

# КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр

(назва освітнього ступеня)

на тему: Метод обробки сигналів для формування команд систем  
ГОЛОСОВОГО керування

Виконав(ла): студент(ка) 6 курсу, групи РРм-61  
спеціальності 172 Телекомунікації та радіотехніка

(шифр і назва спеціальності)

(підпис)

Боїло С.Т.

(прізвище та ініціали)

Керівник

(підпис)

Дедів І.Ю.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Хвостівська Л.В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Дунець В.Л.

(прізвище та ініціали)

Рецензент

(підпис)

Тимків П.О.

(прізвище та ініціали)

Міністерство освіти і науки України  
Тернопільський національний технічний університет імені Івана Пулюя

Факультет прикладних інформаційних технологій та електроінженерії  
(повна назва факультету)

Кафедра радіотехнічних систем  
(повна назва кафедри)

ЗАТВЕРДЖУЮ

Завідувач кафедри

Дунець В.Л.  
(підпис) (прізвище та ініціали)

« 06 » грудня 2022 р.

**ЗАВДАННЯ  
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня магістр  
(назва освітнього ступеня)

за спеціальністю 172 Телекомунікації та радіотехніка  
(шифр і назва спеціальності)

студенту Боїлу Святославу Тарасовичу  
(прізвище, ім'я, по батькові)

1. Тема роботи Метод обробки сигналів для формування команд систем голосового керування

Керівник роботи Дедів Ірина Юріївна, к.т.н., доц.  
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «06» грудня 2022 року № 4/7-989

2. Термін подання студентом завершеної роботи 21 грудня 2022 року

3. Вихідні дані до роботи Технічне завдання, метод обробки сигналів для голосового керування

4. Зміст роботи (перелік питань, які потрібно розробити)

1. Аналітична частина

2. Основна частина

3. Науково-дослідна частина

4. Охорона праці та безпека в надзвичайних ситуаціях

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

1. Задача створення систем голосового керування

2. Приклади систем розпізнавання мовних сигналів

3. Принцип роботи пропонованого способу

4. Висновки



## АНОТАЦІЯ

Боїло С.Т. Метод обробки сигналів для формування команд систем голосового керування. – Рукопис. Кваліфікаційна робота магістра, ТНТУ ім. І.Пулюя, Тернопіль, 2022.

В роботі розглянуто питання розробки методу обробки сигналів для формування команд систем голосового керування. Проаналізовано типову архітектуру систем голосового керування. Розроблено метод обробки голосових сигналів для ідентифікації окремих команд та формування на їх основі сигналів керування. Суть методу зводиться до наступного. Потрібно розпізнати чотири команди в голосовому сигналі: «вліво», «вправо», «взад», «вперед». Для власне опрацювання застосовано метод ковзного вікна. Для визначення тривалості команд проводилось обчислення середнього значення сигналу всередині ковзного вікна та формувалась порогова функція із значенням 1 та 0 залежно від того більше середнє значення за поріг чи менше. Для сегментації сигналу команди на фонемі проведено визначення частоти основного тону методом формантного аналізу та встановлення присутності її в ковзному вікні. Використано порогову функцію яка рівна 1 коли ця частота є присутньою та 0 коли її немає

Ключові слова: голосовий сигнал, метод обробки, команда.

## ANNOTATION

Boilo S.T. Method of signal processing for command generation of voice control systems. - Manuscript. Master's qualification work, TNTU, Ternopil, 2022.

The master's qualification thesis deals with the issue of developing a method of signal processing for the formation of commands for voice control systems. The typical architecture of voice control systems is analyzed. A method of processing voice signals has been developed to identify individual commands and generate control signals based on them. The essence of the method boils down to the following. You need to recognize four commands in the voice signal: "left", "right", "back", "forward". The sliding window method was used for the actual processing. To determine the duration of the commands, the average value of the signal within the sliding window was calculated and a threshold function with a value of 1 and 0 was formed, depending on whether the average value was greater than or less than the threshold. To segment the command signal into phonemes, the frequency of the main tone was determined by the formant analysis method and its presence in the sliding window was established. A threshold function is used, which is equal to 1 when this frequency is present and 0 when it is not

Key words: voice signal, processing method, command.

## ЗМІСТ

ВСТУП.....	7
РОЗДІЛ 1. АНАЛІТИЧНА ЧАСТИНА.....	9
1.1 Задача створення систем голосового керування.....	9
1.2 Голосове керування персональним комп'ютером.....	12
1.3 Аналіз процесу формування мовних сигналів.....	14
1.4 Висновки до розділу 1.....	18
РОЗДІЛ 2. ОСНОВНА ЧАСТИНА.....	20
2.1 Особливості аналогічних систем.....	20
2.2 Приклади систем розпізнавання мовних сигналів.....	22
2.2.1. Біометричний спосіб розпізнавання голосу.....	22
2.2.2. Двохрівневий метод розпізнавання голосової команди.....	28
2.3 Висновки до розділу 2.....	33
РОЗДІЛ 3. НАУКОВО-ДОСЛІДНА ЧАСТИНА.....	34
3.1 Принцип роботи пропонованого способу.....	34
3.2 Результати експериментального дослідження.....	38
3.3 Висновки до розділу 3.....	47
РОЗДІЛ 4. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ.....	48
4.1 Охорона праці.....	48
4.2 Безпека в надзвичайних ситуаціях.....	50
4.3 Висновки до розділу 4.....	52
ВИСНОВКИ.....	53
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	55
ДОДАТКИ	

## ВСТУП

**Актуальність роботи.** Побудова системи голосового керування є актуальним завданням. Ці системи допомагають при роботі з комп'ютеризованими системами, як то голосовий набір тексту, керування електронними приладами в будинку тощо. Актуальною ця технологія є, наприклад, коли йдеться про людей із порушеннями опорно-рухового апарату. Як приклад розглядається задача керування електричним інвалідним кліслом.

Основою для реалізації голосового керування є алгоритми та методи розпізнавання мови. В таких системах проводиться опрацювання голосових сигналів та виділення в їхній структурі ознак, за якими в наступному проводиться ідентифікація чи розрізнення окремих команд, що є певними ключовими словами. Для цього можуть застосовуватись знаки фонемного складу мови, амплітудні та часові параметри голосового сигналу, його склад, переходи між фонемами, складами чи словами.

Результати аналізу поширених методів розпізнавання мови показали, що основні дослідження зосереджено на частотних характеристиках виділення людського вокального каналу, тобто якісних методах, заснованих на обробці кепстрального сигналу.

Недоліком таких підходів є значна обчислювальна складність, неможливість працювати в масштабі реального часу або великі часові затримки.

В дослідженнях проводиться розроблення методу обробки сигналів для формування команд голосового керування.

**Мета.** Розроблення методу обробки сигналів для задачі голосового керування. **Задачі:**

- аналіз задачі голосового керування;
- аналіз способів організації систем голосового керування;

- Розроблення методу опрацювання сигналів для ідентифікації окремих команд;

- експериментальна перевірка працездатності методу.

*Об'єкт дослідження:* процес обробки сигналів для голосового керування.

*Предмет дослідження:* метод обробки голосових сигналів.

**Наукова новизна.** Розроблено метод обробки голосових сигналів для ідентифікації окремих команд.

**Практичне значення отриманих результатів.** Викристання при розробленні автоматизованих систем голосового керування.



## РОЗДІЛ 1

### АНАЛІТИЧНА ЧАСТИНА

#### 1.1 Задача створення систем голосового керування

Побудова системи голосового керування є актуальним завданням. Ці системи допомагають при роботі з комп'ютеризованими системами, як то голосовий набір тексту, керування електронними приладами в будинку тощо. Актуальною ця технологія є, наприклад, коли йдеться про людей із порушеннями опорно-рухового апарату. Як приклад розглядається задача керування електричним інвалідним кріслом.

В таких системах проводиться опрацювання голосових сигналів та виділення в їхній структурі ознак, за якими в наступному проводиться ідентифікація чи розрізнення окремих команд, що є певними ключовими словами. Для цього можуть застосовуватись знаки фонемного складу мови, амплітудні та часові параметри голосового сигналу, його склад, переходи між фонемами, складами чи словами.

Під час створення подібної системи розробник стикається з певними проблемами. По-перше, відсутність математичної моделі семантики мовного сигналу, оскільки з визначення семантики до мовного сигналу можуть застосовуватись лише ймовірнісні і евристичні методи, які не дають точного результату і точність яких обернено пропорційна кількості смислових одиниць. По-друге, індивідуальні характеристики диктора: специфіка вимови, акцент, наголоси тощо. По-третє, робота зі спонтанною мовою та необхідність визначення присутності ключового слова. По-четверте, розбіжності у акустичній обстановці, шуми.

Розглянемо два основних методи голосового керування: метод ковзного вікна та метод моделей заповнювачів. Обидва методи ґрунтуються на алгоритмі розпізнавання мови за допомогою прихованих марківських моделей (ПММ).

Більшість розглянутих систем побудовані за модульною архітектурою.

Як метод розпізнавання більшість сучасних систем використовують метод ПММ. Найчастіше використовуються ПММ із трьома станами (рис. 1).

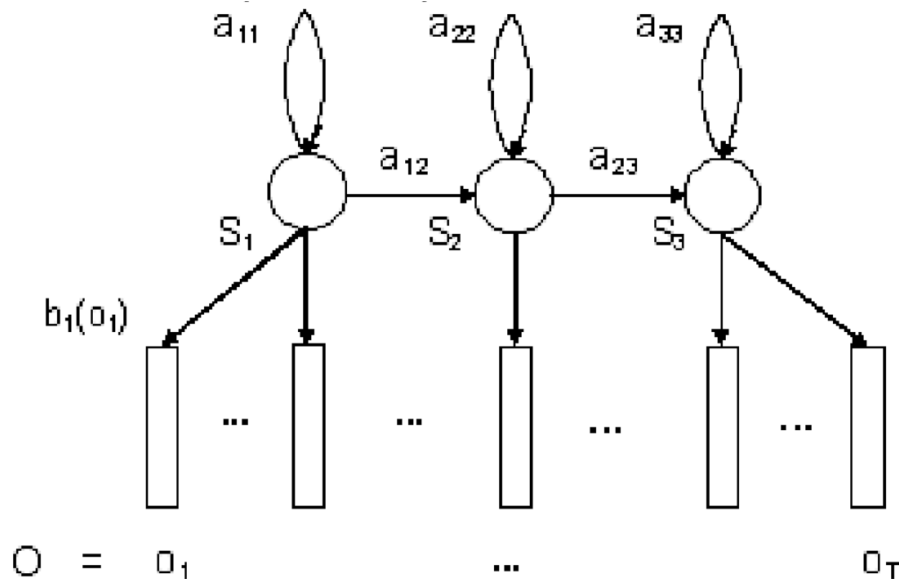


Рис. 1.1. ПММ із трьома станами

ПММ є кінцевим автоматом, що змінює свій стан в кожен дискретний момент часу  $t$ . Перехід із стану  $s_i$  у стан  $s_j$  здійснюється випадковим чином із ймовірністю  $a_{ij}$ . У кожен дискретний момент часу модель породжує вектор спостережень  $o_t$  (який у конкретній задачі є вектором особливостей, отриманим у перетворювачі сигналу) з ймовірністю  $b_j(o_t)$ .

Розподіл густини ймовірності спостережень моделюється кінцевою гаусівською сумішшю з чотирма компонентами. Кожна така модель означає один із звуків мови або відсутність звуку.

Алгоритми використовують ці моделі визначення команд в потоці мови. Найчастіше це завдання вирішується за допомогою ковзного вікна (sliding window) та застосування моделей-заповнювачів (filler models).

Головний недолік такого підходу те що він перебирає всі можливі варіанти входження ключового слова, що створює велику обчислювальну складність. Крім цього, метод розпізнавання команди на основі цього алгоритму полягає у застосуванні його до всієї мовної ділянки для кожної команди. Такий підхід має два істотні недоліки:

- 1) велика обчислювальна складність;
- 2) команди можуть містити слова, які погано розпізнаються.

Перша проблема виникає через необхідність застосування обробки до кожної можливої команди зі словника; друга – з наступних двох причин:

- складові команди містять складні для розпізнавання фонем мови;
- існують дефекти в деяких моделях фонем, отримані з незбалансованості мовної бази даних (МБД), на якій проводилося навчання, або через неправильний процес навчання.

Якщо друге обмеження можна усунути за рахунок правильного вибору ключового слова та якісної МБД, то обчислювальну складність змінити не вдасться. Тим самим метод може застосовуватися тільки в системах голосового керування з невеликим словником команд, які не потребують роботи в режимі реального часу або в системах, що мають значні обчислювальні ресурси.

Метод моделей наповнювачів. Для алгоритмів розпізнавання ключового слова слово розпізнавання представляється вбудованим у сторонню мову. На цій основі методи моделей заповнювачів обробляють цю сторонню мову за допомогою явного моделювання сторонньої мови за рахунок другорядних моделей.

Для цього до словника системи розпізнавання додаються узагальнені слова. Роль цих слів полягає у тому, щоб будь-який сегмент сигналу

незнайомого слова чи немовної акустичної події був розпізнаний системою як одне слово або ланцюжок із узагальнених слів. Для кожного узагальненого слова створюється та навчається акустична модель на основі даних із відповідними розміченими сегментами сигналу.

На виході з декодера отримується ланцюжок слів словника (ключових слів) та узагальнених слів. Узагальнені слова потім відкидаються, і частина ланцюжка, що залишилася, вважається результатом розпізнавання.

Недоліком підходу з використанням слів-заповнювачів є висока ймовірність помилки, коли ключові слова розпізнаються як узагальнені. Крім цього, постає й питання оптимального вибору алфавіту узагальнених слів, тому, що простір акустичних подій, що моделюється альтернативними моделями, дуже великий і складний, тому навчання цільових та альтернативних моделей відіграє важливу роль у підвищенні ефективності методу. У результаті підготовка моделей наповнювачів стає нетривіальним процесом, націленим на певний набір команд. Це не дозволяє динамічно змінювати словник ключових слів зі збереженням колишніх показників розпізнавання.

## 1.2 Голосове керування персональним комп'ютером

В даний час проблема розробки та створення ефективних систем розпізнавання мовлення та голосового керування персональним комп'ютером має велику актуальність. Спільне використання таких систем є основою повнофункціонального голосового інтерфейсу, сфера застосування якого дуже широка на практиці. Дослідження у сфері мовного інтерфейсу ведуться багатьма вченими, а розробка здійснюється найбільшими комп'ютерними організаціями. Використання розпізнавання команд за допомогою голосового керування на персональному комп'ютері може мати широкий діапазон і може бути реалізовано в багатьох галузях суспільства та промисловості.

На сучасному етапі розвитку технологій важко уявити повний перехід до голосового керування, що стане можливим лише тоді, коли персональний комп'ютер навчиться повністю розуміти людську мову, а не просто обмежений набір команд. Але наявність режимів голосового керування у різних додатках та операційних системах справді актуальна, тому що не всі люди мають можливість вільно використовувати свої руки. У принципі, керування голосом за допомогою комп'ютера спрощує роботу з ним, тому що команди передаються швидко та без зусиль через мікрофон.

