

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)

« » 2021 р.

**З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня магістр

(назва освітнього ступеня)

за спеціальністю 172 Телекомунікації та радіотехніка

(шифр і назва спеціальності)

студенту Андрейчуку Богдану Валерійовичу

(прізвище, ім'я, по батькові)

1. Тема роботи Метод розпізнавання голосових сигналів для керування комп'ютерними системами

Керівник роботи Дунець Василь Любомирович, к.т.н.

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

Затверджені наказом ректора від « 30 » листопада 2021 року № 4/7-1019

2. Термін подання студентом завершеної роботи _____

3. Вихідні дані до роботи Об'єкт дослідження: процес опрацювання мовних сигналів у задачах розпізнавання мовлення для керування комп'ютерними системами.

Предмет дослідження: методи розпізнавання мовних сигналів з використанням середньоквадратичного відхилення, евклідова відстань.

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)

6. Консультанти розділів роботи

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|--|---|----------------|------------------|
| | | завдання видав | завдання прийняв |
| <i>Охорона праці та безпека в надзвичайних ситуаціях</i> | <i>Клепчик В.М., стар. викл. каф. ОХ</i> | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

7. Дата видачі завдання 30.10.2021р.

КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів роботи | Термін виконання етапів роботи | Примітка |
|-------|---|--------------------------------|-----------------|
| 1 | <i>Отримання завдання</i> | 30.10.2021 | <i>виконано</i> |
| 2 | <i>Аналіз завдання</i> | 02.11.2021 | <i>виконано</i> |
| 3 | <i>Виконання розділу 1</i> | 09.11.2021 | <i>виконано</i> |
| 4 | <i>Виконання розділу 2</i> | 15.11.2021 | <i>виконано</i> |
| 5 | <i>Виконання розділу 3</i> | 25.11.2021 | <i>виконано</i> |
| 6 | <i>Виконання розділу 4</i> | 28.11.2021 | <i>виконано</i> |
| 7 | <i>Оформлення пояснювальної записки</i> | 30.11.2021 | <i>виконано</i> |
| 8 | <i>Оформлення графічного та презентаційного матеріалу</i> | 02.12.2021 | <i>виконано</i> |
| 9 | <i>Перевірка роботи на антиплагіат</i> | 16.12.2021 | <i>виконано</i> |
| 10 | <i>Попередній захист</i> | 16.12.2021 | <i>виконано</i> |
| 11 | <i>Захист</i> | 20.12.2021 | <i>виконано</i> |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Студент

(підпис)

Андрейчук Б.В.

(прізвище та ініціали)

Керівник роботи

(підпис)

Дунець В.Л.

(прізвище та ініціали)

АНОТАЦІЯ

Тема кваліфікаційної роботи: «Метод розпізнавання голосових сигналів для керування комп'ютерними системами» // Кваліфікаційна робота // Андрейчук Богдан Валерійович // Тернопільський національний технічний університет імені Івана Пулюя, факультет прикладних інформаційних технологій та електроінженерії, група РРм-61 // Тернопіль, 2021 // с. – 95, рис. – 46, табл. – 7, додат. – 1, бібліогр. – 52.

Ключові слова: МОВНИЙ СИГНАЛ, СПЕКТРАЛЬНИЙ АНАЛІЗ, ПЕРЕТВОРЕННЯ ФУР'Є, ВЕЙВЛЕТ АНАЛІЗ, КЕПСТРАЛЬНИЙ АНАЛІЗ, ЄВКЛІДОВА ВІДСТАНЬ.

В кваліфікаційній роботі здійснено порівняльний аналіз застосування різних вимірювань близькості та векторних ознак, який показав, що найбільш придатними для завдань розпізнавання векторами ознак можна вважати: мел-кепстральні коефіцієнти та розподіл інформаційних інтервалів мовного сигналу для керування комп'ютерними системами.

ANNOTATION

Theme of qualification work: " Method of recognition of voice signals for control of computer systems" // Qualification work // Andreychuk Bogdan Valeriyovych // Ternopil National Technical University named after Ivan Pulyuy, Faculty of Applied Information Technologies and Electrical Engineering, group PPM-61 // Ternopil, 2021 // with. - 95, fig. - 46, table. - 7, appendix. - 1, bibliogr. - 52.

Keywords: SPEECH SIGNAL, SPECTRAL ANALYSIS, FOURET'S TRANSFORMATION, WAVELET ANALYSIS, CAPSTRAL ANALYSIS.

In the qualification work, a comparative analysis of the use of different measurements of proximity and vector features, which showed that the most suitable for recognition tasks vector features can be considered: mel-keppstral coefficients and distribution of information intervals of speech signal to control computer systems.

ЗМІСТ

| | |
|--|----|
| ВСТУП..... | 8 |
| РОЗДІ 1. ОСНОВНА ЧАСТИНА..... | 11 |
| 1.1. Сучасний стан напрямку розпізнавання мовних сигналів..... | 11 |
| 1.2. Особливості мовлення та сприйняття мови людиною..... | 16 |
| 1.2.1. Мовний апарат..... | 17 |
| 1.2.2. Сприйняття мовного сигналу людиною..... | 20 |
| 1.3. Методи цифрової обробки сигналів у задачах розпізнавання мовних сигналів..... | 26 |
| 1.3.1. Спектральний аналіз..... | 26 |
| 1.3.2. Віконний аналіз у базисі Фур'є..... | 27 |
| 1.3.3. Вейвлет аналіз..... | 27 |
| 1.3.4. Кепстральний аналіз..... | 29 |
| 1.4. Субсмуговий підхід до обробки мовних сигналів..... | 32 |
| 1.5. Висновки до розділу 1..... | 33 |
| РОЗДІЛ 2. ОСНОВНА ЧАСТИНА..... | 34 |
| 2.1. Акустико-фонетичний підхід до розпізнавання мовних сигналів..... | 34 |
| 2.2. Обчислювальні аспекти субсмугового аналізу мовних сигналів у задачах ідентифікації..... | 37 |
| 2.3. Дослідження просторів ознак у задачах розпізнавання мовних сигналів..... | 43 |
| 2.3.1. Декомпозиція сигналу банком фільтрів..... | 43 |
| 2.3.2. Розподіл миттєвих енергій відрізка МС..... | 45 |
| 2.3.3. Розподіл часток енергії відрізка МС..... | 47 |
| 2.3.4. Розподіл інформаційних інтервалів відрізка МС..... | 49 |
| 2.3.5. Частота переходів через нуль..... | 52 |
| 2.3.6. Ширина частотної області, що займає сигнал..... | 55 |
| 2.3.7. Мел-кепстральні коефіцієнти мовного сигналу..... | 60 |
| 2.4. Заходи близькості у задачах розпізнавання мовних сигналів..... | 63 |

| | |
|--|----|
| 2.4.1. Євклідова відстань..... | 7 |
| 2.4.2. Середньоквадратичне відхилення..... | 63 |
| 2.4.3. Відстань Махаланобіса..... | 63 |
| 2.4.4. Кореляція послідовностей..... | 64 |
| 2.4.5. Динамічна трансформація тимчасової шкали..... | 64 |
| 2.5. Висновки до розділу 2..... | 65 |
| РОЗДІЛ 3. НАУКОВО-ДОСЛІДНА ЧАСТИНА..... | 67 |
| 3.1. Методика оцінки методів розпізнавання мовних сигналів..... | 68 |
| 3.2. Дослідження підходів до розпізнавання мовних сигналів..... | 72 |
| 3.3. Висновки до розділу 3..... | 80 |
| РОЗДІЛ 4. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ..... | 81 |
| 4.1. Охорона праці..... | 81 |
| 4.2. Безпека в надзвичайних ситуаціях..... | 84 |
| 4.3. Висновки до розділу 4..... | 86 |
| ВИСНОВКИ..... | 87 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... | 88 |
| Додаток А. Копія тези конференції..... | 93 |

ВСТУП

Актуальність роботи. З розвитком комп'ютерних систем, у світі, стає дедалі очевидним, використання цих систем у повсякденної діяльності буде розширюватися. Важливим фактором для розвитку комп'ютерних систем є можливість використання людської мови як інтерфейсу для роботи з комп'ютером: управління персональним комп'ютером голосом у реальному часі, а також введення та виведення інформації у вигляді усного мовлення.

