C. Nass, *The media equation: how people treat computers television and new media like real people and places*, Cambridge University Press, 2016.
3. L. L. Yu, Z. X. Cai and M. Y. Chen, "Study on emotion feature analysis and recognition in speech signal an overview", *Journal of Circuits and Systems*, vol. 12, no. 4, pp. 76-84, 2017.
4. J. Nicholson, K. Takahashi and R. Nakatsu, "Emotion recognition in speech using neural networks", *Neural Computing and Applications*, vol. 9, no. 4, pp. 290-296, 2020.
5. C. H. Park and K. B. Sim, "Emotion recognition and acoustic analysis from speech signal", *International Joint Conference on Neural Networks*, pp. 2594-2598, 2013.
6. V. Hozjan and Z. Kacic, "Context-independent multilingual emotion recognition from speech signals", *International Journal of Speech Technology*, vol. 6, pp. 311-320, 2013.
7. W. M. Zheng, M. H. Xin and X. L. Wang, "A novel speech emotion recognition method via incomplete sparse least square regression", *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 569-572, 2014.
8. Y. Attabi and P. Dumouchel, "Anchor models for emotion recognition from speech", *IEEE Trans on Affective Computing*, vol. 4, no. 3, pp. 280-290, 2013.
9. K. X. Wang, N. An and B. N. Li, "Speech emotion recognition using fourier parameters", *IEEE Trans on Affective Computing*, vol. 6, no. 1, pp. 69-75, 2015.
10. M. Abdelwahab and C. Busso, "Domain adversarial for acoustic emotion recognition", *IEEE Trans on Audio Speech and Language Processing*, vol. 26, no. 12, pp. 2423-2435, 2018.

11. P. Song, W. M. Zheng and L. Zhao, "Joint subspace learning and feature selection method for speech emotion recognition", *Journal of Tsinghua University (Science and Technology)*, vol. 58, no. 4, pp. 347-351, 2018.
12. C. Pushpa and M. M. Priya, "A review on deep learning algorithms for speech and facial emotion recognition", *International Journal of Control Theory & Applications*, vol. 9, no. 24, pp. 183-204, 2016.
13. Z. G. Jin, Y. Han and Q. Zhu, "A sentiment analysis model with the combination of deep learning and ensemble learning", *Journal of Harbin Institute of Technology*, vol. 50, no. 11, pp. 1-8, 2018.
14. N. Agarwalla, D. Panda and M. K. Modi, "Deep learning using restricted boltzmann machines", *International Journal of Computer Science & Information Security*, vol. 7, no. 3, pp. 1552-1556, 2016.
15. P. Liu, S. Han, Z. Meng and Y. Tong, "Facial expression recognition via a boosted deep belief network", *IEEE Conference on Computer Vision & Pattern Recognition*, pp. 1805-1812, 2014.
16. G. Ian, B. Yoshua and C. Aaron, *Deep learning*, Beijing: Posts & Telecom Press, 2017.
17. Z. Y. Han, S. X. Lun and J. Wang, "Speech emotion recognition based on Genetic Wavelet Neural Network", *Computer Technology and Development*, no. 1, pp. 75-78, 2013.
18. O.S. Agashin and O.N. Korelin, "Methods of digital processing of a speech signal in the problem of recognizing isolated words with the use of signal processes" in *Proceedings of the Nizhny Novgorod State Technical University*, vol. 4, pp. 32-44, 2019.
19. L. Rabiner and R. Shafer, *Digital processing of speech signals*, Moscow: Radio and communication, 2021.
20. K.A. Makovskin, "Hybrid models: hidden Markov models and neural networks their application in speech recognition systems" in *Computing Center*, Moscow: A. A. Dorodnitsyn, pp. 40-95, 2016.

21. Z. Guohua, "Ant Colony Clustering Algorithm and Improved Markov Random Fusion Algorithm in Image Segmentation of Brain Images", *International Journal Bioautomation*, vol. 20, no. 4, pp. 505-514, 2016.
22. Li Junbing, "Application of BP Neural Network Algorithm in Biomedical Diagnostic Analysis", *International Journal Bioautomation*, vol. 20, no. 3, pp. 417-426, 2016.
23. M. Schultz, *MATLAB 14. Programming numerical methods*, Moscow: BHV-Petersburg, pp. 928, 2016.
24. L.F. Chaparro, *Signals and Systems Using MATLAB*, pp. 768, 2021.
25. D. Chikhachev, *World market forecasts and research*, [online] Available: <http://www.techportal.ru/marketsandmarkets/>.
26. S. Biryukov, *The neural network zoo* [Online], Oct. 2016.
27. S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural Comput.*, vol. 9, pp. 1735-1780, 1997.
28. M. Xu, F. Zhang and S. U. Khan, "Improve accuracy of speech emotion recognition with attention head fusion", *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 1058-1064, 2020.
29. Y. Zhang, J. Du, Z. Wang, J. Zhang and Y. Tu, "Attention based fully convolutional network for speech emotion recognition", *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1771-1775, 2018.
30. S. Mirsamadi, E. Barsoum and C. Zhang, "Automatic speech emotion recognition using recurrent neural networks with local attention", *2017 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 2227-2231, 2017.
31. Y. Li, T. Zhao and T. Kawahara, "Improved end-to-end speech emotion recognition using self attention mechanism and multitask learning", *INTERSPEECH*, pp. 2803-2807, 2019.
32. E. Takeishi, T. Nose, Y. Chiba and A. Ito, "Construction and analysis of phonetically and prosodically balanced emotional speech database", *2016 Conference*

of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA), pp. 16-21, 2016.