Провідні компанії з розробки програмного забезпечення, розробки операційних систем та впровадження нових високотехнологічних пристроїв розуміють усі переваги включення голосового керування до списку можливостей своїх продуктів, тому в цьому напрямі ведеться інтенсивна робота.

Щоб настроїти голосові команди в операційній системі, на персональному комп'ютері має бути встановлена відповідна програма. Взаємодія з комп'ютером може бути створена не лише за допомогою клавіатури та миші. Тепер стали доступними і голосові команди. Існує спеціальний інструментарій, який дозволяє це робити. Його функція включає не тільки написання тексту для диктування або декодування аудіозаписів. За допомогою голосових команд можна запускати програми та використовувати їх.

Основою для створення систем голосового керування є алгоритми та методи розпізнавання мови. Найбільш поширеними та ефективними є такі методи:

1. MFCC – це метод кепстральних коефіцієнтів, який полягає у обчисленні коефіцієнтів спектру Фур'є, накладання на отриманий спектр набору фільтрів шкали  $mel$ , виконання логарифмування зміненого спектра та реалізації дискретного косинусного перетворення.

2. LPCC – це метод кепстральних коефіцієнтів лінійного передбачення, який ґрунтується на обчисленні коефіцієнтів авторегресійної моделі для кожного кадру аудіо сигналу.

3. PLP – це метод коефіцієнтів перцептивного лінійного передбачення, який відрізняється від методу LPCC тим, що враховуються особливості сприйняття різних частот людиною – перед обчисленням параметрів авторегресійної моделі сигнал проходить певну передобробку.

Результати аналізу наведених методів показали, що більшість зусиль розробників зосереджено на частотних характеристиках виділення людського вокального каналу і, отже, якісних методах, заснованих на обробці кепстрального сигналу.

Налаштування системи розпізнавання мовлення передбачає здатність керувати шумопоглинанням, визначати час затримки та результати розпізнавання, а також відсоток злиття кількох результатів розпізнавання. Підсумковий результат виникає у вікні та дозволяє редагувати мовну базу без спеціального знання системи.

### 1.3 Аналіз процесу формування мовних сигналів

Мовний сигнал переносить декілька видів інформації, а саме лінгвістичну, що відображає повідомлення диктора, представлення індивідуальної інформації про те, хто говорить, і емоційної інформації, що відображає емоції диктора. Перший інформаційний тип є найбільш важливим.

Створення мови починається з початкової концептуалізації, яку диктор хоче передати слухачу. Диктор перетворює цю ідею в лінгвістичну структуру шляхом вибору відповідних слів чи фраз які ясно (чітко) представляють її і тоді впорядковує їх відповідно до певних граматичних правил, залежно від взаємовідносин диктора і слухача.

Після цих процесів людський мозок виробляє руховим нервом команди, що приводять в рух різні м'язи голосових органів. Цей процес складається з фізіологічного процесу, що включає в себе нерви і м'язи, і фізичного процесу, в якому продукується і поширюється мовна хвиля.

Речення утворюється з допомогою основних слівних одиниць, кожне слово складається з складів, і кожен склад складається з фонем, які можуть бути класифіковані на голосні та приголосні звуки. Число голосних і приголосних різне, залежно від методу класифікації і залученої мови.

Як протилежність до фонем, що є найменшою одиницею мовлення з лінгвістичної чи фонемної точки зору, фізичною одиницею дійсної мови є фон.

Хоча число слів в кожній мові є дуже великим і постійно додаються нові слова, їх повне число є набагато меншим за можливе число комбінацій складів чи фонем.

Вважається, що число часто вживаних слів становить 2000-3000 і число слів вживаних окремою особою становить в середньому 5000-10000. Стрес і інтонація також грають критичну роль в визначенні важливих слів, в створенні питальних речень і в передачі емоцій від диктора.

Мова вимовляється намірено і в припущенні, що вона одержана і зрозуміла відповідному слухачу. Це очевидно означає, що продукування мови залежить від здатності слухати.

Мовна хвиля, створена голосовим органом передається через повітря до вух слухачів. В вусі вона активує органи чуття, що виробляють нервові імпульси, які подаються до мозку слухачів через систему слухових нервів.

Мовленнєвий процес включає три під процеси: генерування джерела, артикуляцію та випромінювання. Комплекс людського мовного апарату складається з легенів, трахей, гортані, носової та ротової порожнин. Разом це все утворює з'єднану трубу. Верхня частина починається з гортані і називається голосовим трактом, який приймає різні форми через рухи щелепою, язиком,

губами і іншими внутрішніми частинами. Носова порожнина відділяється від ротової шляхом піднімання піднебіння.

Коли черевні м'язи підносять діафрагму вгору, виштовхується повітря і виходить через легені з повітряним переходом через трахею і глотку в гортань. Глотка чи прорив між лівою і правою голосовими зв'язками, які зазвичай відкриті під час дихання, стає вузкою коли диктор хоче створити звук. Потім повітря через глотку періодично переривається відкриванням і закриванням щілини відповідно до потоку повітря і голосових зв'язок. Цей перемінний потік, що називається гортанним потоком чи мовним потоком, може бути зсимульований асиметричними трикутними хвилями.

Механізм звукових коливань звичайно є дуже складним. Коли голосові зв'язки сильно напружені і тиск повітря, що виходить з легенів (субглотковий повітряний тиск) є високим, періоди відкривання і закривання (періоди коливання голосових зв'язок) стають короткими, висота джерела звуку стає вищою. І навпаки, низький повітряний тиск створює менш високий звук. Цей період коливання голосових зв'язок називається основним періодом, а його обернена величина – основною частотою. Акцент і інтонація утворюються через часові коливання основного періоду. Джерело звуку, що складається з основної та гармонічної компонент, змінюється через голосовий тракт для створення тональних видів, таких як /a/ чи /o/, при створенні голосних. Під час створення звуків (голосних) голосовий тракт зберігається в відносно стабільній конфігурації (стабільній формі) в всіх відношеннях щодо вимови.

Два інші механізми відповідають за зміну повітряного потоку від губ в мовні звуки. Ці механізми лежать в основі створення двох видів приголосних: фрикативних і вибухових. Фрикативи, такі як /s/, /f/ і /S/ є шумні звуки, створені турбулентним потоком, який утворюється коли повітряний потік проходить через звуження голосового тракту, утвореного за допомогою язика чи губ.



Тональна різниця фрикативних звуків відповідає достатньо точно розміщеній конструкції і формі голосового тракту.

Вибухові, такі як /p/, /t/, /k/ є імпульсні звуки, які проходять з раптовим вивільненням високо стиснутого повітря, створеного стримуванням повітряного потоку в голосовому тракті, знову ж використовуючи язик чи губи. Тональна різниця відповідає різниці між стримуючою позицією та формою голосового тракту.

Утворення цих приголосних є цілком залежним від вібрації голосових зв'язок. Приголосні, які супроводжуються вібрацією голосових зв'язок називаються вокалізованими, ті які не супроводжуються такими вібраціями називаються невокалізованими. Придихання чи шепіт утворюється коли турбулентний потік створений в глотці незначним привідкриттям голосових зв'язок так, що вібрація голосових зв'язок не відбувається.

Напівголосні, назальні і африкативні звуки також входять в родину приголосних. Напівприголосні утворюються подібним чином як і голосні, але їхні фізичні властивості поступово змінюються без постійного вираженого періоду.

Щодо створення носових (назальних) звуків, носова впадина стає розтягнутою перегородкою ротової порожнини, з повітряним потоком стає запасним до ротової порожнини, через опускання піднебіння і затримку повітряного потоку в деякому особливому місці в ротовій порожнині. Коли носова впадина формує частину голосового тракту разом з ротовою порожниною протягом створення голосних голосний тип одержує назалізацію і утворюється носовий звук.

Африкати створюються послідовністю вибухових і фрикативних звуків, поки зберігається закрита конструкція в деякій позиції.

Налагодження форми голосового тракту для створення певних лінгвістичних звуків називається артикуляцією, тоді як рухи кожної з частин

голосового тракту відомі як артикуляційні рухи. Частини голосового тракту, що використовуються для артикуляції називаються артикуляційними органами, а ті, що можуть активно рухатися, такі як язик, губи, піднебіння називаються артикуляторами.

Різниця між артикуляційними способами створення фрикативних, вибухових, назальних і інших звуків ґрунтується на манері артикуляції. Звужене місце в голосовому тракті, утворене артикуляційними рухами визначається як місце артикуляції. Змінюючи тон видів звуків створюються зміною форми голосового тракту із зміною передавальних характеристик (таких як резонансні характеристики) голосового тракту.

Мовні звуки можуть бути класифіковані відносно комбінацій (поєднань) джерела і резонансних характеристик артикуляційного органу голосового тракту, що ґрунтуються на механізмі утворення, описаному вище.

#### 1.4 Висновки до розділу 1

Побудова системи голосового керування є актуальним завданням. Ці системи допомагають при роботі з комп'ютеризованими системами, як то голосовий набір тексту, керування електронними приладами в будинку тощо. Актуальною ця технологія є, наприклад, коли йдеться про людей із порушеннями опорно-рухового апарату. Як приклад розглядається задача керування електричним інвалідним кліслом.

В таких системах проводиться опрацювання голосових сигналів та виділення в їхній структурі ознак, за якими в наступному проводиться ідентифікація чи розрізнення окремих команд, що є певними ключовими словами. Для цього можуть застосовуватись знаки фонемного складу мови, амплітудні та часові параметри голосового сигналу, його склад, переходи між фонемами, складами чи словами.

Проаналізовано типову архітектуру систем голосового керування.

Встановлено, що як метод розпізнавання більшість сучасних систем використовують метод прихованих марківських моделей.

Для розроблення методу обробки мовних сигналів для формування сигналів керування проведено аналіз процесу формування мовних сигналів та виділено їхні параметри і характеристики, що можуть бути використані для розпізнавання окремих команд.

## РОЗДІЛ 2

### ОСНОВНА ЧАСТИНА

#### 2.1 Особливості аналогічних систем

Системи розпізнавання мови розрізняють за такими ознаками.

Інтервал між окремими словами. Якщо система розпізнає неперервне мовлення, користувач може вимовляти мовні фрази природно, не роблячи паузи між словами. Безперервне розпізнавання краще, проте воно вимагає більшої обчислювальної потужності комп'ютерів, що призводить поки до малого числа таких систем. У системах, що працюють з дискретною мовою, користувач при диктуванні повинен робити паузу між окремими словами, що зазвичай становить не менше частини секунди. Третім різновидом є системи, що виділяють одне слово з інтервалу мови, навіть якщо вона складається з кількох безперервно сказаних слів.

Залежність від диктора. Системи, що мають відносну незалежність від диктора, дозволяють користувачеві працювати з системою без попереднього налаштування, однак покращують надійність розпізнавання після навчання. Незалежність від диктора таких систем зазвичай досягається за рахунок зберігання звукових еталонів для всіх типових голосів носіїв даної мови. Це, безумовно, вимагає у кілька разів більшої продуктивності та обсягу пам'яті. Налаштування на голос диктора дикторозалежних систем займає зазвичай від 30 хвилин до кількох годин. Це є головною незручністю для користувачів.

Третім різновидом систем за цією ознакою є системи, що автоматично налаштовуються на голос диктора в міру їх використання.

Системи останнього типу мають дві особливості – їм потрібно знати, чи зробив користувач помилку, вимовивши конкретне слово (інакше навчання буде

неправильним); після налаштування одного диктора такі системи перестають надійно працювати з іншими голосами.

Ступінь деталізації при заданні еталонів. Розрізняють алгоритми, у яких як еталони використовуються цілі слова, та алгоритми, які використовують еталони елементів слів. Порівняння цілих слів дає більшу точність, швидкість, проте потребує більшого обсягу пам'яті (пропорційно до кількості слів у словнику) та навчання кожного слова.

Алгоритми порівняння елементів слів (фонем, складів тощо) доводиться застосовувати у разі великих словників, тому що обсяг пам'яті пропорційний кількості цих еталонних елементів слів (наприклад, звуків) і залежить від обсягу словника.

Розмір словника. Системи розпізнавання мовлення можуть використовувати великі чи малі словники. Розмір словника системи розпізнавання майже пов'язаний з реальною кількістю слів, які система може розпізнати. Він визначається кількістю слів, необхідних для розпізнавання у цьому конкретному стані системи. Системи, які працюють із маленькими словниками (близько 50 слів) дозволяють користувачеві давати прості команди комп'ютеру. Для диктування текстів потрібні великі словники (кілька десятків тисяч слів). Якщо системи диктування враховують контекст визначення активного підсловника у конкретному стані, то вони працюють зі словниками середнього розміру (близько 1000 слів).

Незважаючи на те, що в принципі можлива будь-яка комбінація цих характеристик, найбільш популярними є системи голосового управління комп'ютером і системи дискретного диктування тексту.

Системи голосового управління комп'ютером – це системи диктонеалежного розпізнавання команд, що безперервно вимовляються, складених зі слів обмеженого (до кількох сотень слів) словника.

Для подібних систем, якщо користувач вимовляє команду, що не входить до списку, система або видає відмову від розпізнавання, або видає як відповідь схожу "на слух" команду.

## 2.2 Приклади систем розпізнавання мовних сигналів

### 2.2.1. Біометричний спосіб розпізнавання голосу.

Розглянемо спосіб розпізнавання голосових команд, що ґрунтується на біометричній системі.

Процес розпізнавання у біометричній системі проходить дві стадії:

- Навчання – фізичний або поведінковий зразок запам'ятовується системою до бази даних, визначається унікальна інформація та складається “відбиток” голосу;
- Порівняння – представлений зразок порівнюється з вихідними та на основі отриманого результату формується відповідь.

Навчання. Весь процес занесення даних не займає багато часу. Для кожного питання користувач вимовляє кілька разів свою відповідь.

Вибір фрагментів фонограми. Існує кілька підходів щодо вибору ділянок сигналу для виділення параметрів. Наприклад, методи, що часто застосовуються - проводять використання всього мовного сигналу за винятком пауз або шумоочищення. Також існує спосіб вибору опорних сегментів - найбільш інформативних ділянок мовного сигналу. Вибираються найбільш енергетично сильні звуки, оскільки вони менш залежні від шумів та адитивних спотворень. Здебільшого це голосні та дзвінкі приголосні, тобто вокалізовані ділянки мови, що добре відображають роботу голосових зв'язок та мовного тракту. Слід зазначити, що багато авторів підкреслюють підвищену інформативність параметрів окремих груп звуків, наприклад, носових чи шиплячих в окремих областях амплітудного спектра.

Первинна обробка. Оцінка спектральних параметрів вимови проводиться з використанням вікон частотного зважування. Існує безліч різних часових вікон. Найбільш відомі з них: вікно Хеммінга, косинусне вікно, вікно Гауса, прямокутне вікно, вікно Блекмана та інші. Найбільший інтерес у дослідженні мовлення становить вікно Гаусса, оскільки воно визначено на всій числовій осі, містить лише один максимум в амплітудному спектрі і не містить жодного артефакту, а також легко застосовується при різних аналітичних обчисленнях.