На даний момент, повсюдно, ведуться роботи зі створення систем обробки мовлення, серед яких особливе місце займає розпізнавання мови. Система розпізнавання мови отримує інформацію про акустичні коливання повітря через мікрофон, порівнює отримані дані з наявними в системі і у разі збігу ідентифікує ділянку сигналу. Для обліку варіативності та навчання моделей фонем та слів потрібні великі обсяги тексту та мовного матеріалу, підготовка яких потребує величезних трудовитрат. Сучасні системи розпізнавання навчання на обмежених мовних корпусах мають ряд недоліків.

Сьогодні існує два основних види технологій розпізнавання мови. Один із них — це розпізнавання мови, яка залежить від диктора, тобто користувач повинен спочатку навчити систему розпізнавати його голос, і лише після цього система може функціонувати. Другий - це розпізнавання мови, що не залежить від диктора, тобто система здатна розпізнати будь-яку мову, незалежно від того, хто говорить. Системи розпізнавання ізольованих слів працюють із дискретними словами — у разі потрібна пауза між словами.

Існуючі системи розпізнавання створені для роботи з англійською, німецькою, іспанською та іншими популярними мовами та малозастосовні у роботі з російською мовою. Це пов'язано з тим, що російська мова принципово відрізняється від інших мов не тільки фонетично, а й вільним порядком слів у реченні, що значно ускладнює математичну мовну модель. Важливим завданням при розробці систем розпізнавання голосових сигналів, є виділення таких ознак, які мали 1) властивістю інваріантності на відрізках, отриманих при

виголошенні одних і тих же звуків російської мови, 2) і варіативності на відрізках, що містять різні звуки. З іншого боку, потрібні підходи до порівняння даних ознак – тобто. визначення адекватних заходів близькості.

В основі багатьох з розроблених підходів [25,26,49,50] використовуються частотні уявлення, так як породжувані звуками мови відрізки РС мають властивість концентрації енергії в досить вузьких смугах частотної осі. У випадку з цим можна згадати розбиття частотної смуги, що розглядається в літературних джерелах, на так звані критичні смуги слуху, які опосередковано відбиваються на частотних властивостях РС.

Слід зазначити, що запропоновані нині методи розпізнавання РС з урахуванням аналізу їх частотних властивостей, як ознакового простору, або відбивають властивості концентрації енергії, або недостатньо точно відображають характер зміни у мовному сигналі.

Метою роботи є визначення важливих, з погляду розв'язання задачі розпізнавання мови – ознак мовних сигналів та заходів їхньої близькості.

Для досягнення мети необхідно вирішити такі завдання:

1. Проаналізувати особливості обробки мовних сигналів у задачах розпізнавання мови;
2. Вивчити існуючі методи представлення мовних сигналів у завданнях розпізнавання та провести їх порівняльний аналіз;
3. Вивчити заходи близькості, що застосовуються порівняння ознак мовних сигналів у системах розпізнавання промови;
4. Визначити важливі завдання розпізнавання мови характеристики мовного сигналу.

Об'єкт дослідження: процес опрацювання мовних сигналів у задачах розпізнавання мовлення для керування комп'ютерними системами.

Предмет дослідження: методи розпізнавання мовних сигналів з використанням середньоквадратичного відхилення, евклідова відстань.

Методи дослідження: цифрова обробка сигналів. Matlab.

Наукова новизна отриманих результатів. Проведено порівняльний аналіз застосування різних методів опрацювання сигналів та їх ознак, який

показав, що найбільш придатними для завдань розпізнавання векторами ознак можна вважати: мел-кепстральні коефіцієнти та розподіл інформаційних інтервалів мовного сигналу для керування комп'ютерними системами.

Публікації.

Викладені в роботі результати доповідалися та обговорюватися на 9-ій Науково-технічній конференції «Інформаційні моделі, системи та технології» (м. Тернопіль, 8-9 грудня 2021 р.).

РОЗДІЛ 1

АНАЛІТИЧНА ЧАСТИНА

1.1. Сучасний стан напряму розпізнавання мовних сигналів

З початку розвитку людства до моменту експансії сучасних медіа технологій, людська мова залишається основним видом комунікаційної діяльності людей та використовується для обміну інформації та соціальної взаємодії щодня. Зважаючи на процеси комп'ютеризації та інформатизації суспільства, мова використовується в багатьох технологічних аспектах соціальної та інформаційної взаємодії людини: телефонні переговори, радіо, телебачення, Інтернет. Це говорить про безумовну важливість мовних технологій для суспільства, так і про провідне місце промови – як засоби комунікації для людини загалом.

На даний момент спостерігається необхідність розвитку методів та алгоритмів розпізнавання мовних сигналів для широкого спектру завдань та додатків:

- 1) мовні командні системи (управління транспортом, роботами, тощо);
- 2) системи перекладу мови в текст;
- 3) системи синхронного перекладу
- 4) системи допомоги людям з обмеженими можливостями;
- 5) розумна телефонія – голосовий набір, голосове меню; і багато іншого.

Процес розпізнавання мови можна описати наступною моделлю, що представлена на рисунку 1.1. Повідомлення (W) сформоване джерелом повідомлень передається на джерело мовного сигналу, де формується акустичний сигнал (x), що відповідає повідомленню яке кодується. Мовний сигнал (x) передається через комунікаційний канал (в даному випадку – ідеальний) надходить на приймач, де відбувається його обробка з метою отримати повідомлення (W), яке має відповідати переданому повідомленню (W). [50]



Рис. 1.1. Загальна схема розпізнавання мовних повідомлень

Реалізація системи розпізнавання мови може бути представлена наступною блок-схемою, зображеною рисунку 1.2. Мовний сигнал подається на блок обробки сигналу, де виконуються такі операції: оцифрування сигналу із заданою частотою дискретизації; представлення сигналу як деякий вектор ознак. Далі, на декодері, виконується операції зі зіставлення вхідного мовного сигналу з деякими зразками мовних повідомлень, клас яких наперед відомий.

Акустична модель служить визначення фонем, у тому числі складається мовної сигнал, шляхом їх класифікації виходячи з підходів теорії ймовірностей і математичної статистики. При цьому фонема є мінімальним елементом мови в мові, який може служити для визначення різниці одного слова від іншого.

Мовна модель служить для побудови слів і речень, отриманих в результаті роботи акустичної моделі, транскрипцією, використовуючи закладені в ній правила граматики та лінгвістики.

Таким чином, для розпізнавання мови необхідно виконати наступні етапи:

- 1) оцифрування мовного сигналу;
- 2) уявлення сигналу у вигляді деякого вектора ознак;
- 3) акустичний аналіз відрізків на підставі обраних ознак;
- 4) мовний аналіз отриманої транскрипції з метою формування слів і які мають сенс словосполучень та речень. [50].

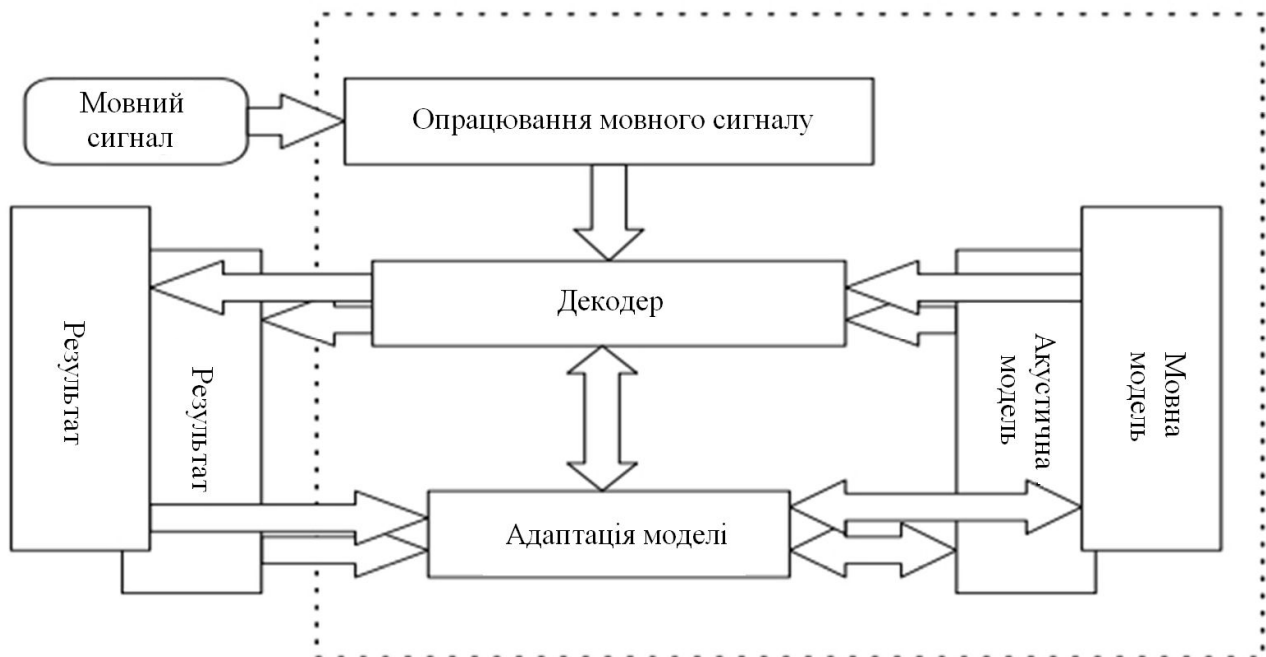


Рис. 1.2. Схема реалізації системи розпізнавання мовлення

Ефективність методів як розпізнавання мовлення загалом, прямо залежить від якості отриманого мовного сигналу (x), і навіть від типу уявлення мовного сигналу як певного набору його ознак, щоб забезпечити максимальну відстань між різними класами, тобто. від вибору ознакового простору, котрій формується вирішальна функція.