33. Zhaojuan Song, English speech recognition based on deep learning with multiple features[J], Springer Vienna, vol. 102, no. 1, 2020.

34. Thales Aguiar de Lima and Márjory Da Costa-Abreu, A survey on automatic speech recognition systems for Portuguese language and its variations[J], Elsevier Ltd, vol. 62, 2020.

35. Aye Nyein Mon, Win Pa Pa and Ye Kyaw Thu, Improving Myanmar Automatic Speech Recognition with Optimization of Convolutional Neural Network Parameters[J], AIRCC Publishing Corporation, vol. 7, no. 6, 2018.

36. Wang Junchao, Huang Hao, Xu Haihua and Hu Ying, "Speech recognition of low-resource uyghur language based on transfer learning[J]", Computer engineering, vol. 44, no. 10, pp. 281-285, 2018.

37. "Sensitivity and specificity of the Percentage of Consonants Correct-Revised in the identification of speech sound disorder[J]", CoDAS, vol. 29, no. 3, 2017.

38. A. Graves, A.R. Mohamed and G. Hinton, «Speech recognition with deep recurrent neural networks», Acoustics speech and signal processing (icassp) 2013 ieee international conference on, pp. 6645-6649, May 2013.

39. A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, et al., Deep speech: Scaling up end-to-end speech recognition, 2014.

40. G. Dhande and Z. Shaikh, "Recurrent Neural Networks for End-to-End Speech Recognition: A Comparative Analysis", International Journal on Recent and Innovation Trends in Computing and Communication, vol. 6, no. 4, pp. 88-93, 2018.

41. НПАОП 0.00-1.28-10 Правила охорони праці під час експлуатації електронно-обчислювальних машин.

42. ДСН 3.3.6.037-99 «Санітарні норми виробничого шуму, ультразвуку та інфразвуку».

43. Державні санітарні норми та правила «Гігієнічна класифікація праці за показниками шкідливості та небезпечності факторів виробничого середовища, важкості та напруженості трудового процесу».
44. «ДСН 3.3.6.042-99 Санітарних норми мікроклімату виробничих приміщень».
45. ДБН 13.2.5-28-2006 «Природне і штучне освітлення».
46. ДСН 3.3.6.037-99 «Санітарні норми виробничого шуму, ультразвуку та інфразвуку».
47. ДСТУ 12.1.005-88 «ССБП. Загальні санітарно-гігієнічні вимоги до повітря робочої зони».
48. ДБН В.2.5-28-2006 «Інженерне обладнання будинків і споруд. Природне і штучне освітлення».
49. ДСТУ Б В.2.5-82:2016 «Електробезпека в будівлях і спорудах. Вимоги до захисних заходів від ураження електричним струмом».
50. ДСТУ 8604:2015 «Дизайн і ергономіка. Робоче місце для виконання робіт у положенні сидячи. Загальні ергономічні вимоги».
51. НАПБ Б.03.002-2007 «Норми визначення категорій приміщень, будинків та зовнішніх установок за вибухопожежною та пожежною небезпекою».
52. Прогнозування екологічних ризиків з використанням аналізу ієрархів та теорії нечітких множин: міжнародна науково-практична конференція «І-й всеукраїнський з'їзд екологів»: Тези доповідей. Україна, м. Вінниця, 4-7 жовтня 2016. – 2016. – С.25.
53. Клап Я. А., Яремкевич О. С., Червецова В. Г., Заярнюк Н. Л., Новіков В. П., Дослідження впливу електромагнітних, постійних магнітних та акустичних полів на організм людини // Вісник Нац. ун-ту “Львівська політехніка”. – 2016 – № 812. – С. 365–372.
54. Сучасний стан досліджень впливу електромагнітних випромінювань на організм людини [Електронний ресурс]/[А. П. Чорний, В. В. Никифоров, Д. І. Ро-

дьякін, В. Ю. Ноженко] // Інженерні та освітні технології в електротехнічних та комп'ютерних системах: щоквартальний науково-практичний журнал. – Кременчук: КрНУ, 2013.

55. Екологія та охорона навколишнього природного середовища: навч. посібник для вузів / В. С. Джигирей. - 6-те вид., випр. і доп. - К. : Знання, 2017. - 422 с.

56. Боротьба з шумом на виробництві: Довідник / Під ред. О. Я. Юдіна. – М: Машинобудування, 2015. – 297 с.