Побудова "відбитка" голосу. Будується "відбиток" голосу (це може бути, наприклад, мультимодальний розподіл параметрів, інтегральні оцінки параметрів чи набір характерних проявів голосу). Записані відповіді накладають одна на одну, прибирається сторонній шум і на основі отриманого результату будується еталонний зразок голосу.

Збереження. Отриманий зразок зберігається в базі даних і може зберігатися в повному обсязі або стискатися ефективними алгоритмами, які дозволяють зберігати індивідуальні параметри без спотворення (методи лінійного передбачення).

Порівняння. На цьому етапі проводиться порівняльний аналіз відбитків голосів. Для порівняння можуть використовуватися кілька способів для забезпечення більш високої точності розпізнавання.

Найпростіший із способів - це порівняння відстаней між векторами - вихідними та представленими. Мінімальна відстань між векторами показує схожість зразка, що розпізнається, і еталона.

Вектори будуються з урахуванням амплітудно-частотної характеристики отриманого сигналу. При побудові вектора може бути використана різна кількість ознак. При збільшенні кількості ознак можна отримати достовірніший результат, але часу та ресурсів пам'яті при цьому потрібно більше. При зменшенні достовірність погіршується пропорційно, але результат досягається швидше. Також, якщо при розпізнаванні обмежити максимально можливу

відстань між векторами, представлений зразок при великій кількості ознак може не потрапити в один з наявних класів.

Існують різні методи побудови векторів та порівняння відстаней між ними:

1. Метод побудови еталонів. Для кожного класу за навчальною вибіркою будується стандарт, що містить значення ознак

$$\bar{x}^0 = \{x_1^0, x_2^0, \dots, x_N^0\}$$

Де 
$$x_i^0 = \frac{1}{K} \sum_{k=1}^K x_{ik},$$

$K$  – кількість об'єктів даного образу у навчальній вибірці,

$i$  – номер ознаки.

Фактично, еталон – це усереднений по навчальній вибірці абстрактний об'єкт. Абстрактним він називається тому, що може не збігатися не тільки з жодним об'єктом навчальної вибірки, але і з жодним об'єктом генеральної сукупності.

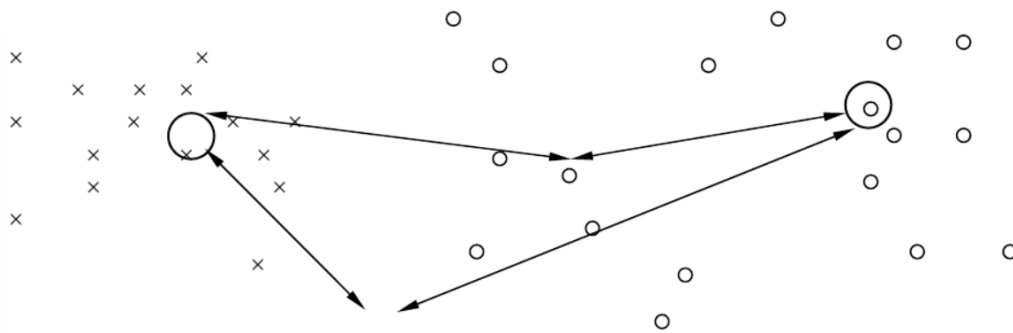


Рис. 2.1. Метод побудови еталонів



Розпізнавання здійснюється в такий спосіб. На вхід системи надходить об'єкт  $\overline{x^*}$ , приналежність якого до того чи іншого образу системи невідома. Від цього об'єкта вимірюються відстані до еталонів всіх образів, і система відносить  $\overline{x^*}$  до того образу, відстань до зразка якого мінімальна. Відстань вимірюється у тій метриці, яка введена для вирішення певного завдання розпізнавання.

2. Метод «подрібнених» еталонів. На першому етапі у навчальній вибірці охоплюються всі об'єкти кожного класу гіперсферою можливо меншого радіусу. Зробити це можна, наприклад, так. Будується зразок кожного класу. Обчислюється відстань від зразка до всіх об'єктів даного класу, які входять у навчальну вибірку. Вибирається максимальна з цих відстаней  $r_{\max}$ . Будується гіперсфера з центром в еталоні та радіусом  $r_{\max} + \varepsilon$ . Вона охоплює всі об'єкти цього класу. Така процедура проводиться для всіх класів (образів).

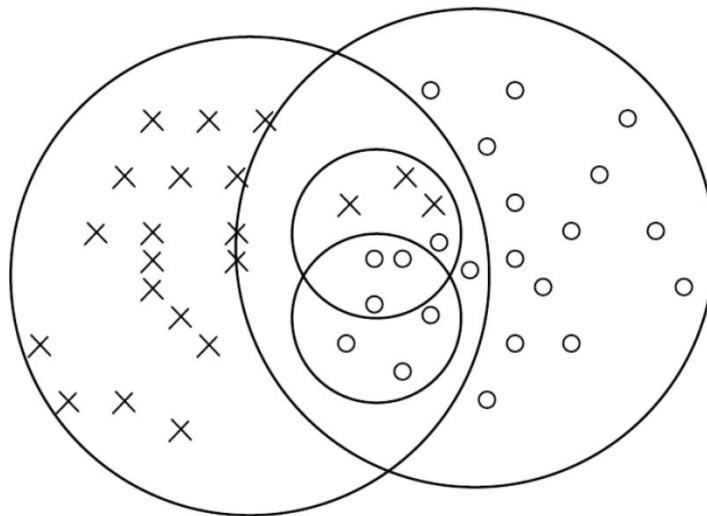


Рис. 2.2. Метод «подрібнених» еталонів

Якщо гіперсфери різних образів перетинаються й у області перекриття виявляються об'єкти більш як одного образу, то для них будуються гіперсфери

другого рівня, потім третього тощо, до тих пір, поки області не виявляться такими, що не перетинаються, або в області перетину будуть присутні об'єкти тільки одного образу.

Розпізнавання здійснюється в такий спосіб. Визначається місцезнаходження об'єкта відносно гіперсфер першого рівня. При попаданні об'єкта в гіперсферу, що відповідає одному і лише одному образу, процедура розпізнавання припиняється. Якщо ж об'єкт опинився в області перекриття гіперсфер, яка при навчанні містила об'єкти більш ніж одного образу, то переходимо до гіперсфер другого рівня і проводимо такі ж дії, як для гіперсфер першого рівня.

Цей процес триває до того часу, поки належність невідомого об'єкта конкретному образу не визначиться однозначно. Щоправда, ця подія може не наступити. Зокрема, невідомий об'єкт може не потрапити до жодної гіперсфери будь-якого рівня. У цих випадках "учитель" повинен включити до вирішальних правил відповідні дії. Наприклад, система може або відмовитися від рішення про однозначне віднесення об'єкта до будь-якого образу, або використовувати критерій мінімуму відстані до еталонів цього або попереднього рівня.

3. Метод «найближчого сусіда». Нерозпізнаний об'єкт  $\bar{x}^*$  відноситься до того образу, чий «представник» виявився найближчим до  $\bar{x}^*$ .

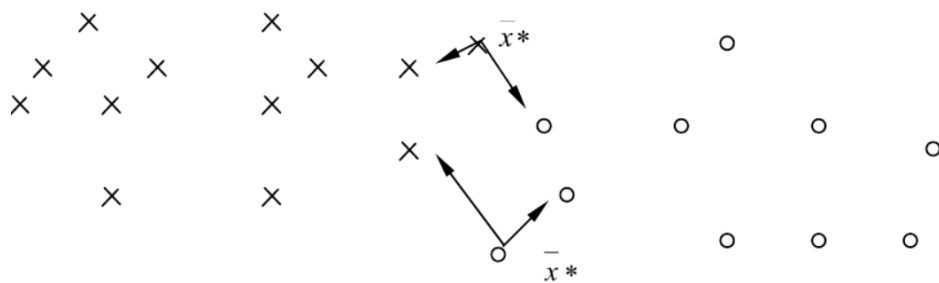


Рис. 2.3. Метод «найближчого сусіда»

Правило k найближчих сусідів полягає в тому, що будується гіперсфера об'єму  $V$  з центром  $\bar{x}^*$ . Розпізнавання здійснюється з більшості «представників» будь-якого образу, що опинилися всередині гіперсфери.

Ще один із способів розпізнати приналежність об'єкта до одного із класів – це вирішальні правила. Межа, що розділяє ознаковому просторі області різних образів, описується лінійною функцією.

$$D(\bar{x}) = \sum_{i=1}^N a_i x_i + a_0.$$

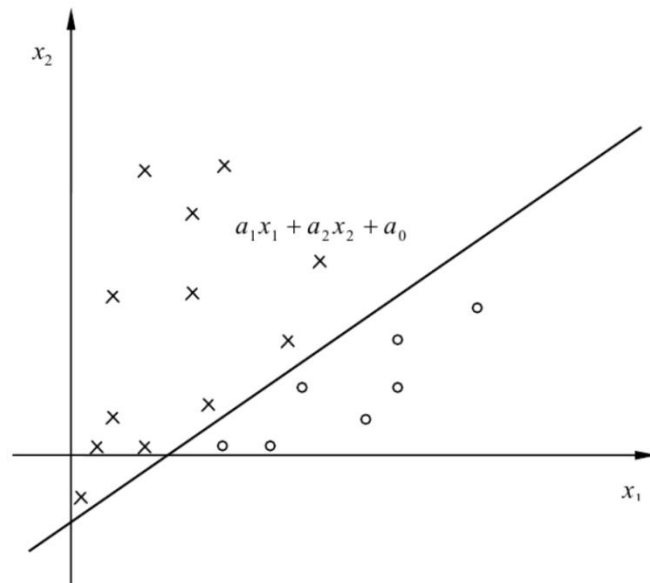


Рис. 2.4. Метод «вирішальних правил»

Одна межа при цьому поділяє області двох образів. Якщо кількість класів більше двох, то потрібно кілька лінійних функцій і межа є кусково-лінійною.

Процес розпізнавання. За вибраними значеннями частот будуються вектори. Розпізнавання відбувається двома методами:

- методом побудови еталонів

- методом «найближчого сусіда»

На початку першого методу із векторів-шаблонів кожного образу будуються вектори-еталони. Насправді, значеннями стандартів є усереднені значення за кожною ознакою вихідних векторів. Далі обчислюються відстані між вектором невідомого образу та кожним з еталонних за формулою:

$$\sqrt{\sum_{i=1}^n (\chi_2^i - \chi_1^i)^2}$$

де n = кількість ознак;

Ухвалення рішення обумовлюється мінімальною відстанню між векторами.

Другий спосіб обчислює значення відстаней аналогічним способом, але між кожним векторів-шаблонів і вектором невідомого образу. Ухвалення рішення здійснюється за тими ж параметрами.

### 2.2.2. Двохрівневий метод розпізнавання голосової команди

Розглянемо новий двохрівневий метод розпізнавання команди, який дозволяє вирішити ці проблеми за рахунок комбінації методу розпізнавання ключового слова з методом розпізнавання злитого мовлення.

Двохрівневий метод розпізнавання команди. Розроблено наступний підхід [2]: в доданні в команду заздалегідь встановленого ключового слова та застосування алгоритму розпізнавання ключового слова тільки для цього вступного (ключового) слова, а для частини команди, що залишилася, використання стандартного алгоритму розпізнавання злитого мовлення (CSR).

Таким чином, використання запропонованої системи полягає у вимові команди, яка складається з ключового слова (за допомогою нього система розуміє, що звертаються до неї) та власне команди (назва дії, яку хоче виконати

користувач). Наприклад: «Агент, увімкнути світло», якщо «Агент» – ключове слово, а «включити світло» – безпосередньо команда.

Виділення двох частин спрямоване на збільшення швидкості роботи і якості розпізнавання.

Збільшення швидкості досягається за рахунок скорочення кількості ключових слів, що розпізнаються, до одного, заздалегідь встановленого, і використання для розпізнавання безпосередньо команди алгоритму розпізнавання зливої мови (CSR), який на порядок швидше алгоритму розпізнавання ключового слова, часом роботи якого в даному випадку можна знехтувати.

Поліпшення якості розпізнавання досягається за рахунок введення заздалегідь встановленого, єдиного ключового слова, що дає можливість в якості такого слова вибрати слово, що легко розпізнається, або провести додаткові заходи для поліпшення розпізнавання цього слова: підготовка додаткового мовного матеріалу, калібрування параметрів алгоритму розпізнавання ключового слова.

Крім цього, можна використовувати назву системи або ім'я як ключове слово, що дає наступні переваги:

- є для людини інтуїтивніше звичним і зручним способом вимови команд;
- додає однакову всім користувачам командну інтонацію, що зменшує кількість інтонаційних варіацій вимови команди, тим самим спрощує процес розпізнавання;
- зменшується кількість помилкових спрацьовувань за рахунок скорочення кількості ключових слів до одного.

Недоліком запропонованого підходу розпізнавання команди є некоректні спрацьовування системи, якщо після ключового слова впливають команди чи фрази не зі словника команд. Наприклад, якщо користувач скомандує «Агент, увімкнути телевізор» і такої команди немає у словнику команд, то алгоритм

розпізнає цю команду як команду зі словника, найбільш співзвучну зі сказаною, наприклад «Агент, увімкнути світло». Це відбувається через застосування алгоритму розпізнавання злитого мовлення для розпізнавання безпосередньо команди, який лише розпізнає голосову фразу відповідно до заздалегідь визначеного словника і може визначити, чи є ця фраза не з словника. Таке завдання вирішує алгоритм розпізнавання ключових слів. Таким чином, пропонується підхід вносить наступне обмеження на використання системи: користувач повинен знати набір команд, що підтримуються системою, і забороняти їх вгадування.

Архітектурною відмінністю підходу є об'єднання двох різних алгоритмів розпізнавання мови в єдиний блок. У цьому блоці (модулі) спочатку робиться висновок про наявність вступного слова і потім, якщо це слово є присутнім, відбувається розпізнавання самої команди. Таким чином, на виході головного модуля ми отримуємо інформацію про команду або її відсутність.

Алгоритм розпізнавання ключового слова. Одним з популярних алгоритмів розпізнавання ключового слова є метод ковзного вікна. Аналіз методу, і навіть експериментальні дані показують, що метод містить істотний недолік – велику обчислювальну складність, що створює незручності та заважає застосуванню системи практично.

Для усунення цього недоліку був розроблений новий однопрохідний алгоритм розпізнавання ключового слова, що включає новий алгоритм поширення і нову функцію правдоподібності.

Для кожної фонемі визначено свій власний поріг спрацьовування. Використання вищого порога для добре розпізнаваних фонем і нижчого для поганорозпізнаних дозволило суттєво покращити якість розпізнавання. Крім цього, для скорочення явно хибних шляхів функція правдоподібності використовує спеціальний загальний поріг на кожен стан шляху.

Шлях видаляється зі списку поточних шляхів, якщо не виконується хоча б одна з наступних умов:

- 1) показник відповідності всіх станів шляху більший за загальний поріг;
- 2) середнє значення показника відповідності поточної фонемі більше за поріг для цієї фонемі.