Мова являє собою послідовність звуків, що йдуть один за одним з розділеними паузами. Інформація у мовному сигналі представлена як акустичного коливання складної форми. У традиційних або вузькосмугових телефонних розмовах межа звукових частот знаходиться в діапазоні від 300 до 3400 Герц. Для передачі одного каналу голосової частоти, включаючи захисну смугу частот зазвичай виділяють смугу пропускання 4 кГц, що допускає частоту дискретизації 8 кГц.

Виходячи з теореми Котельникова, відомо, що частота дискретизації в 16 кГц є достатньою для кодування мовних коливань із частотою до 8 кГц. Зниження частоти дискретизації тягне за собою збільшення помилок при розпізнаванні мови через зниження розбірливості через втрату частини енергетичних складових звуків, розташованих у області високих частот, які стосуються наступним буквам українського алфавіту: “с”, “ч”, “ш”, “ф”.

Проведені американськими вченими дослідження [46] доводять, що оптимальною, з погляду розпізнавання мови є частота дискретизації 16 кГц, за глибини дискретизації 16 біт.

Таблиця 1.1

**Залежність розбірливості слів від частоти дискретизації
оцифрованого мовного сигналу**

| Частота | Рівень розбірливості слів |
|---------|---------------------------|
| 8 кГц | опорний рівень |
| 11 кГц | + 10 % |
| 16 кГц | + 10 % |
| 22 кГц | + 0 % |

Як видно з представленої таблиці, виграш у розпізнаванні порівняно з опорним рівнем 8 кГц дає використання частоти дискретизації 11 кГц та дає виграш 10 %. Збільшення частоти дискретизації рівня 16 кГц, дає додатковий виграш в 10%, але подальше збільшення частоти дискретизації не дає виграшу в розбірливості мови. [50]

Для вибору ознакового простору розробки алгоритму ідентифікації мовних даних необхідно вивчити їх структуру, природу формування, і навіть модель сприйняття людиною вимову речення.

Як було описано вище, при розпізнаванні мовних сигналів, як правило, оперують не з аналоговим мовним сигналом, а з так званим описом мовного сигналу, який економно представляє мовний сигнал і містить смислову та біометричну інформацію. Йдеться про обробку послідовностей відліків, що реєструються через певні проміжки часу:

$$\vec{x} = (x_1, x_2, \dots, x_N)^T, \quad (1.1)$$

$$x_i = x(i\Delta t), \quad i = 1 \dots N, \quad (1.2)$$

де N – тривалість сигналу у відліках; Δt – період дискретизації, що дорівнює

величині зворотної частоти дискретизації f_d :

$$\Delta t = 1/f_d \quad (1.3)$$

При аналізі мовного сигналу, він піддається сегментації ділянки певної тривалості – фрейми. Тривалість кадру становить близько 10-30 мс, це значення обрано у вигляді стаціонарності мовного сигналу на ділянках цієї тривалості.

Далі фрейми перетворюють з часової на частотну область за допомогою перетворення Фур'є :

$$X(z) = \sum_{k=1}^N x_k e^{-j \cdot z \cdot (k-1)} \quad (1.4)$$

де x_k – відліки аналізованого відрізка сигналу; N – тривалість вікна аналізу; j – уявна одиниця ($j^2 = -1$).

Так щоб близькість фреймів щодо Евклідова відстані:

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1.5)$$

відповідала близькості ділянок сигналу у часовій області.

Далі проводиться операція виділення векторів ознак і встановлення відповідності між фреймами певного прецеденту і досліджуваного мовного сигналу для здійснення класифікації. Проблема даного етапу у тому, що тривалості однієї й тієї ж слова (складу чи фонем) відрізняються ступенем стиснення чи розтягнення у часі.

Можливості комп'ютерних систем з розпізнавання мови обмежені, оскільки людина використовує для розпізнавання не тільки акустичну інформацію. Для розпізнавання мови людина використовує семантичну і синтаксичну зв'язку слів, і навіть візуальну інформацію (читання по губам,

міміка та інших.) – загальновідомо що, перебуваючи у галасливій обстановці мова співрозмовника легше розпізнавати якщо стежити за його губами.

Фонетичні моделі, що використовуються при створенні алгоритмів ідентифікації мови, не є точними, оскільки неможливо охопити всю різноманітність факторів. При задані фонетичних прецедентів використовують статистичні методи, які часто припускають, що параметри фонем розподілені за законом. Але насправді, точна модель прецедентів звуків і слів повинна включати безліч різноманітних варіантів, що не здійснимо в реальних системах.

Крім того, існує проблема дикторозалежності систем ідентифікації, будучи навченими за зразками одного диктора системи показують менш надійні результати при розпізнаванні мовних сигналів, сказаних іншим голосом.

Вище перелічені факти, свідчать, що розпізнавання відрізків мовних сигналів з допомогою ЕОМ має обмежену надійність, яку неможливо істотно підвищити ні шляхом вдосконалення алгоритмів, ні шляхом збільшення обчислювальних потужностей. Маючи на увазі це твердження, зробимо висновок що розробка нових підходів до ідентифікації повинна ґрунтуватися на чітко поставленому завданні, відштовхуючись від якого необхідно зробити вибір простору ознак для порівняння, а також алгоритмів та методів, що застосовуються для виявлення тотожності між прецедентом та об'єктом ідентифікації.

Як ознаки в задачах розпізнавання можуть використовуватися часові, частотні та інші характеристики мовного сигналу, які, різною мірою, відбивають особливості мовного апарату, характерні конкретного диктора. Щоб отримати уявлення про такі особливості, необхідно ознайомитися з особливостями сприйняття мови.

1.2. Особливості мовлення та сприйняття мови людиною

Однією з основних тем, пов'язаних із звуком, є сприйняття звуку людиною. Є недоцільним говорити про розпізнавання мови, не торкнувшись

озвученого питання.

1.2.1. Мовний апарат.

Відповідно до теорії мовоутворення мова є струмінь повітря, яка випромінюється системою органів: легкими, бронхами і трахеєю, та був перетворюється голосовим трактом на набір звуків.

У мовний апарат людини входять ротова та носова порожнини з підрядними порожнинами, горло (верхні резонатори), гортань з голосовими складками, трахея та бронхи (нижній резонатор), легені, грудна клітка з дихальними м'язами та нервова система організує їх функції у єдиний, цілий звукоутворення, що є складним психофізичним актом [1, 2].

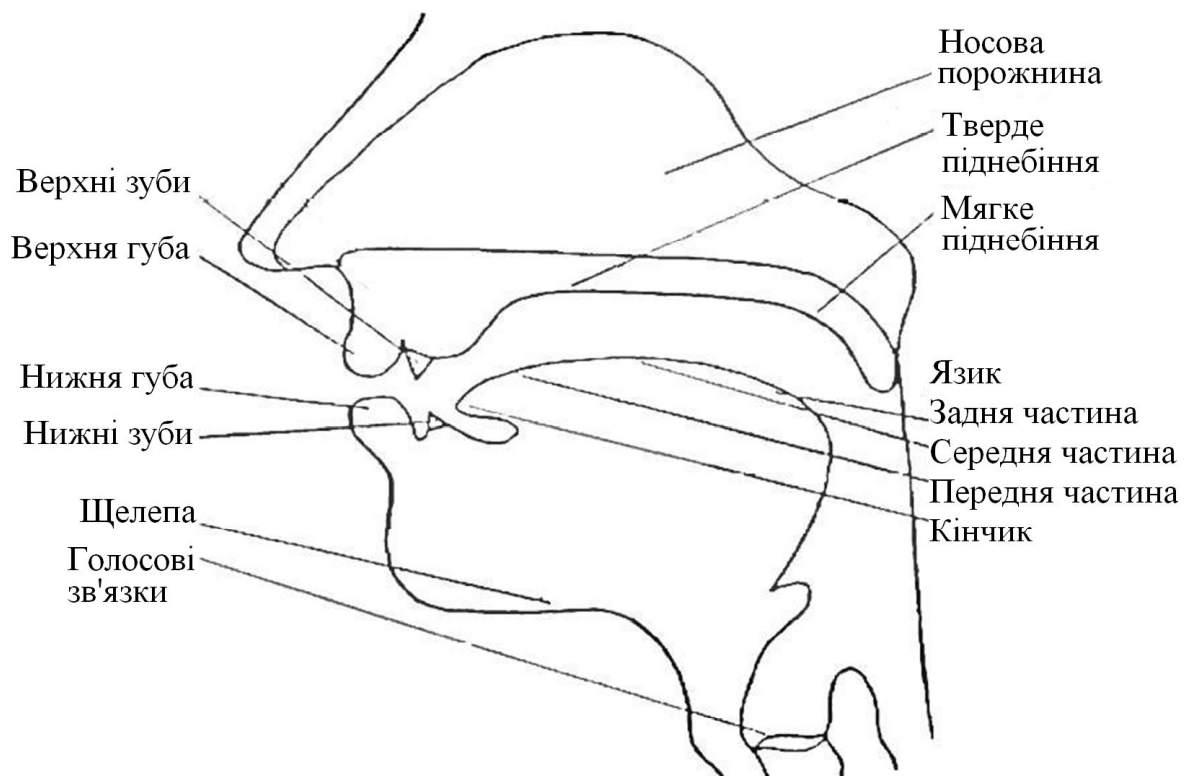


Рис. 1.3. Мовоутворюючі органи людини

Відповідно до рисунку 1.3, мовний тракт починається з голосових зв'язок і закінчується ротовою щілиною, його довжина у дорослої людини становить приблизно 17 см, площа поперечного перерізу тракту, яка визначається виходячи з положення мови, губ, щелеп та піднебінної фіранки, може змінюватися від 0 до 20 см².

Мовний апарат людини здатний породжувати акустичні коливання різної форми (шумоподібні або невокалізовані та квазіперіодичні або вокалізовані). Характер акустичних коливань, що породжуються, залежить від анатомії різних артикуляторів людини та їх точок торкання мовного тракту. [40]

Легкі – це джерело повітря та тиску у процесі мовоутворення.

Голосові зв'язки: коли голосові зв'язки знаходяться на маленькій відстані один від одного і коливаються один щодо одного в мові, кажуть, що звук – вокалізований. Якщо ж зв'язки не коливаються, то кажуть, що звук – невокалізований.

М'яке піднебіння: працює як заслінка, яка відкриває прохід повітря в носову порожнину. Тверде піднебіння: довга, відносно тверда поверхня верхньої стінки ротової порожнини, у поєднанні з язиком дозволяє вимовляти приголосні звуки.

Мова: гнучкий артикулятор. При віддаленні від піднебіння дозволяє вимовляти голосні звуки, при наближенні до піднебіння – приголосні. Зуби: у поєднанні з мовою використовуються при вимові деяких приголосних звуків. Губи: можуть округлятися або розтягуватися, змінюючи звучання голосних звуків, або змикатися для зупинки повітряного потоку при вимові деяких приголосних звуків.

Основною відмінністю між звуками є їхнє розмежування на вокалізовані та невокалізовані звуки.

Вокалізовані звуки у своїй частотній та часовій структурі мають квазіперіодичну складову. Вона вноситься, коли при вимові звуку беруть участь голосові зв'язки, що вібрують із різною частотою (від 60 Гц у дорослого чоловіка до 300 Гц або вище у дівчини чи дитини). Частота вібрації голосових зв'язок називається основною частотою звуку, так як вона є базовою частотою для інших високочастотних гармонік, створюваних у гортанній та ротовій порожнині. Також, основна частота більша, ніж будь-який інший фактор впливає на основний тон мови. [38]

На рисунку 1.4 зображено етапи циклу стану голосових зв'язок людини при проходженні через них повітряного потоку. На стадії (а), голосова щілина

зімкнута, повітряний потік зупиняється перед голосовими зв'язками.

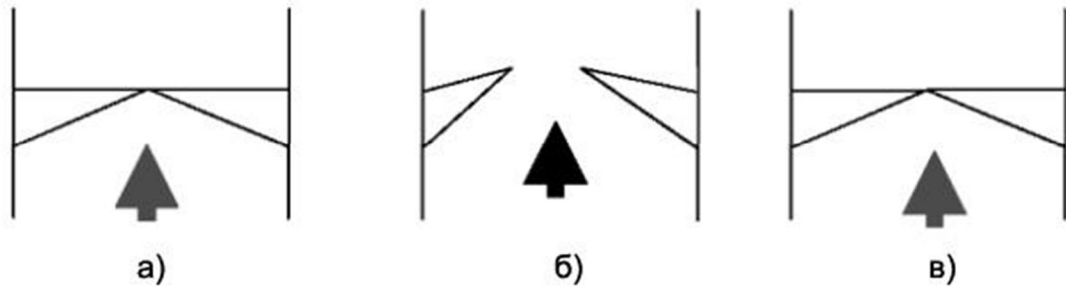


Рис. 1.4. Цикл скорочення голосових зв'язок. а) перекриття голосової щілини, наростання тиску повітряного потоку; б) відкриття щілини під впливом тиску; в) закриття голосової щілини за рахунок вирівнювання тиску та еластичності тканин.

У якийсь момент (стадія б) тиск повітря перед зв'язками долає бар'єр, і повітря виривається назовні через голосову щілину.

Тим не менш, тканини та м'язи голосових зв'язок, завдяки природній еластичності, повертаються у вихідний стан, закриваючи голосову щілину (стадія в). Таким чином, створюється послідовність звукових коливань, що є джерелом енергії всім вокалізованих звуків.

При вимові невокалізованих звуків голосові зв'язки або розслаблені або сильно напружені, внаслідок чого не виробляють звукових коливань. Повітря вільно проходить з легенів у ротову та/або носову порожнину мовного тракту. Внаслідок взаємодії повітря з різними артикуляторами відбувається перетворення повітряного потоку, що призводить до вимови того чи іншого звуку. [38]

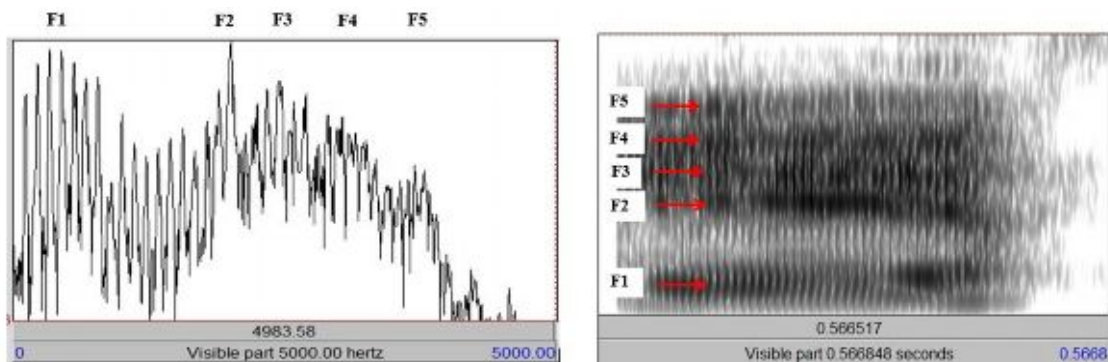


Рис. 1.5. Сигналограма (ліворуч) та спектрограма (праворуч) мовного сигналу, що відповідає вимовленому звуку “А”

Таким чином, мовний сигнал є періодичний процес, що складається з основної частоти F_0 і деякої кількості гармонік. Окремо варто відзначити посилені гармоніки - форманти (F_1, F_2, F_3, F_4, F_5). Форманта являє собою певну частотну область (рисунок 1.5), в якій внаслідок резонансу посилюється деяка кількість гармонік тону, що виробляється голосовими зв'язками, тобто в спектрі звуку форманта є областю посилених частот, що досить чітко виділяється. [40] Практично парадокс форманти є прояв роботи активного смугового фільтра у складі мовного тракту. З вище переліченого, мовний сигнал є природним зображати як суму синусоїдальних коливань.