Детектор голосу складається з двох частин, які в сумі дають уявлення про наявність мовної складової сигналу. Перша частина полягає в зміні енергії сигналу, друга – на періодичності, яка розраховується з допомогою алгоритму, заснованого методі найменших квадратів.

Для перетворення сигналу у вектори особливостей за основу взято алгоритм, запропонований європейським інститутом стандартів телекомунікації, блок-схема якого зображена на рис. 2.

Для виділення особливостей алгоритм використовує властивості людського сприйняття. Крім того, до нього були внесені такі модифікації, спрямовані на покращення якості роботи: ліфтрація (Liftering); відрахування середнього кепстрального значення (Spectral Mean Subtraction); нормалізація енергії (Energy Normalization). На виході даного блоку для кожної ділянки сигналу 25 мс формується вектор з 39 параметрів. Першими 13-ма є кепстральні коефіцієнти (12 mel-frequency cepstral coefficient) і логарифм енергії ( $\log E$ ), а рештою – похідні 1-го і 2-го порядку цих коефіцієнтів (вони показують динаміку зміни).

Тестування проводилося для розпізнавання зливої мови для словника 20 слів. Як матеріал для тестування роботи модуля шумоочищення використовувався шум зі спеціального набору шумів. На першому етапі було проведено порівняння роботи алгоритму розпізнавання злого мовлення для трьох різних конфігурацій модуля шумоочищення:

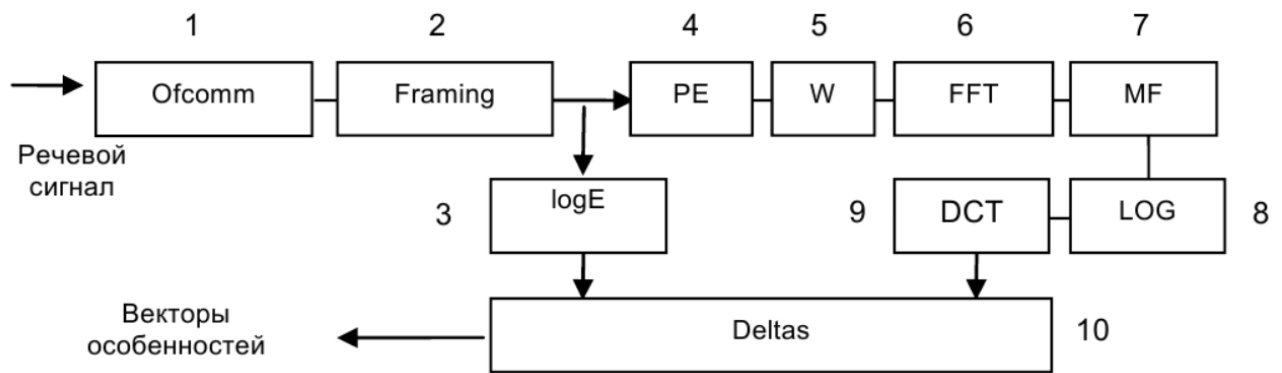


Рис. 2.5. Блок діаграма перетворювача сигналу: 1 – компенсація усунення; 2 – розбиття на блоки; 3 – обчислення енергії; 4 – попередня корекція; 5 – застосування віконної функції; 6 – розкладання на частотні складові; 7 – застосування фільтрів; 8 – нелінійне перетворення; 9 – косинус перетворення; 10 – обчислення 1-ї та 2-ї похідної за часом без шумоочищення;

- з використанням алгоритму шумоочищення OM-LSA;
- з використанням алгоритму шумоочищення OM-LSA та із застосуванням розроблених евристик.

Другий етап тестування полягав у експериментальному порівнянні розробленого алгоритму розпізнавання ключового слова з методом ковзного вікна. Як показники для порівняння було обрано швидкість роботи та якість розпізнавання. При цьому якість розпізнавання вимірювалася двома величинами:

- правильністю спрацювання та розпізнавання голосової команди;
- неправильне спрацювання на ділянці сигналу, що не містить голосової команди.

Результати експериментів показують перевагу розробленого методу визначення ключового слова по швидкості і по якості розпізнавання.

Відмінною рисою розробленого дворівневого методу розпізнавання голосової команди є поділ мовної команди на частини: ключового слова і безпосередньо команди. Недоліком цього підходу є таке обмеження:



користувач повинен знати набір команд, що підтримуються системою, і забороняти їх вгадування. Запропонований однопрохідний алгоритм розпізнавання ключового слова та проведене експериментальне порівняння розробленого алгоритму з методом ковзного вікна показують перевагу розробленого методу як у плані швидкості, так і в плані якості.

### 2.3 Висновки до розділу 2

Розглянуто особливості систем розпізнавання мови. При цьому розглянуто такі системи, які враховують інтервали між окремими словами, залежність від диктора та системи, що автоматично налаштовуються на голос диктора в міру їх використання. Системи останнього типу мають дві особливості – їм потрібно знати, чи зробив користувач помилку, вимовивши конкретне слово. Після налаштування одного диктора такі системи перестають надійно працювати з іншими голосами.

Розглянуто алгоритми виділення елементів слів (фонем, складів тощо)

Спрощено процес розпізнавання включає шумоочищення та відділення корисного сигналу.

Проаналізовано спосіб розпізнавання голосових команд, що ґрунтується на біометричній системі, та двохрівневий метод розпізнавання голосової команди. Використання останньої системи полягає у вимові команди, яка складається з ключового слова (за допомогою нього система розуміє, що звертаються до неї) та власне команди.

## РОЗДІЛ 3

### НАУКОВО-ДОСЛІДНА ЧАСТИНА

#### 3.1 Принцип роботи пропонованого способу

Суть методу обробки голосових сигналів для виділення слів-команд і відповідно формування на їх основі сигналів керування зводиться до наступного. Потрібно розпізнати чотири команди в голосовому сигналі: «вліво», «вправо», «взад», «вперед».

На першому етапі користувач промовляє ці слова в мікрофон, система опрацювання їх обробляє та виділяє інформативні ознаки, за якими в наступному буде проводити ідентифікацію цих команд. Власне для розробки методу використано описані в попередніх розділах методи із певною модифікацією. Спочатку проводиться запис сигналу, потім проводиться його фільтрація адаптивними методами. Далше проводиться виділення довжини слова кожної команди. На наступному етапі проводиться сегментація кожного слова на фонемі. Власне на останньому етапі проводиться виділення ділянок з ознаками голосних та приголосних вокалізованих звуків.

Усі чотири команди відрізняються фонемним складом. Так, перша команда містить 5 фонем, з яких 2 голосні та 3 приголосні вокалізовані. Друга команда містить 6 фонем, з яких 2 голосні, 3 приголосні вокалізовані та одна глуха. Третя команда містить 4 фонемі з однією голосною та трьома приголосними вокалізованими. І четверта команда містить 6 фонем з двома голосними однією глухою та трьома приголосними вокалізованими фонемами. Чергування фонем в командах є різним. Тому, для фонемної сегментації застосовано як інформативну ознаку значення частоти основного тону.

На першому етапі проводиться виявлення часових меж кожної команди. Для цього застосовано ковзне вікно, в межах якого проводиться обчислення середнього значення голосового сигналу. Далше це значення порівнюється з порогом, і якщо воно рівне або перевищує поріг, то значення порогової функції  $h$  приймається рівним 1, в іншому випадку – 0. Алгоритм такої обробки наведено на рис. 3.1.

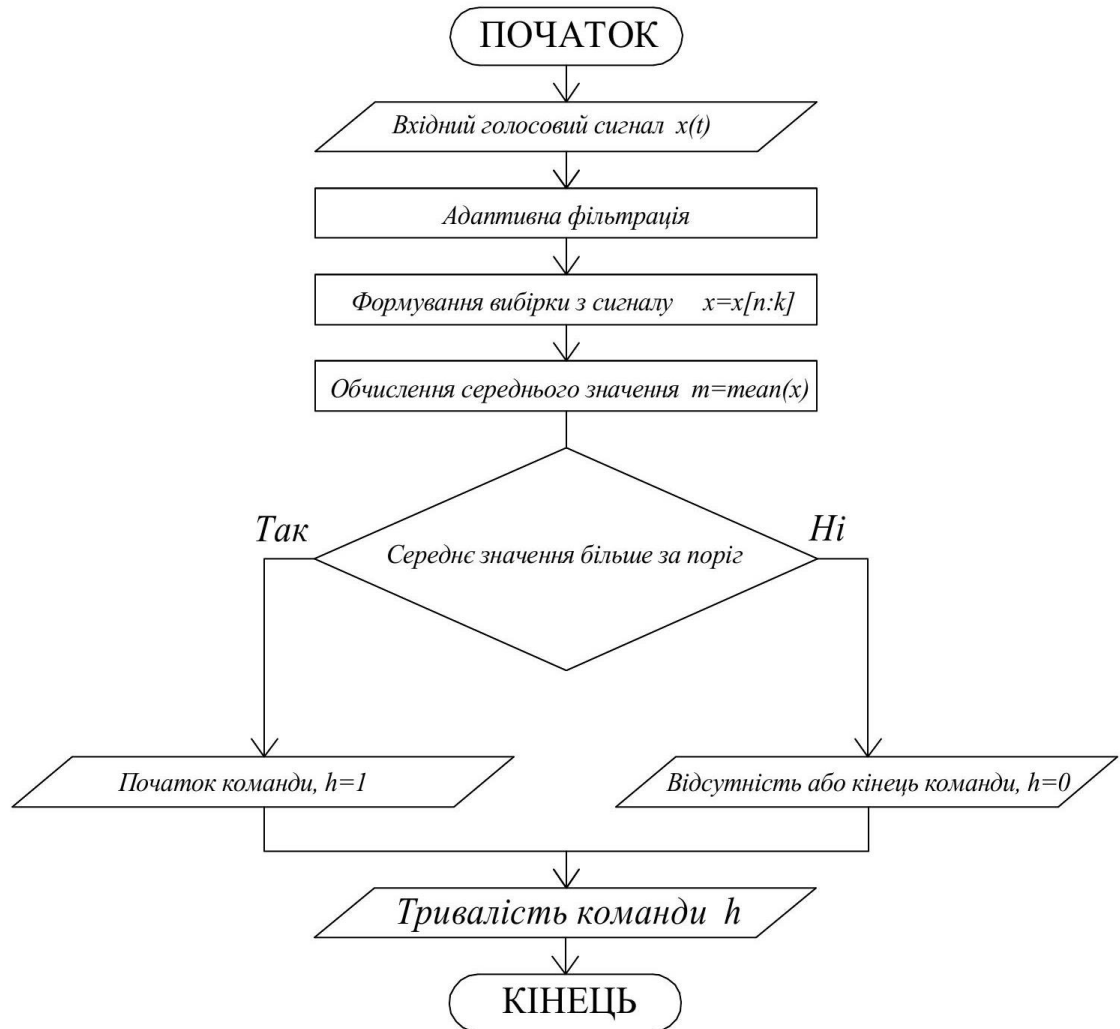


Рис. 3.1. Алгоритм знаходження тривалості команди

На наступному етапі проводиться визначення частоти основного тону, яка є індивідуальною характеристикою та є присутньою в голосних та приголосних вокалізованих фонемах. Для цього використано метод формантного аналізу,

коли частоті основного тону відповідає частота першого максимуму в спектрі потужності. Для цього також обробка проводилась з застосуванням ковзного вікна. Алгоритм такої обробки наведено на рис. 3.2.

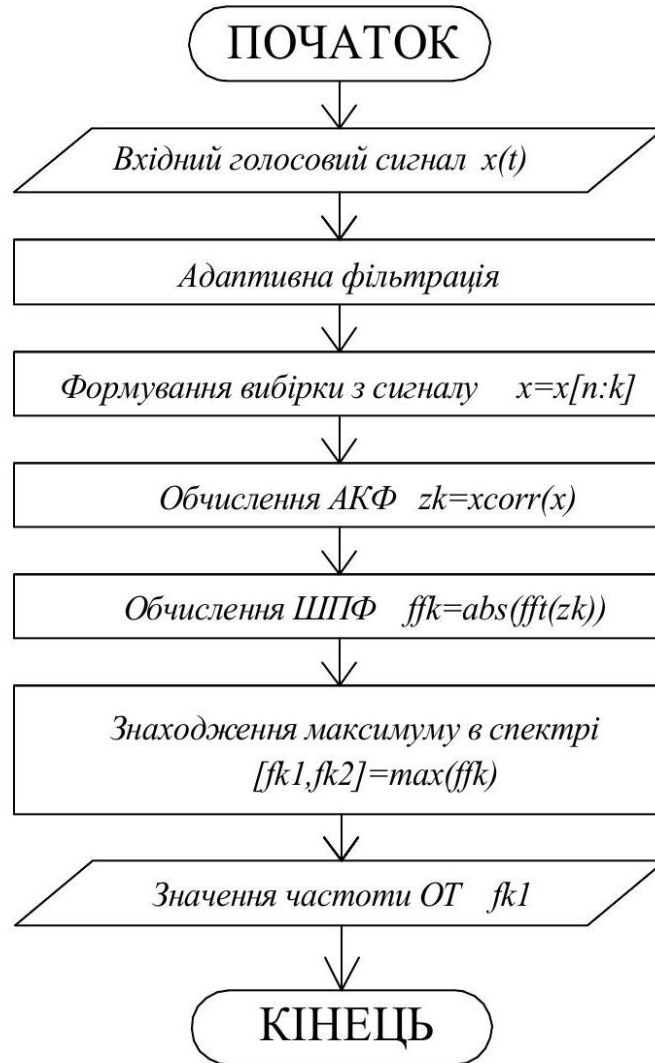


Рис. 3.2. Алгоритм оцінювання частоти основного тону

І на останньому етапі проводиться ідентифікація команди. Для цього невідома команда обробляється в межах ковзного вікна наступним чином. Проводиться оцінювання наявності частоти основного тону. Вона буде присутня лише на ділянках в межах тривалості команди, які відповідають голосним та приголосним вокалізованим звукам. Далше формуються інтервали з наявністю такої частоти і вже за їхньою послідовністю та розміром

проводиться ідентифікація команди, оскільки для чотирьох розглянутих команд послідовності цих ділянок і тривалості мають бути різними. Ідентифікувавши ж команди згенерувати сигнали керування не є складною задачею. Алгоритм розпізнавання команд наведено на рис. 3.3.

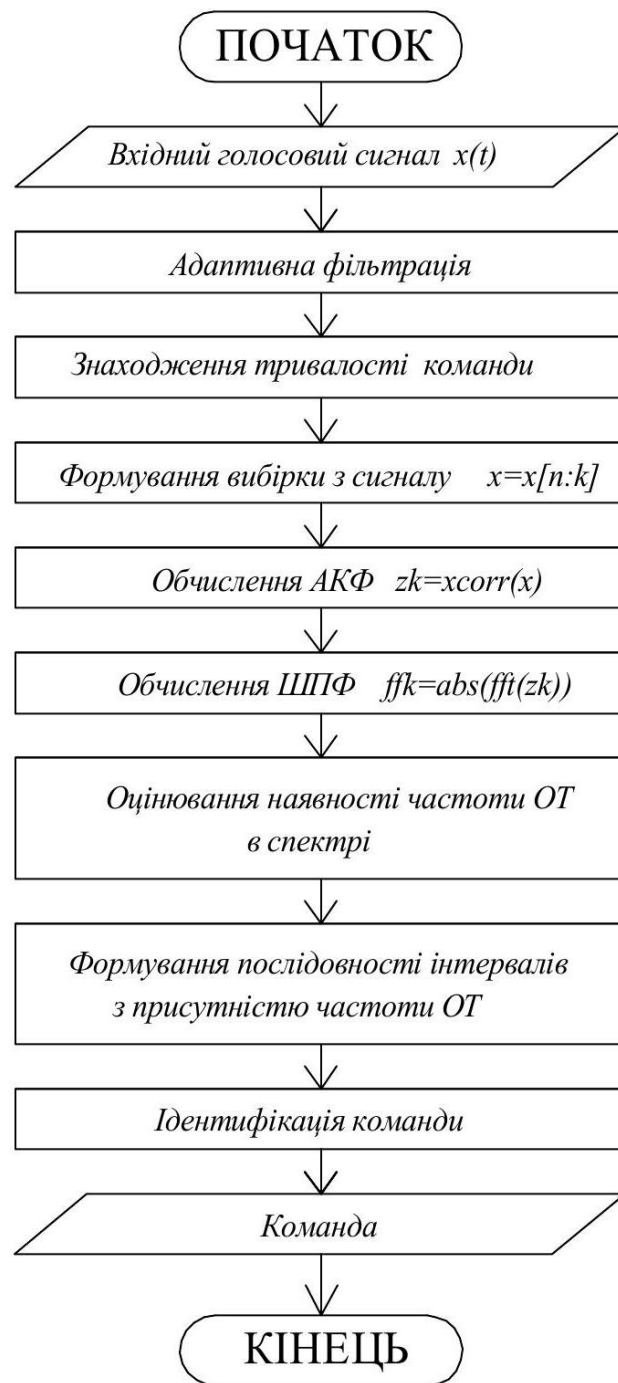


Рис. 3.3. Алгоритм розпізнавання команд

### 3.2 Результати експериментального дослідження

Для відбору голосових сигналів з командами використано стандартну комп'ютерну гарнітуру та програму Adobe Audition. Вигляд записанного сигналу із чотирьма командами наведено на рис. 3.4.

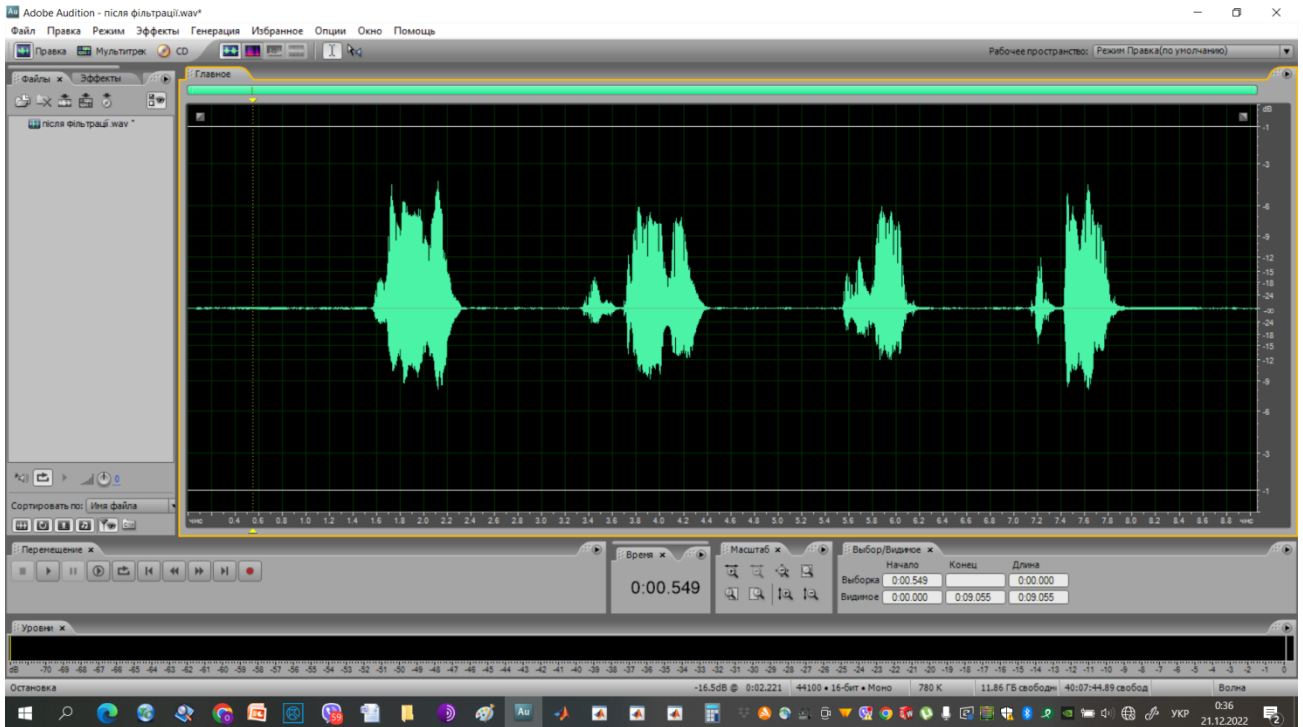


Рис. 3.4. Вигляд записанного сигналу із чотирьма командами

Можна помітити, що в сигналі є присутні шуми, які можуть мати значно вищий рівень. На рис. 3.5 наведено вигляд динамічного спектра сигналу.

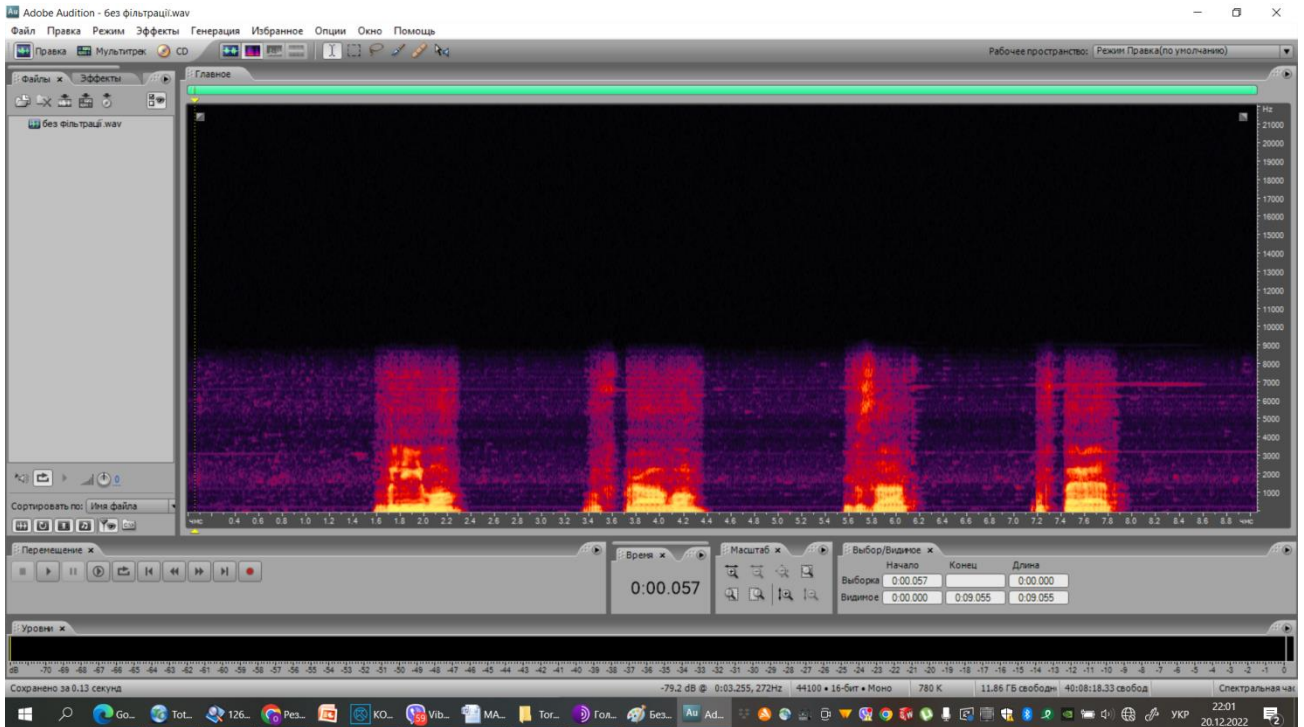


Рис. 3.5. Вигляд динамічного спектра сигналу

Тому на кожному етапі обробки проводиться адаптивна фільтрація. Застосовано метод спектрального віднімання, який є реалізований в програмі Adobe Audition. Сигнал після такої фільтрації та його динамічний спектр наведено на рис. 3.6 та рис. 3.7.

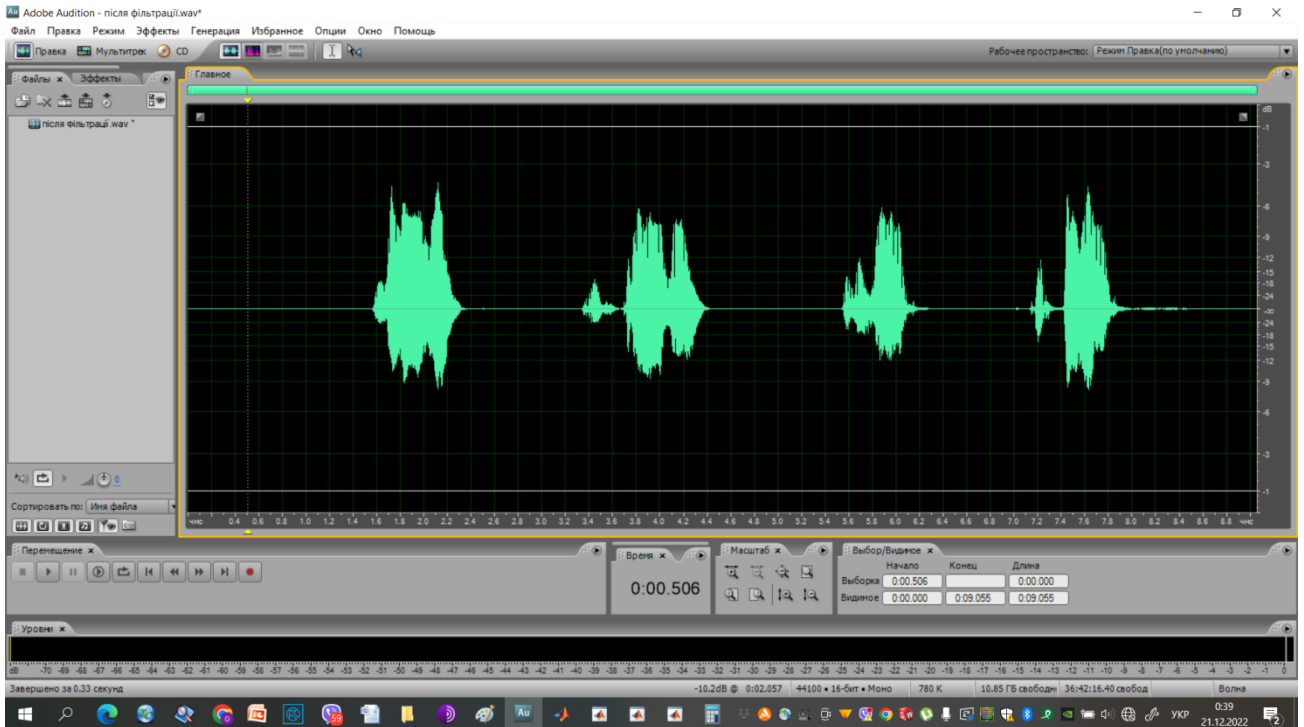


Рис. 3.6. Сигнал після адаптивної фільтрації

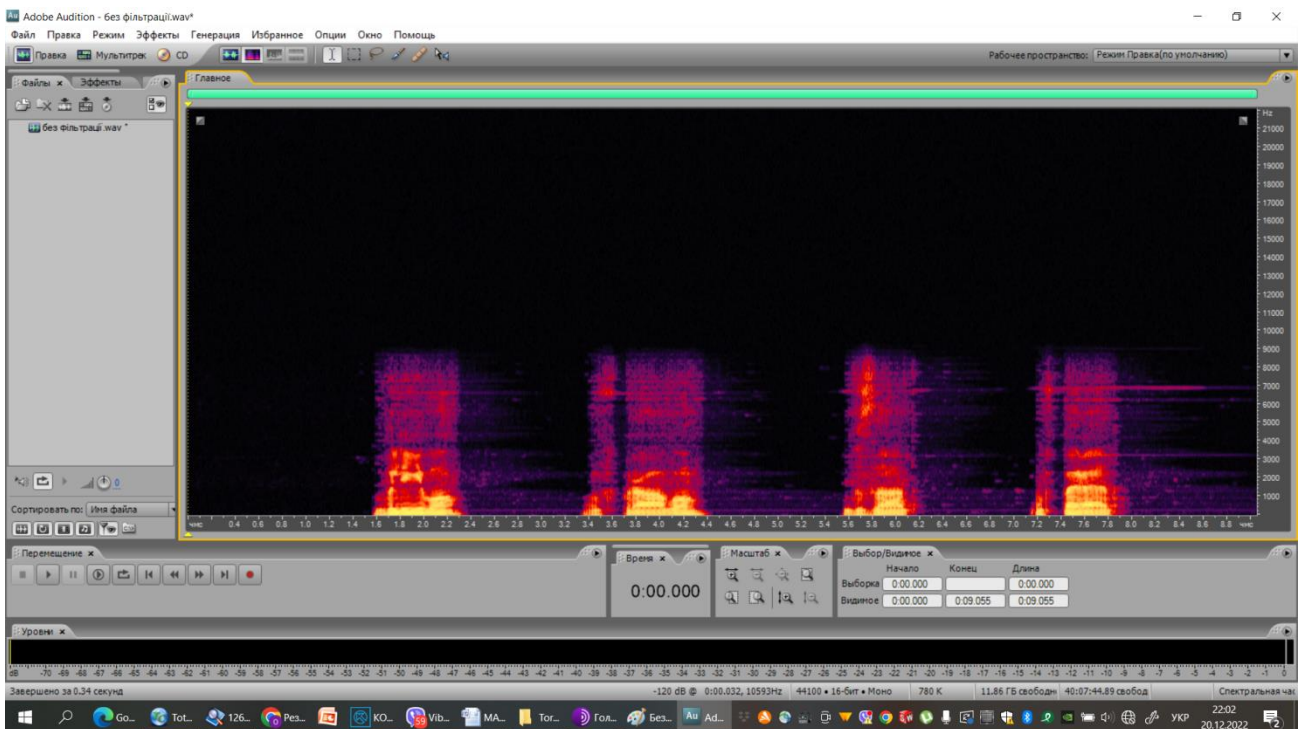


Рис. 3.7. Динамічний спектр сигналу після фільтрації



Дальше очищений сигнал було завантажено в середвище Matlab, в якому проводилась наступна його обробка. Вигляд такого сигналу наведено на рис. 3.8.