1.2.2. Сприйняття мовного сигналу людиною

У системі сприйняття мови є дві основні складові частини: зовнішні слухові органи та слуховий відділ мозку. Вухо обробляє сигнал, який несе в собі звукова хвиля шляхом перетворення його в механічну вібрацію барабанної перетинки і подальшим відображенням цієї вібрації в послідовність імпульсів, що передаються слуховим нервом. Корисна інформація витягується у різних ділянках слухового відділу мозку людини. Рисунок 1.6 зображує анатомію вуха людини. [40]

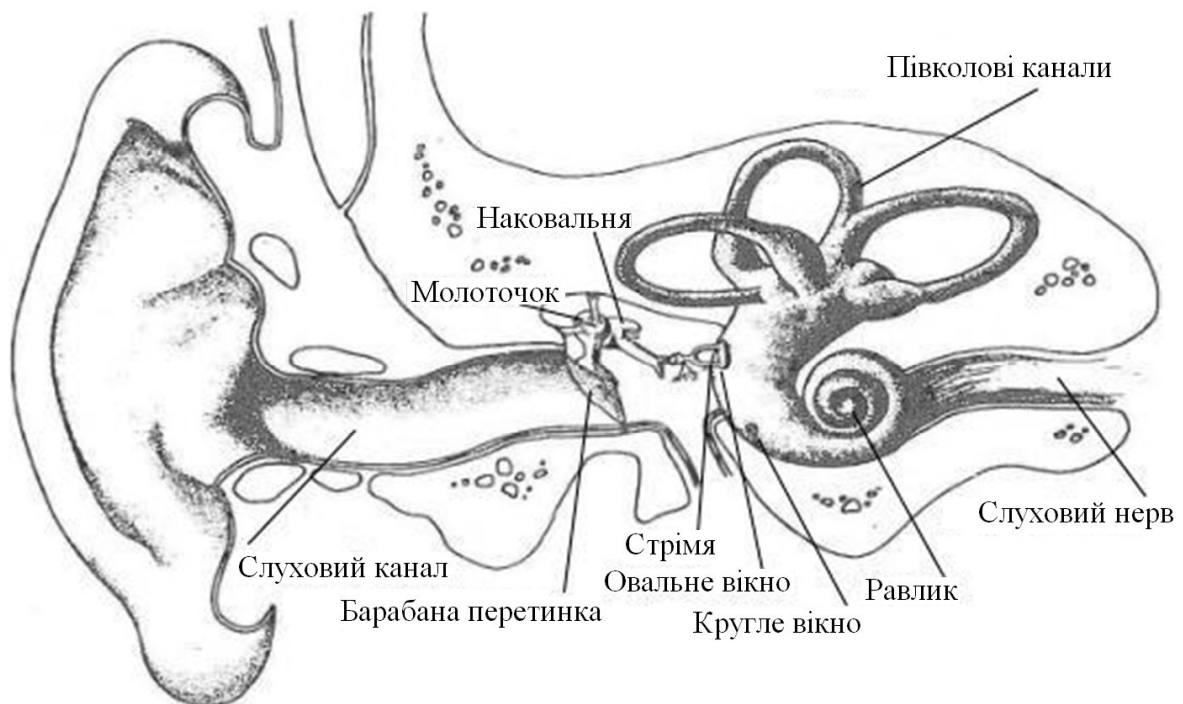


Рис. 1.6. Будова групи слухових органів [40].

Структура внутрішнього вуха влаштована так, що за різних частот початкового сигналу, максимальна амплітуда зміни тиску рідини в равлику буде реєструватися на певній відстані від її основи. Таким чином, равлика можна уявити, як гребінець фільтрів, вихідний сигнал якого впорядкований на відстані від підстави равлика. Фільтри, ближчі до основи равлики відповідають більш високі частоти. Таким чином, равлик є аналізатором спектра. [3]

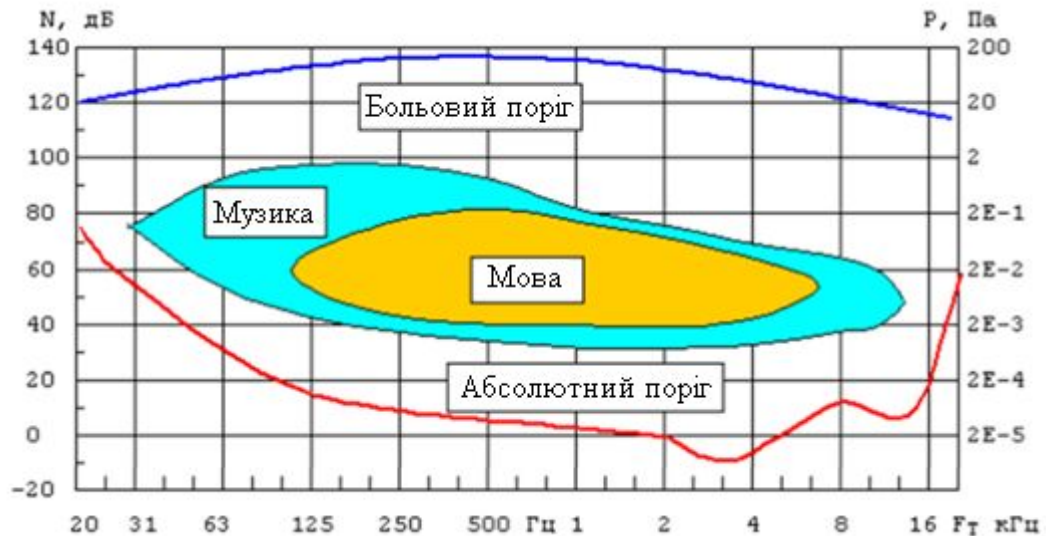


Рис. 1.7. Графік сприйняття звуку людиною [3]

Слуховий нерв є набір частотних каналів. До кожного частотного каналу входить група нейронів, з'єднаних з одним або сусідніми фільтрами равлики, тобто ті, які мають однакові або близькі характеристичні частоти. Цей набір ознак подається як миттєве зображення сигналу в мозок людини, в якому, за допомогою складної нейронної мережі, відбувається виділення корисної інформації з отриманого сигналу. На жаль, точних даних про те, як ця інформація отримується всередині людського мозку, немає. Є лише ряд теорій, які по-різному описують можливі нейронні структури всередині мозку та їхню взаємодію.