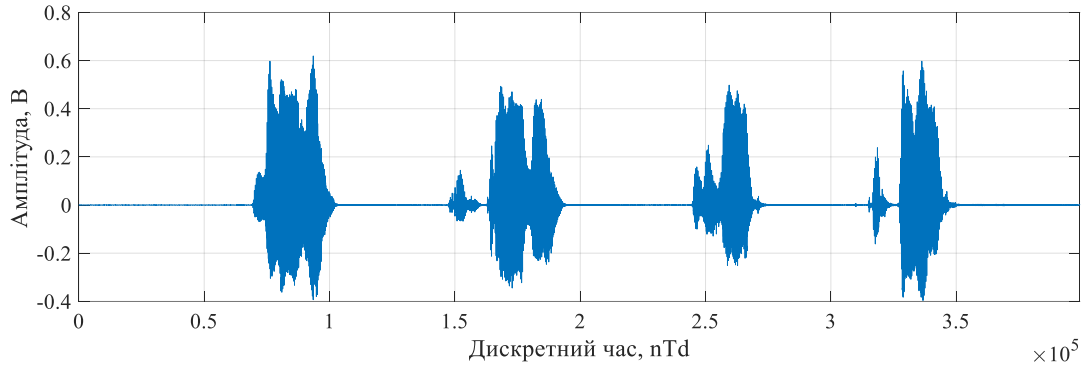


Рис. 3.8. Завантажений в Matlab голосовий сигнал із командами

Дальше відповідно до алгоритму на рис. 3.1 проведено визначення тривалості кожної команди. Для цього застосовано ковзне вікно з тривалістю 5мс, яка вибиралась апріорно і була в більше як в 100 разів коротшою за реалізацію команди. Поріг також вибирався апріорно. Вигляд значення порогової функції наведено на рис. 3.9.

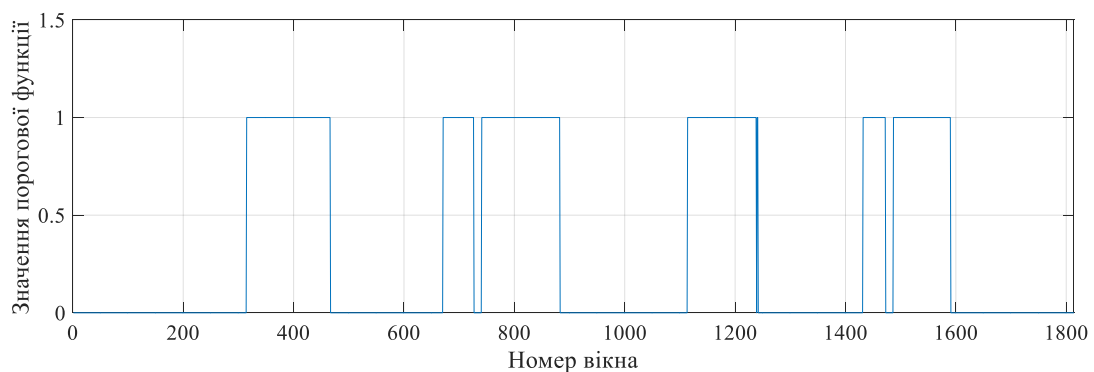


Рис. 3.9. Вигляд порогової функції

Як можна побачити, в такий спосіб за пороговою функцією справді можна визначити тривалості голосових команд.

На наступному етапі було визначено частоту основного тону, яка становила 172 Гц для цього випадку (відповідно до алгоритму на рис. 3.2).

На останньому етапі проведено фонемну сегментацію кожної окремої команди. На рис. 3.10 наведено вигляд голосового сигналу першої команди «Вліво».

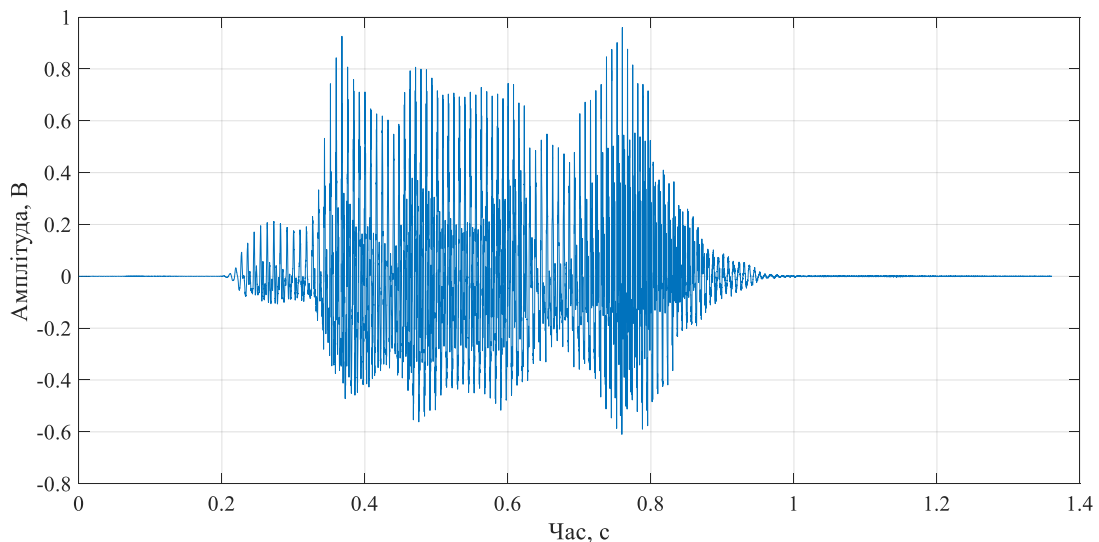


Рис. 3.10. Вигляд голосового сигналу команди «Вліво»

Дальше за описаним вище алгоритмом побудовано оцінки спектру та проведено його усереднення в околі значень частоти основного тону. Опрацювання проводилось в межах ковзного вікна. Якщо в цих межах була присутня частота основного тону, то це проявлялось в спектрі сигналу. Ці усереднені значення спектра в околі частоти основного тону відкладались на одній часовій осі відповідно до того де розміщувалось ковзне вікно. Отриманий графік наведено на рис. 3.11.

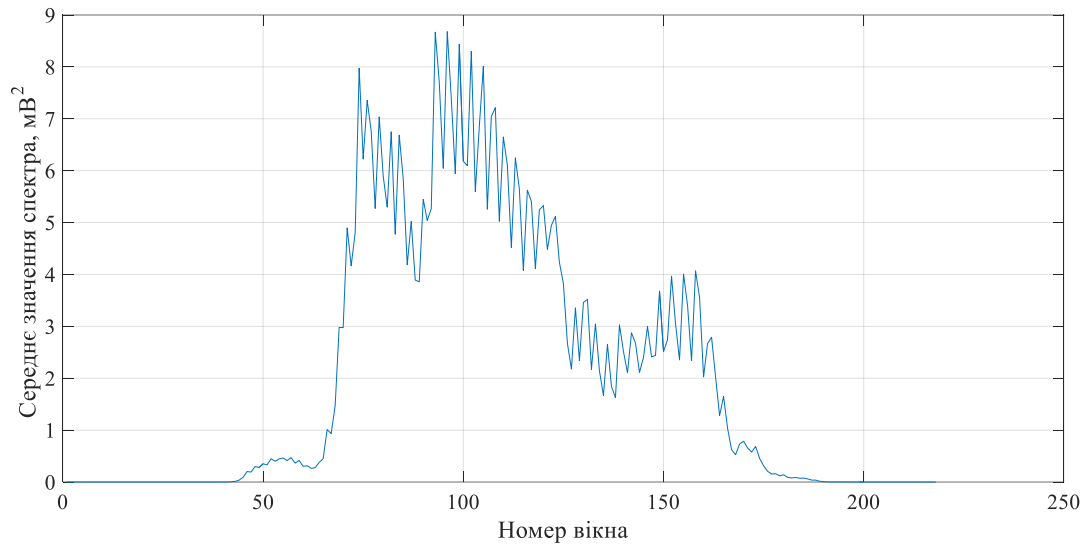


Рис. 3.11. Присутність частоти основного тону в голосовому сигналі

Дальше було побудовано порогову функцію із порогом на рівні 0,1, який вибирався апріорно. Вона наведена на рис. 3.12.

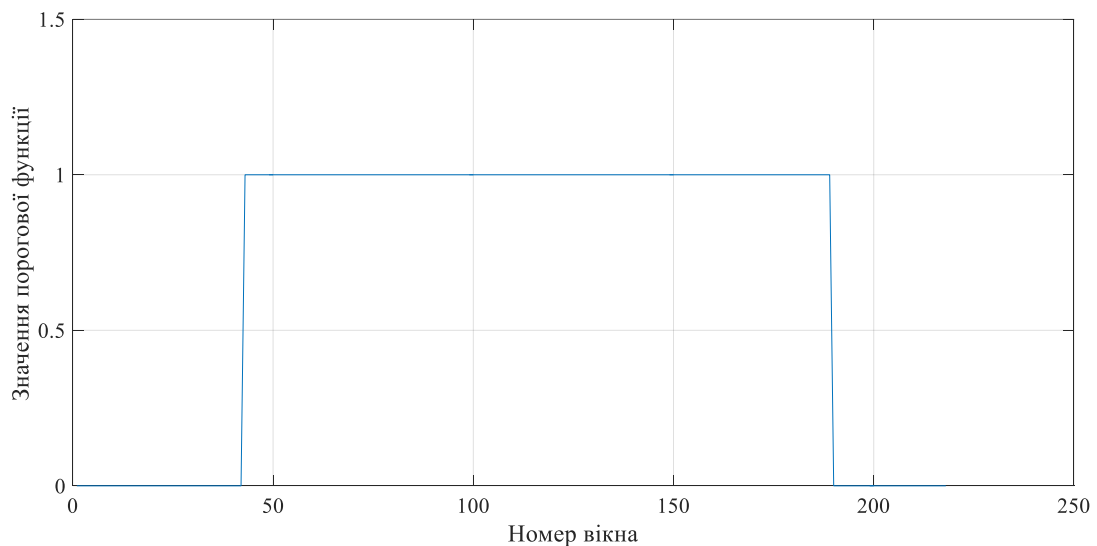
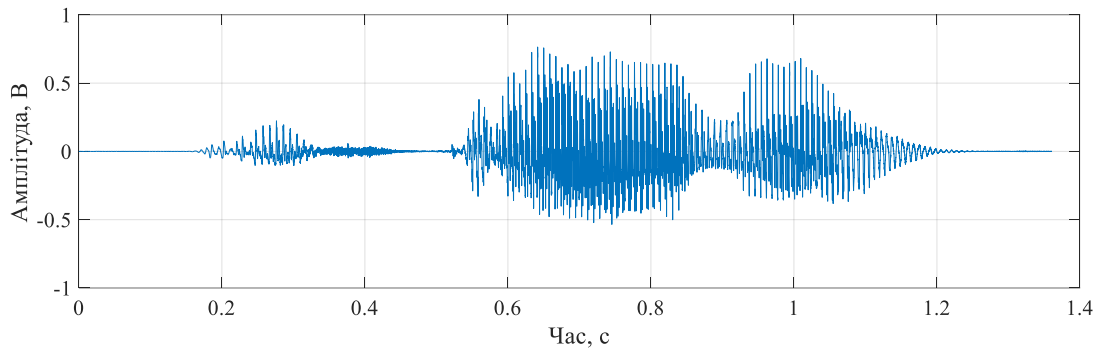


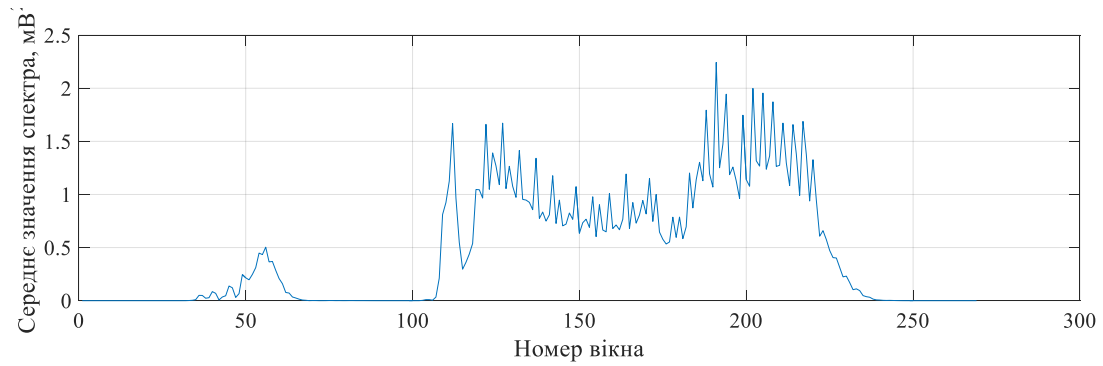
Рис. 3.12. Вигляд порогової функції

Звідси можна зробити висновок, що в сигналі частота основного тону є присутня в межах повної тривалості команди, що відповідає дійсності, оскільки всі фонемі є або голосними або приголосними локалізованими.

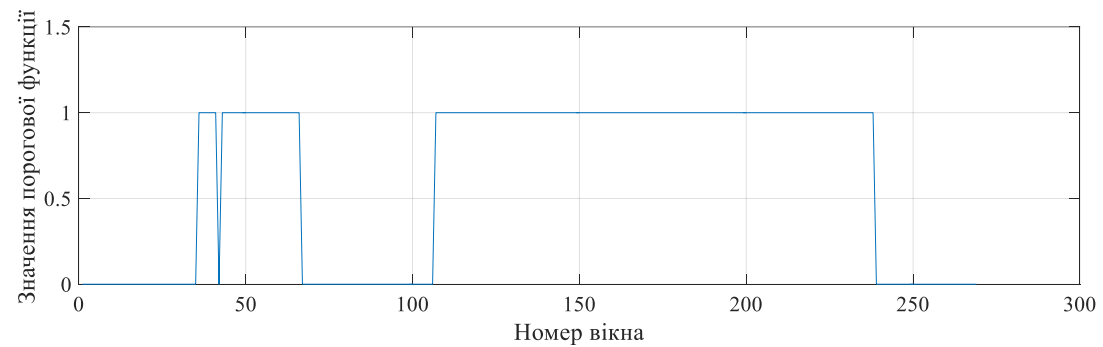
На рис. 3.13 наведено вигляд сигналу другої команди «вправо», графік присутності частоти основного тону в сигналі та вигляд порогової функції.