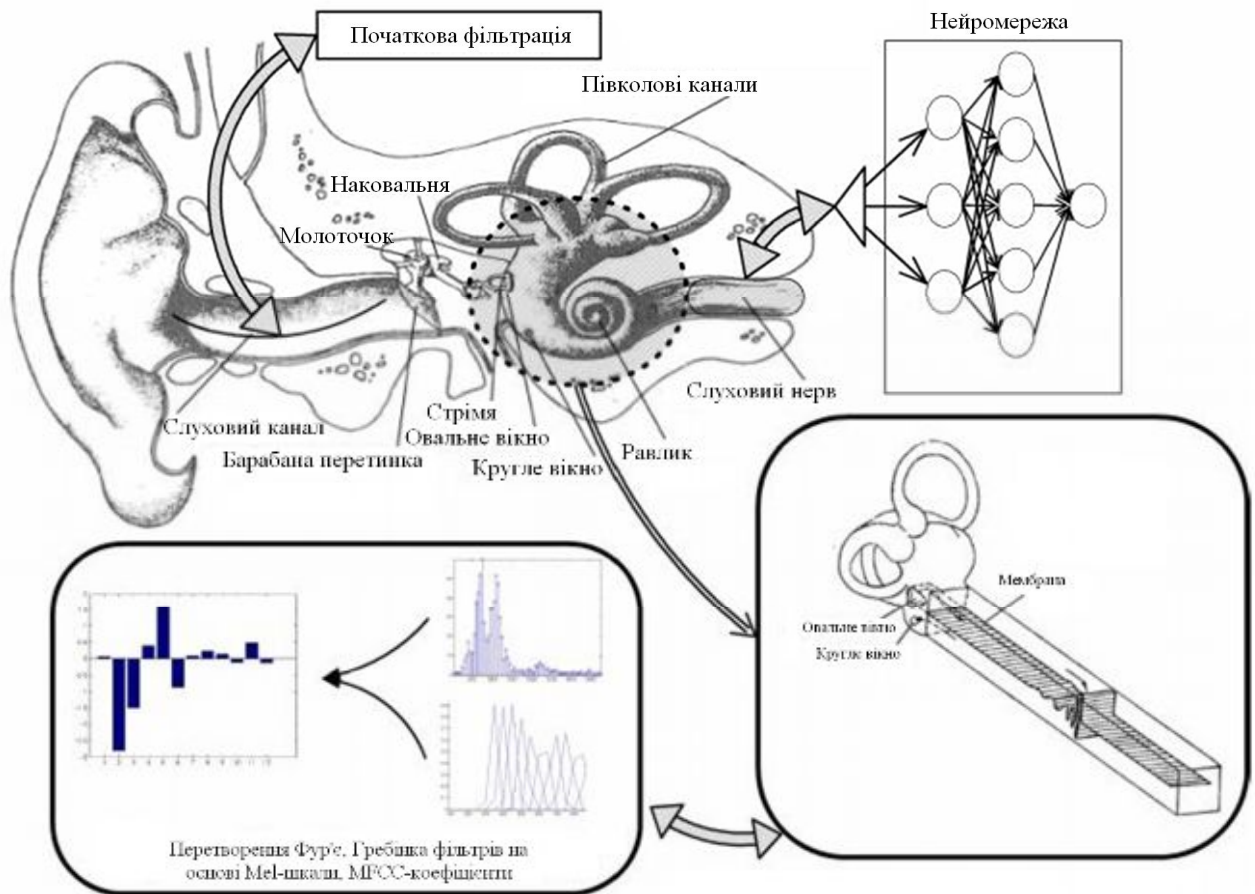


Рис. 1.8. Гіпотетична схема розпізнавання мови людиною

На рисунку 1.8 представлено спрощену, гіпотетичну схему розпізнавання мови людиною. Багато елементів різних систем розпізнавання мови ґрунтуються на слуховому тракті людини та намагаються імітувати механізми його роботи. Так, одна з популярних на сьогоднішній день характеристична ознака мовного сигналу (Мел-кепстральні коефіцієнти MFCC) заснована на вивченні методів перетворення сигналу у внутрішньому вусі людини. Також, розробка та розвиток неймережевих алгоритмів пов'язані з дослідженнями мозку людини. [23,24, 50]

Вченими були проведені дослідження, для визначення градації частот, яка б моделювала природну реакцію людської системи сприйняття мови, в якій равлик діє як спектральний аналізатор [40]. Складний механізм внутрішнього вуха та слухового нерва передбачає, що властивості отримання звуків на змінних частотах не можуть бути, очевидно, простими чи лінійними. Широко відомо, що музичний тон поділяється на октави та півтони.

Частота f_1 вище частоти f_2 на октаву і тоді, коли $f_1=2f_2$. У 1 октаві 12 півтонів, отже, f_1 вище частоти f_2 на півтон тоді і лише тоді, коли виконується рівність (1.6)

$$f_1 = 2^{(1/12)} f_2 \quad (1.6)$$

В результаті різних досліджень [52], що ґрунтуються на людських відчуттях звуків різних частот, було виведено ряд шкал, які дозволяли уявити частоту звуку в ближчих до людського сприйняття величинах.

При розпізнаванні людської мови набула поширення - мел-шкала [52], лінійна при частотах нижче 1кГц і логарифмічна при частотах вище 1кГц. Мел-шкала була отримана в результаті експериментів із взірцевими тонами (синусоїдами) у яких з піддослідних потрібно розділити дані діапазони частот на 4 рівних інтервалу або налаштувати частоту потрібного тону так, щоб він був в половину частоти вихідного. 1 крейда визначається як 1 тисячна рівня тону в 1 кГц. Як і будь-яких інших спробах створити подібні шкали, розраховується, що шкала крейда більш точно моделює чутливість людського вуха. Обчислення мел-значень можна приблизно подати такою формулою (1.9):

$$m(f) = 2595 \cdot \log_{10} (1 + f/700), \quad (1.7)$$

Зворотне перетворення можна здійснити за допомогою формули:

$$f(m) = 700 \cdot \left(10^{\frac{m}{2595}} - 1 \right), \quad (1.8)$$

де f – частота звуку у Гц, m – висота звуку.

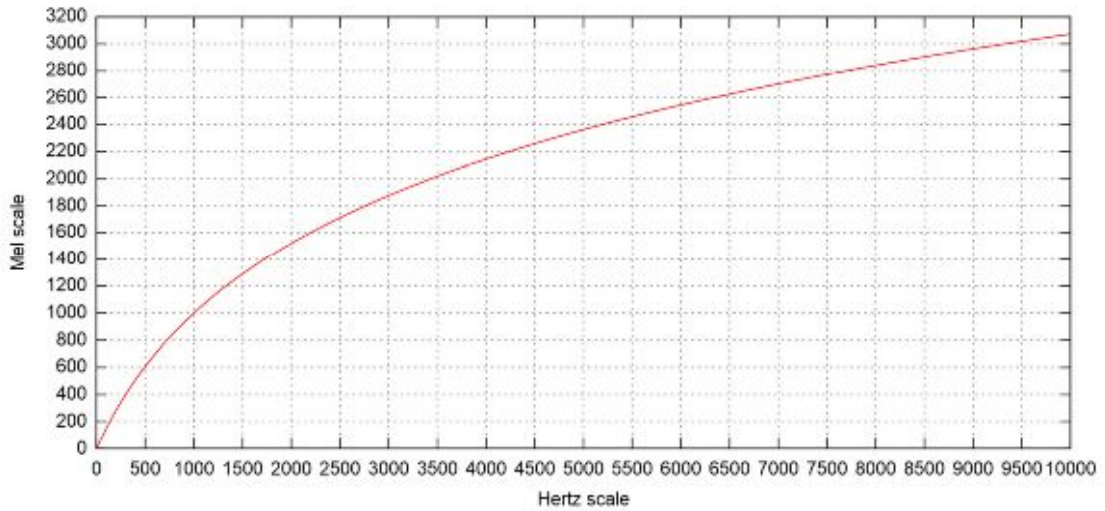


Рис. 1.9. Шкала Герц/Мел

Ряд сучасних технік обробки мовного сигналу ґрунтується на застосуванні таких шкал. На рисунку 1.9 представлена шкала відповідності одиниць частоти звуку крейда одиницям Герц

Доцільно відзначити деякі основні параметри, властиві мовним сигналам, такі як основний тон, висота тону та тембр звуку.

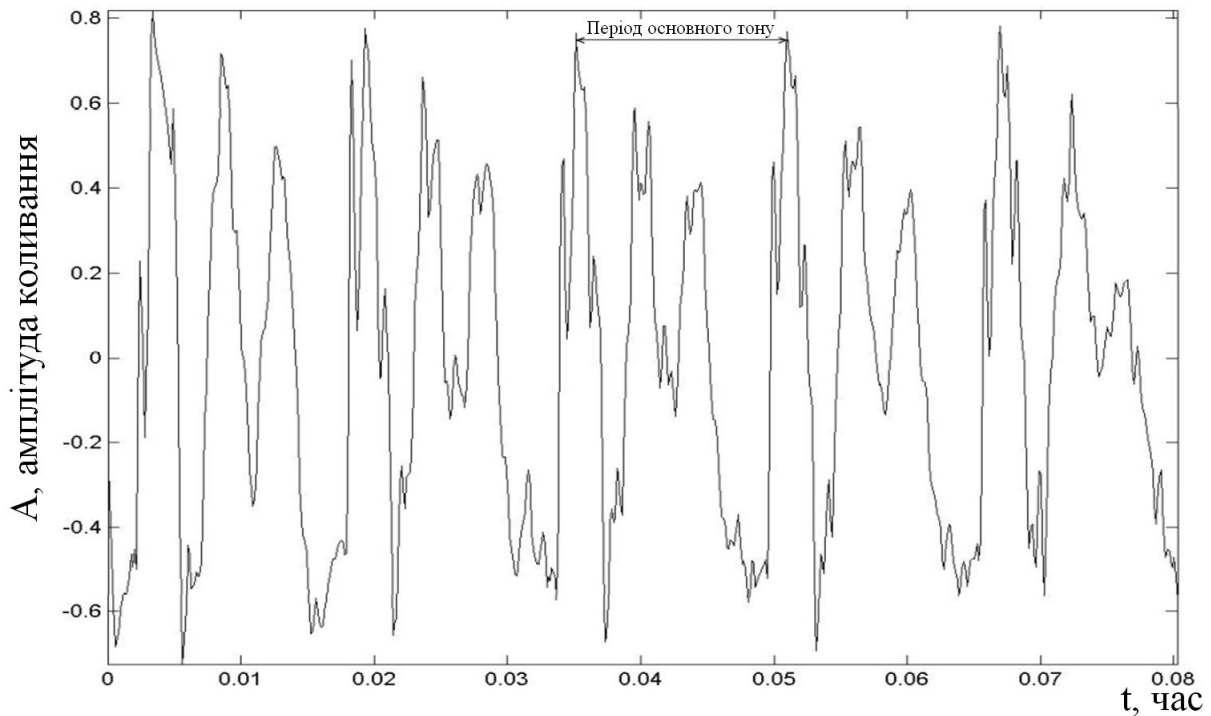


Рис. 1.10. Сигналограма фонемі, що відповідає звуку "А" з виділеним періодом основного тону

У спектрі мовлення найчастіше присутній частотна складова (відповідає найменшій частоті і найбільшому періоду в даному спектрі), що прийнято називати основним тоном (ВІД). На рисунку 1.10 представлений період основного тону фонем, що відповідає звуку "А".

Висота звуку – характеристика, що умовно розподіляє звуки за деякою шкалою від низьких до високих. На сприймається висоту звуку впливає частота основного тону, в той же час форма і період звукової хвилі можуть впливати на висоту звуку. Висота звуку може бути визначена слуховий системою людини і в складних сигналів, але у разі періодичності сигналу. [40]

Залежно від співвідношення амплітуд, частотних складових сигналу звук може набувати різного забарвлення і сприймається як тон або шум. Якщо спектр сигналу має яскраво виражені піки або один пік, то ця ділянка є тоном, якщо спектр звуку є суцільним, то ця ділянка є шумовим. [40]

1.3. Методи цифрової обробки сигналів у задачах розпізнавання мовних сигналів

Системи розпізнавання мови або вимови, на сьогодні, базуються на отриманні інформації, яка необхідна для розпізнавання мови. Про те, в даний час неможливо здійснити нормальне перетворення мови в лінгвістичні символи. Початковою стадією розпізнавання мови є первісне трансформування мовного коливання, що надходить на мікрофони, так що його можна піддати комп'ютерного аналізу, тобто. дискретизація та квантування. Наступним етапом є аналіз мовного сигналу, який може бути виконаний як у базисі Фур'є, так і за допомогою вейвлетів або шляхом обчислення кепстру, що не тільки дає змогу стиснути інформацію, а ще й має можливість сконцентруватися на аспектах мовного сигналу.

1.3.1 Спектральний аналіз.

Для аналізу оцифрованих сигналів використовують найчастіше спектральний аналіз у базисі Фур'є [33]. Даний метод обробки дозволяє характеризувати частотну складову аналізованого сигналу. Для реалізації

частотного аналізу можна скористатися дискретним прямим (1.11) та зворотним (1.12) перетвореннями Фур'є. Для одномірного масиву мовних даних $x(i)$, де $i=1,2,\dots,N$, коефіцієнти ряду Фур'є визначаються таким чином:

$$X(k) = \sum_{i=1}^N x(i) e^{-j \frac{2\pi}{N} (i-1)(k-1)} \quad (1.9)$$

де $k=1,2,\dots,N$, j - комплексна одиниця.

Зворотне перетворення на часову область:

$$x(i) = \frac{1}{N} \sum_{k=1}^N X(k) e^{j \frac{2\pi}{N} (i-1)(k-1)} \quad (1.10)$$

Вагова частина трансформантів Фур'є (косинусна) визначається так:

$\text{Re}(X(k) = X \cos(k)$, уявна (синусна) так: $\text{Im}(X(k) = X \sin(k)$.

Потім визначаються такі значення:

Амплітуда

$$A(k) = \sqrt{\text{Re}(X(k))^2 + \text{Im}(X(k))^2} \quad (1.11)$$

Фаза

$$\Psi(k) = \text{arctg} \left(\frac{\text{Im}(X(k))}{\text{Re}(X(k))} \right) \quad (1.12)$$

Енергія

$$P(k) = \text{Re}(X(k))^2 + \text{Im}(X(k))^2 \quad (1.13)$$

Оскільки змінна k має сенс частоти, можна аналізувати частотні характеристики сигналу досліджуючи функції $X(k)$ і $\Psi(k)$.

Основний недолік перетворення Фур'є полягає у відсутності можливості відрізнити стаціонарний сигнал від нестаціонарного сигналу, з тим же спектральним складом, що проілюстровано на рисунку 1.11. Дане явище накладає ряд обмежень застосування перетворення Фур'є при аналізі сигналів.

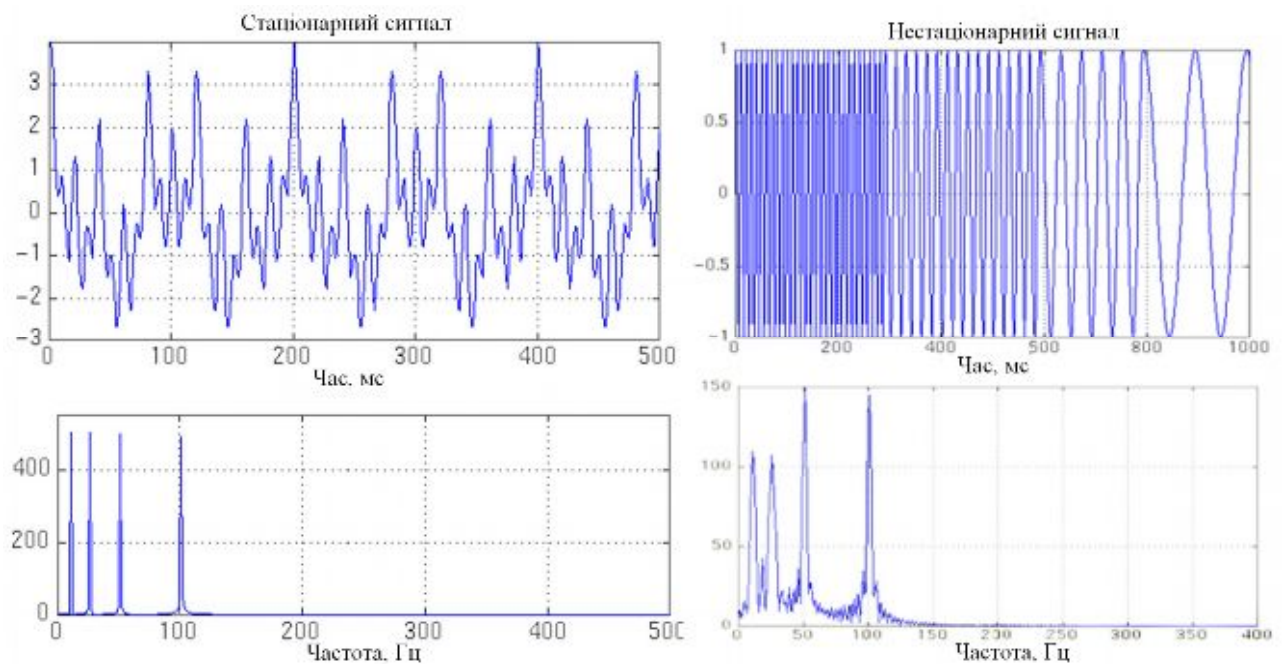


Рис. 1.11. Демонстрація перетворення Фур'є для стаціонарного та нестаціонарного сигналів

1.3.2 Віконний аналіз у базисі Фур'є

Оскільки перетворення Фур'є дає інформацію лише про частоту, що є у сигналі і дає жодної інформації у тому, який проміжок часу ця частота є у сигналі. Отже, для аналізу кусково-стаціонарного сигналу, яким і є мовний сигнал, необхідно використовувати вікно, досить вузьке для того, щоб сигнал усередині нього виглядав стаціонарним, тобто для мовного сигналу тривалістю щонайменше період основного тону. Таке подання сигналу отримав назву віконного.

1.3.3 Вейвлет аналіз.