а)



б)

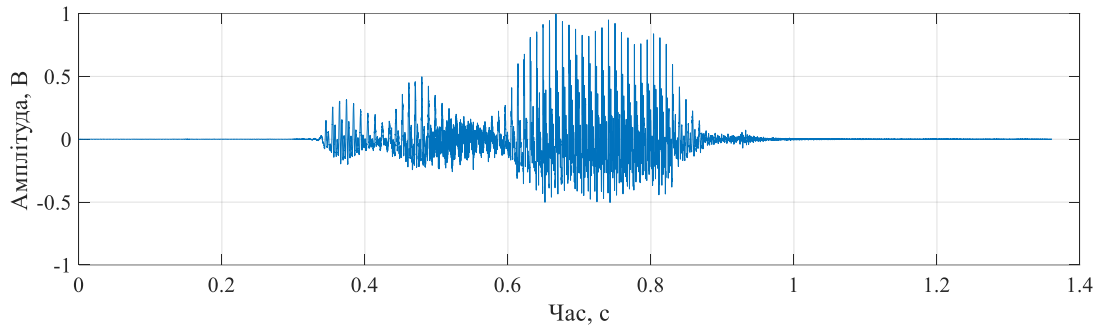


в)

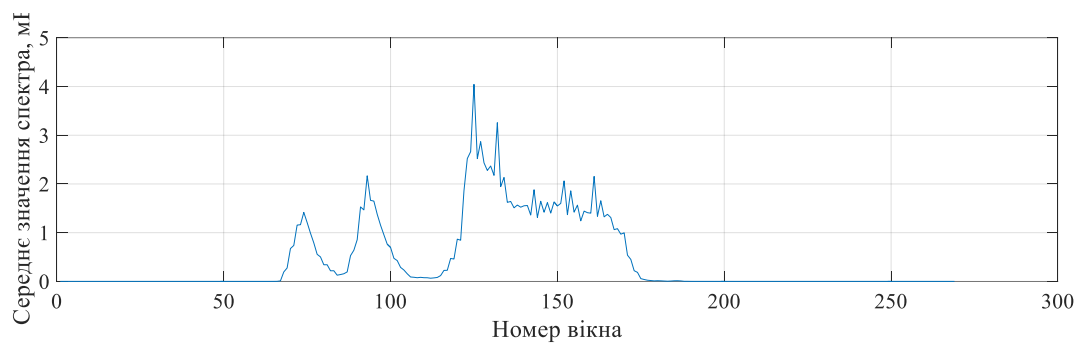
Рис. 3.13. Вигляд сигналу команди «вправо» (а), графік присутності частоти основного тону в сигналі (б) та вигляд порогової функції (в)

В цьому випадку є пропуск, який відповідає глухій фонемі [п].

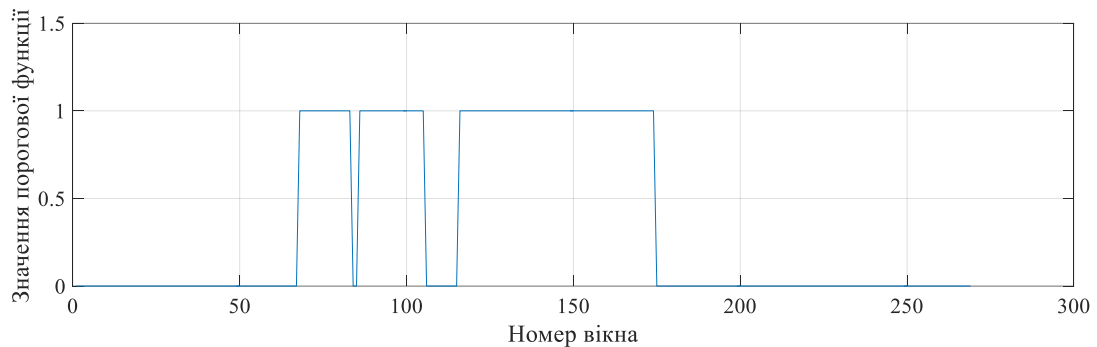
На рис. 3.14 наведено вигляд сигналу третьої команди «взад», графік присутності частоти основного тону в сигналі та вигляд порогової функції.



а)



б)

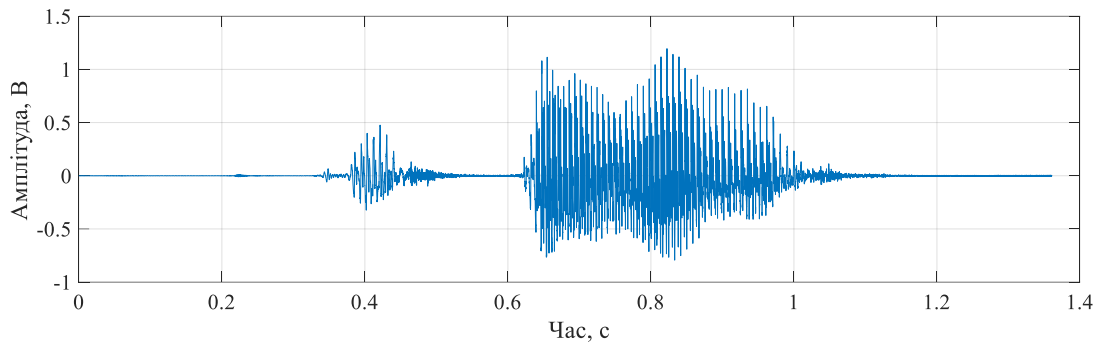


в)

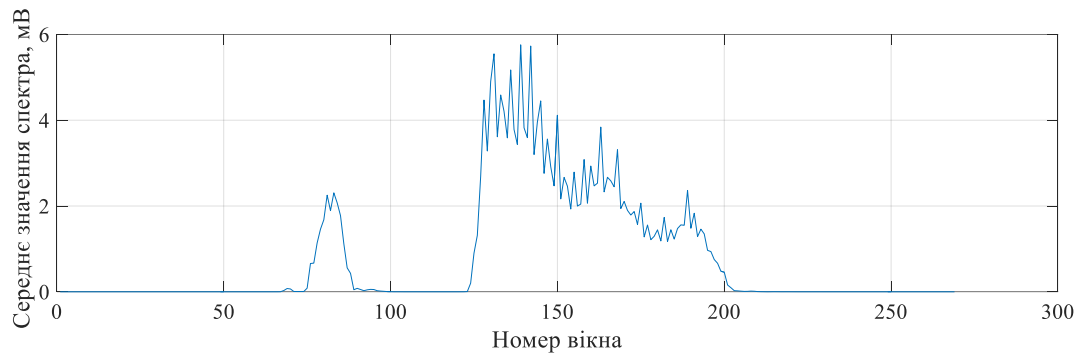
Рис. 3.14. Вигляд сигналу команди «взад» (а), графік присутності частоти основного тону в сигналі (б) та вигляд порогової функції (в)

На рисунку можна побачити міжфонемний пропуск.

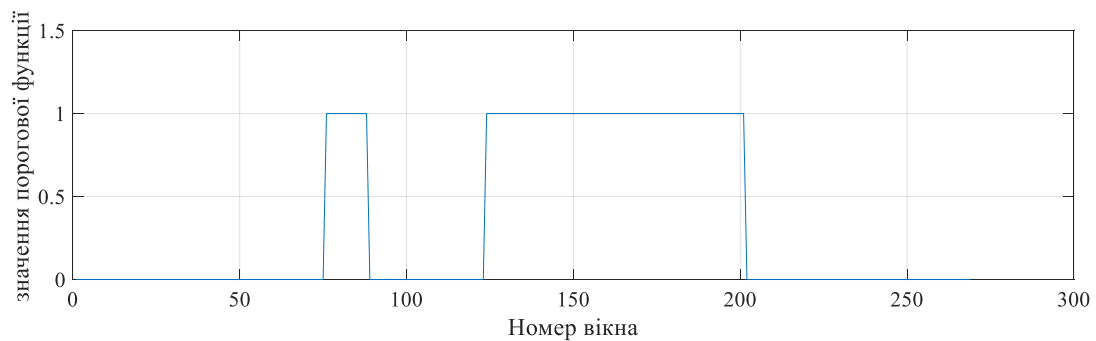
На рис. 3.15 наведено вигляд сигналу третьої команди «вперед», графік присутності частоти основного тону в сигналі та вигляд порогової функції.



а)



б)



в)

Рис. 3.15. Вигляд сигналу команди «вперед» (а), графік присутності частоти основного тону в сигналі (б) та вигляд порогової функції (в)

На рисунку також чітко видно пропуск через присутність глухої фонемі. Окрім цього аналізуючи наведені графіки для чотирьох команд можна прийти до висновку, що на значеннями порогової функції та тривалостями команд можна проводити їх ідентифікацію і наступне формування відповідних сигналів керування.

### 3.3 Висновки до розділу 3

Розроблено метод обробки голосових сигналів для ідентифікації окремих команд та формування на їх основі сигналів керування. Суть методу зводиться до наступного. Потрібно розпізнати чотири команди в голосовому сигналі: «вліво», «вправо», «взад», «вперед». Спочатку користувач промовляє ці слова в мікрофон, система опрацювання їх обробляє та виділяє інформативні ознаки, за якими в наступному буде проводити ідентифікацію цих команд.

Спочатку проводиться запис сигналу, потім проводиться його фільтрація адаптивними методами. Далше проводиться виділення довжини слова кожної команди. На наступному етапі проводиться сегментація кожного слова на фонемі. Власне на останньому етапі проводиться виділення ділянок з ознаками голосних та приголосних вокалізованих звуків

Для відбору голосових сигналів з командами використано стандартну комп'ютерну гарнітуру та програму Adobe Audition. Для адаптивної фільтрації застосовано метод спектрального віднімання.

Для власне опрацювання застосовано метод ковзного вікна. Для визначення тривалості команд проводилось обчислення середнього значення сигналу всередині ковзного вікна та формувалась порогова функція із значенням 1 та 0 залежно від того більше середнє значення за поріг чи менше.

Для сегментації сигналу команди на фонемі проведено визначення частоти основного тону методом формантного аналізу та встановлення присутності її в ковзному вікні. Використано порогову функцію яка рівна 1 коли ця частота є присутньою та 0 коли її немає. Аналізуючи результати опрацювання чотирьох команд зроблено висновок, що на значеннями порогової функції та тривалостями команд можна проводити їх ідентифікацію і наступне формування відповідних сигналів керування.

## РОЗДІЛ 4

### ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

#### 4.1 Охорона праці

Електромагнітне випромінювання в оптичній області, що примикає з боку коротких хвиль до видимого світла і має довжину хвиль в діапазоні 200...400 нм, називають ультрафіолетовим випромінюванням (УФВ). Вплив його на людину оцінюють еритемною дією (почервоніння шкіри, що приводить через 48 годин до її пігментації - засмазі).

Проблема ультрафіолетового випромінювання як виробничого та екологічного чинника обумовлена широким використанням джерел постачання в народному господарстві, збільшенням рівнів сонячного випромінювання у зв'язку зі зменшенням озонового шару, зростанням кількості захворювань, зокрема злоякісних і доброякісних пухлин шкіри, та інших порушень стану здоров'я, що викликаються ультрафіолетовою радіацією.

При тривалій відсутності УФВ в організмі розвивається «світлове голодування». Тому воно необхідно для нормальної життєдіяльності людини. Однак, при тривалому впливі великих доз УФВ можуть наступити серйозні поразки очей і шкіри. Зокрема, це може призвести до розвитку раку шкіри, кератитів (запалень рогівки) і помутніння кристалика очей (фотокератиту, який характеризується прихованим періодом від 0,5 до 24 годин). Для профілактики несприятливих наслідків, викликаних дефіцитом УФВ, використовують сонячне випромінювання, влаштовуючи солярії, інсоляцію приміщень, а також застосовуючи штучні джерела УФВ.

На промислових підприємствах джерелами ультрафіолетових випромінювань є дуга електрозварювання, ртутно-кварцові лампи, лазери, інші прилади та установки. Формування й вплив на працюючих оптичного



випромінювання в ультрафіолетовій області відбувається при електрогазозварювальних процесах, на роботах з плазменними технологіями (різка металу, напилювання, наплавлення металу), при використанні різних світильників та випромінювачів з кварцовими, ртутними, галогенними лампами, інших спектральних джерел. У різних галузях економіки та народного господарства широке застосування знаходять такі сучасні технології, як ультрафіолетове сушіння, установки для знезараження повітря, поверхонь та води, різні медичні та інші випромінювачі (перукарське устаткування, манікюрні лампи, солярії та інші).

Професіональному впливу УФВ піддаються електрогазозварювальники, копіювальники друкованих форм, працівники тепличних господарств, медичний персонал (фізіотерапевти, стоматологи, педіатри) та інші працівники, обслуговуючі різні джерела ультрафіолетового випромінювання. З іншого боку, при дорожніх, сільськогосподарських, будівельних та інших видах робіт, виконуваних на свіжому повітрі, відбувається вплив на працюючих природного УФ-випромінювання, як складової сонячної радіації. Окремо слід виділити групу працівників різних професій (звані "прихвачувальники"), які виконують спільні зі зварником роботи з фіксації деталей великогабаритних конструкцій в останній момент накладення первинного шва. Ці роботи виконують самі зварювальники (різних спеціальностей), і працівники інших професій - слюсарі механозбиральних робіт, монтажники та інші. Особливість таких робіт - короткочасність використання зварювальної дуги, її "імпульсний" характер під час "прихвачування" деталей зварювальної конструкції. Зазначені роботи, необхідно виконувати в окулярах із захисними фільтрами.

При впливі надлишкового ультрафіолетового випромінювання можливий розвиток низки захворювань і патологічних станів, насамперед, із боку органу зору, серед яких найчастіше відзначаються катаракта чи помутніння кришталика очі, запалення роговиці (кератит), слизових оболонок

(фотоофтальмія). УФ-переопромінювання може призвести до хвороб шкірних покривів: запалювальне почервоніння шкіри чи еритема, прискорення старіння шкіри, алергічні реакції, пухлини шкіри, в тому числі злоякісні (рак шкіри, меланома).

До засобів колективного захисту від УФВ відносяться різні пристрої (огороджувальні, вентиляційні, автоматичного контролю і сигналізації, дистанційного управління), а також знаки безпеки. Індивідуальний захист від УФВ здійснюють різними екранами: фізичними (у вигляді різних предметів, що поглинають, розсіюють або відображають промені) і хімічними (хімічні речовини та захисні креми, що містять інгредієнти, які поглинають УФВ). Для захисту також використовують виготовлений із тканини (попліну та ін) спеціальний одяг, окуляри із захисними фільтрами. Повний захист від УФВ усіх хвиль забезпечує флінтглас (скло, що містить окис свинцю) товщиною 2 мм. При влаштуванні приміщень враховують, що відображуюча властивість різних оздоблювальних матеріалів для УФВ і видимого світла різна.

## 4.2 Безпека в надзвичайних ситуаціях

### *Режим зони надзвичайної екологічної ситуації*

Режим зони надзвичайної екологічної ситуації - це особливий правовий режим, який може тимчасово запроваджуватися в окремих місцевостях у разі виникнення надзвичайних екологічних ситуацій і спрямовується для попередження людських і матеріальних витрат, відвернення загрози життю і здоров'ю громадян, а також усунення негативних наслідків надзвичайної екологічної ситуації.

Запровадження відповідного правового режиму передбачає виділення державою (або органами місцевого самоврядування) додаткових фінансових та інших матеріальних ресурсів, достатніх для нормалізації екологічного стану і

відшкодування завданих збитків, запровадження спеціального режиму поставок продукції для державних потреб, реалізації комплексних та цільових програм громадських робіт.

Законодавство про зону надзвичайної екологічної ситуації становлять:

- Закон України від 25 червня 1991 року "Про охорону навколишнього природного середовища";
- від 14 грудня 1999 року "Про аварійно-рятувальні служби";
- від 16 березня 2000 року "Про правовий режим надзвичайного стану";
- від 13 липня 2000 року "Про зону надзвичайної екологічної ситуації";
- а також прийняті відповідно до них нормативно-правові акти.

Підставами для оголошення окремої місцевості зоною надзвичайної екологічної ситуації можуть бути:

- значне перевищення гранично допустимих норм показників якості навколишнього природного середовища, визначених законодавством;
- виникнення реальної загрози життю та здоров'ю великої кількості людей або заподіяння значної матеріальної шкоди юридичним, фізичним особам чи навколишньому природному середовищу внаслідок надмірного забруднення навколишнього природного середовища, руйнівного впливу стихійних сил природи чи інших факторів;
- негативні зміни, які сталися у навколишньому природному середовищі на значній території і які неможливо усунути без застосування надзвичайних заходів з боку держави, або які суттєво обмежують чи виключають можливість проживання населення і провадження господарської діяльності на відповідній території;
- значне збільшення рівня захворюваності населення внаслідок негативних змін у навколишньому природному середовищі. Окрема місцевість України оголошується зоною надзвичайної екологічної ситуації Указом Президента України, затвердженим Верховною Радою України за пропозицією

Ради національної безпеки і оборони України або за поданням Кабінету Міністрів України.

В такому Указі Президента України має бути зазначено:

- обставини, що стали причиною та обґрунтуванням необхідності оголошення окремої місцевості зоною надзвичайної екологічної ситуації;
- межі території, на якій вона оголошується;
- заходи щодо організаційного, фінансового та матеріально-технічного забезпечення життєдіяльності населення в такій зоні;
- основні заходи, що запроваджуються для подолання наслідків надзвичайної екологічної ситуації;
- обмеження на певні види діяльності в цій зоні;
- час, з якого окрема місцевість оголошується зоною надзвичайної екологічної ситуації;
- строк, на який ця територія оголошується такою зоною.

За наявності достатніх підстав у межах зони надзвичайної екологічної ситуації може бути введений правовий режим надзвичайного стану в порядку, встановленому відповідним законом із запровадженням додаткових заходів.

Юридичні та фізичні особи, винні у порушенні правового режиму в зоні надзвичайної екологічної ситуації, несуть відповідальність згідно з законами України.

#### 4.3 Висновки до розділу

В розділі «Охорона праці та безпека в надзвичайних ситуаціях» описано вплив ультразвуку на організм людини, також режими зони надзвичайної екологічної ситуації.

## ВИСНОВКИ

Побудова системи голосового керування є актуальним завданням. Ці системи допомагають при роботі з комп'ютеризованими системами, як то голосовий набір тексту, керування електронними приладами в будинку тощо. Актуальною ця технологія є, наприклад, коли йдеться про людей із порушеннями опорно-рухового апарату. Як приклад розглядається задача керування електричним інвалідним кліслом.

В таких системах проводиться опрацювання голосових сигналів та виділення в їхній структурі ознак, за якими в наступному проводиться ідентифікація чи розрізнення окремих команд, що є певними ключовими словами. Для цього можуть застосовуватись знаки фонемного складу мови, амплітудні та часові параметри голосового сигналу, його склад, переходи між фонемами, складами чи словами.

Проаналізовано типову архітектуру систем голосового керування. Встановлено, що як метод розпізнавання більшість сучасних систем використовують метод прихованих марківських моделей.

Для розроблення методу обробки мовних сигналів для формування сигналів керування проведено аналіз процесу формування мовних сигналів та виділено їхні параметри і характеристики, що можуть бути використані для розпізнавання окремих команд.

Розглянуто особливості систем розпізнавання мови. Розглянуто алгоритми виділення елементів слів (фонем, складів тощо). Спрощено процес розпізнавання включає шумоочищення та відділення корисного сигналу. Проаналізовано спосіб розпізнавання голосових команд, що ґрунтується на біометричній системі, та двохрівневий метод розпізнавання голосової команди. Використання останньої системи полягає у вимові команди, яка складається з

ключового слова (за допомогою нього система розуміє, що звертаються до неї) та власне команди.

Розроблено метод обробки голосових сигналів для ідентифікації окремих команд та формування на їх основі сигналів керування. Суть методу зводиться до наступного. Потрібно розпізнати чотири команди в голосовому сигналі: «вліво», «вправо», «взад», «вперед». Спочатку користувач промовляє ці слова в мікрофон, система опрацювання їх обробляє та виділяє інформативні ознаки, за якими в наступному буде проводити ідентифікацію цих команд.

Спочатку проводиться запис сигналу, потім проводиться його фільтрація адаптивними методами. Далше проводиться виділення довжини слова кожної команди. На наступному етапі проводиться сегментація кожного слова на фонемі. Власне на останньому етапі проводиться виділення ділянок з ознаками голосних та приголосних вокалізованих звуків

Для відбору голосових сигналів з командами використано стандартну комп'ютерну гарнітуру та програму Adobe Audition. Для адаптивної фільтрації застосовано метод спектрального віднімання.

Для власне опрацювання застосовано метод ковзного вікна. Для визначення тривалості команд проводилось обчислення середнього значення сигналу всередині ковзного вікна та формувалась порогова функція із значенням 1 та 0 залежно від того більше середнє значення за поріг чи менше.

Для сегментації сигналу команди на фонемі проведено визначення частоти основного тону методом формантного аналізу та встановлення присутності її в ковзному вікні. Використано порогову функцію яка рівна 1 коли ця частота є присутньою та 0 коли її немає. Аналізуючи результати опрацювання чотирьох команд зроблено висновок, що на значеннями порогової функції та тривалостями команд можна проводити їх ідентифікацію і наступне формування відповідних сигналів керування.



## ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Чумаков В. І., Таранчук А. А., Харченко О. І. Моделювання пристроїв радіоавтоматики в системі MathCAD : навч. посібник / В. І. Чумаков, А. А. Таранчук, О. І. Харченко. -Хмельницький : ХНУ, 2011. - 151 с.
2. Лозинський А.О. Системи керування електропобутовими приладами: навч. посібник / А.О. Лозинський, Б.Л. Копчак, В.В. Бушер. - Львів: Видавництво Львівської політехніки, 2010.-304 с.
3. Локазюк В. М., Іванов О. В., Тітова В. Ю. Засади систем підтримки прийняття рішень на основі комп'ютерних систем та їх компонентів : Навч. посібник для вузів. -Хмельницький: ПП Гонта А.С., 2010. - 337 с.
4. Дідковський В.С. та ін. Шуми і вібрації: Підручник / В. С. Дідковський, О. В. Коржик, О. Г. Лейко. - К.: ТОВ "Імекс-ЛТД", 2010. - 336 с: іл.
5. Волошин, О. Ф. Моделі та методи прийняття рішень : навч. посіб. для студ. вищ. навч. закл. / О. Ф. Волошин, С. О. Мащенко. - 2-ге вид., перероб. та допов. - К. : Видавничо-поліграфічний центр "Київський університет", 2010. - 336 с.
6. Калюжний О.Я. Моделювання систем передачі сигналів в обчислювальному середовищі MATLAB-Simulink. – К., “Політехніка”, 2004. – 135 с.
7. F. Jelinek «Continuous Speech Recognition by Statisical Methods.» IEEE Proceedings 64:4(1976): 532-556 с.
8. Дедів І. Обґрунтування математичної моделі дихальних шумів у вигляді періодично корельованого випадкового процесу / Я. Драган, І. Дедів // Науковий вісник Чернівецького університету. Вип. 426: Фізика. Електроніка. ЧНУ – Чернівці: Рута. – 2008. – Ч. II. – С. 93-97.



9. Джичка Н., Дедів І., Дозорський В., Драган Я. Модель акустичного сигналу для виявлення порушень стану дихальної системи та голосового апарату як частковий випадок стохастичної коливної системи. Вісн. Нац. ун-ту "Львів. політехніка". 2011. № 710. С. 155-158.

10. Бачинський, М.В. Комп'ютерна імітаційна модель вокалізованих фрикативних звуків / М.В. Бачинський, Л.Є. Дедів, В.Г. Дозорський // Вісник Сумського державного університету. Технічні науки. – Суми : видавництво СумДУ, 2012. – № 1. – С. 149–156.

11. Дедів Л.Є., Дозорський В.Г., Бачинський М.В. Математична модель електроенцефалографічного сигналу для задач побудови комп'ютерних діагностичних систем. Вісник Хмельницького національного університету. 2012. №2. С. 186-189.

12. Драган, Я. Метод опрацювання фрикативних звуків для діагностики захворювань органів голосового апарату на ранніх стадіях / Я. Драган, В. Дозорський, М. Хвостівський, І. Дедів // Вісник Національного університету "Львівська політехніка". Комп'ютерні науки та інформаційні технології. – Львів : НУЛП, 2011. – № 694. – С. 376–382.

13. Palaniza Y.B., Shadrina H.M., Khvostivskiy M.O., Dediv L.Ye., Dozorska O.F. Main theoretical basis of biosignals modeling. Znanstvena misel. Slovenia. 2018. №16. P. 39-44.

14. Vyacheslav Nykytyuk, Vasyl Dozorskyi, Oksana Dozorska. Detection of biomedical signals disruption using a sliding window. Scientific journal of the Ternopil National Technical University. 2018. Vol. 91. № 3. P. 125–133.

15. Oksana Dozorska, Evhenia Yavorska, Vasil Dozorskyi, Iryna Pankiv, Iryna Dediv, Leonid Dediv. The Method of Indirect Restoration of Human Communicative Function. 15th International Conference "The Experience of Designing and Application of CAD Systems" (CADSM) (Polyana, Svalyava, 26.02.2019-02.03.2019). Zakarpattya, 2019. P.19-22.

## ДОДАТКИ

## Текст програми обробки голосового сигналу

```
%clear all
x11=data;
xrr=x11;
t1=(0:(length(x11)-1))./44100;
figure(1);
plot(xrr);
grid on;
ta=220;
mm=0;
for kk=0:1813
    x112=abs(xrr((kk*ta+1):(kk*ta+250)));
    hhh=max(x112);
    if hhh>=0.006
        ggg=1;
    else
        ggg=0;
    end
    mm=mm+1;
    p(mm)=ggg;
end
figure(2)
plot(p)
grid on
x1=data(300000:360000);
xr=x1.*2;
```

```
t11=(0:(length(xr)-1))./44100;
figure(3);
plot(t11,xr);
grid on;
tau=221;
m=0;
for k=0:268
    x12=xr((k*tau+1):(k*tau+500));
    z1=xcorr(x12);
    fd=44100;
    N=length(z1);
    y1=fft(z1,N);
    y=abs(y1)/N*2;
    df=fd/N;
    f=0:df:fd-df;
    n1=fix(200/df);
    n2=fix(400/df);
    hh=mean(y(n1:n2));
    if hh>=0.1
        gg=1;
    else
        gg=0;
    end
    m=m+1;
    b(m)=gg;
    r(m)=hh;
end
figure(4)
```

```
plot(b)
```

```
grid on
```

```
figure(5)
```

```
plot(r)
```

```
grid on
```

УДК 057.087

**М.П. Мотелюк, С.Т. Боїло, І.Ю. Дедів, к.т.н., доцент, В.Г. Дозорський, к.т.н., доцент**  
Тернопільський національний технічний університет імені Івана Пулюя, Україна

## **МЕТОДИ ОБРОБКИ МОВНИХ СИГНАЛІВ ДЛЯ БЕЗПЕКОВИХ СИСТЕМ**

**M.P. Motelyuk, S.T. Boilo, I.Yu. Dediv, Ph.D., Assoc. Prof., V.G. Dozorskyi, Ph.D., Assoc. Prof.**

## **METHODS OF PROCESSING SPEECH SIGNALS FOR SECURITY SYSTEMS**

Голос і мова людини несуть індивідуальну інформацію через унікальність фізіологічної будови її артикуляторного апарату та специфіки мови. Цей тип сигналів стає сьогодні особливо поширеним при розробці безпекових біометричних систем, що реалізують процедури верифікації та ідентифікації диктора для різних сервісів. Особливість голосової біометрики полягає в тому, що вона допускає віддалену і приховану автентифікацію за допомогою простих і доступних сенсорів, що іноді неможливо або дорого для отримання іншої біометричної інформації. Зручність для користувача, простота, здатність легко інтегруватися з іншими методами - також важливі фактори, що визначають перспективність мовних технологій у біометричних системах як окремо, так і комплексно з іншими методами верифікації/ідентифікації особистості.

Додатки систем верифікації та ідентифікації можуть бути найрізноманітнішими — від систем локальної або віддаленої (телефоном) авторизації особистості, пов'язаної з наданням прав (використовуються при наданні допуску до охоронних об'єктів, або, наприклад, до інформації та фінансових операцій) до юридичних аспектів автентифікації особи у судовій практиці. Надійність верифікації чи ідентифікації, а також вартість реалізації – важливі питання, вирішення яких залежить від конкретної програми та наявних альтернатив. Також, системи автоматичної верифікації диктора за мовним сигналом забезпечують надійність, яка є порівнянна з надійністю прийняття рішення людиною, добре знайомою із голосом диктора, а в деяких ситуаціях перевищують по точності рішення людини (особливо при верифікації телефоном).

В загальному випадку мовленнєва фраза, що є об'єктом аналізу та прийняття рішення при розпізнаванні диктора, може мати фіксований характер (пароль), бути обраною системою за випадковим законом із заданого набору або бути довільною. Також, враховуючи природу та характер утворення мовний сигнал описується складним стохастичним процесом із відповідною морфологією та спектральним складом.

В дослідженнях проводиться розроблення методу виділення обвідної складової мовних сигналів та алгоритмів сегментації його на окремі звуки для часткового випадку розпізнавання мови в області розробки безпекових біометричних систем.

### **Література**

1. Біометрія / М.П. Горошко, С.І. Миклуш, П.Г. Хомюк. – Львів : Вид-во "Камула", 2004. – 236 с.
2. Біометричні технології в ХХІ столітті та їх використання правоохоронними органами: посібник / В. П. Захаров, В. І. Рудешко; Львів. держ. ун-т внутр. справ. – 2-ге вид., допов. – Львів: ЛьвДУВС, 2015. – 491 с.
3. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов /Пер. с англ.; Под ред. М.В. Назарова, Ю.Н. Прохорова. К.: Радио и связь, 1981
4. Левин К.Е., Никитин О.Р. Моделирование устройств обработки речевых сигналов. Сборник научных трудов «Методы и устройства».