Вейвлет-перетворення (ВП) [10,29] забезпечує частотно-часове представлення сигналів. Сигнал пропускається через два фільтри –

низькочастотний та високочастотний, і процедура повторюється. ВП в порівнянні з віконним перетворенням Фур'є, яке має постійний масштаб в абиякий момент часу для всіх частот, має найкраще уявлення часу та найгірше уявлення частоти на низьких частотах сигналу та найкраще уявлення частоти з найгіршим уявленням часу на високих частотах сигналу . [29]

На рисунку 1.12 добре видно, що отримане вейвлет-перетворення є більш деталізованим за часом у сфері високих значень масштабу (низьких частот) і менш деталізовано області низьких значень масштабу (високих частот).

З цього випливає, що ВП дозволяє зменшити вплив принципу невизначеності на отриманому частотно-часовому поданні сигналу. З його допомогою низькі частоти мають більш детальне уявлення щодо часу, а високі щодо частоти. [29]

Неможливо одночасно забезпечити гарний дозвіл за часом і частотою. Чим вужче вікно, тим вище роздільна здатність за часом і нижче роздільна здатність за частотою. На відміну від постійного дозволу по осях у перетворенні Фур'є (рис. 1.12).

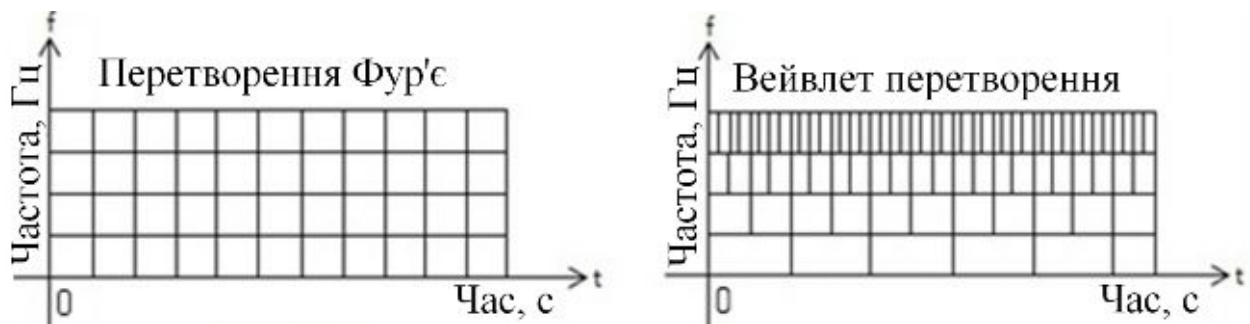


Рис. 1.12. Порівняння графіків перетворенням Фур'є із Вейвлет-перетворенням

З рисунку 1.12 видно що змінюються тільки часові інтервали та частотні смуги.

ВП є альтернативою перетворенню Фур'є у тих випадках, коли сигнал не носить періодичного характеру. Розрізняють безперервне та дискретне ВП. Передбачається, що це інтеграли, розглянуті нижче, існують.

Існує можливість аналізу сигналу з допомогою альтернативного підходу, ім'я якому – кратномасштабний аналіз (КМА). КМА дозволяє отримати

хороший за часом (поганий за частотою) на високих частотах і хороший дозвіл за частотою (поганий за часом) на низьких частотах. [29]

Витоки Дискретного Вейвлет-перетворення (ДВП) сягають 1976 року, коли Кройцер, Естебан і Галанд розробили метод декомпозиції дискретних сигналів. Кройхер, Веббер і Фланган того ж року опублікували аналогічну роботу з кодування мовних сигналів. Вони назвали свій метод аналізу субсмуговим кодуванням. Основна ідея - та сама, що і при НВП. Масштабно-часове уявлення сигналу виходить з використанням методів цифрової фільтрації. У дискретному випадку для аналізу сигналу різних масштабах використовуються фільтри з різними частотами зрізу. Сигнал пропускається через деревоподібно з'єднані ВЧ та НЧ фільтри [29].

1.3.4 Кепстральний аналіз.

Основою кепстрального аналізу мовних сигналів є припущення, що мовний сигнал є сигналом на виході лінійної системи з параметрами, що поступово змінюються.

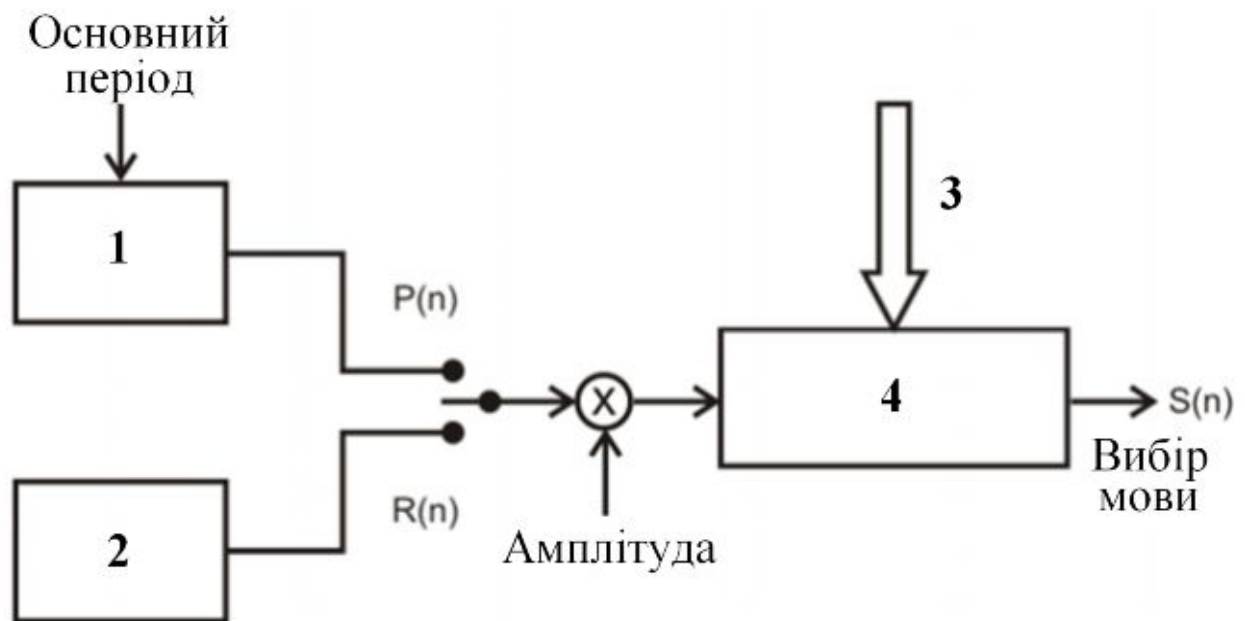


Рис. 1.13. Модель мовного апарату як лінійної системи: 1 – генератор імпульсної послідовності, 2 – генератор ПСП, 3 – коефіцієнти цифрового фільтра (параметри голосового тракту), 4 – нестационарний цифровий фільтр.

Таким чином, на коротких відрізках мовний сигнал можна розглядати як

сигнал на виході лінійної системи з постійними параметрами, що збуджується послідовністю імпульсів або випадковим шумом. Якщо джерела збудження та форма голосового тракту щодо незалежні мовний апарат можна подати у вигляді наступної сукупності елементів, зображених на рисунку 1.13.

Таким чином, проблема аналізу сигналу зводиться до вимірювання параметрів моделі та оцінки зміни параметрів з часом. Відомо, що імпульсна характеристика фільтра та сигнал збудження взаємодіють через операцію згортки. Аналіз мови може розсмартуватися як завдання поділу компонент, що у операції згортки, тобто. задача зворотного згортки [31,33]. Один із способів вирішення поставленого завдання є кепстральний аналіз.

У представленій моделі фільтри мають постійні характеристики на часовому інтервалі ~10-20 мс. Тому на кожному інтервалі фільтр можна характеризувати імпульсною або частотною характеристикою або набором коефіцієнтів. Така модель дозволяє використовувати для аналізу мовних сигналів гомоморфну розгортку. [33].

Нехай буде заданий сигнал $S_{вих}(t)$ на виході фільтра. Необхідно отримати інформацію про вхідний сигнал $S_{вх}(t)$ і сам фільтр, шляхом визначення його імпульсної характеристики $h(t)$.

Вихідний сигнал визначається згорткою:

$$S_{вих}(t) = S_{вх}(t) h(t) \quad (1.14)$$

Так як $S_{вих}(t) = S_{вх}(t)H(t)$ у частотній області, то прологарифмувавши отримуємо вираз:

$$\text{Ln}[S_{вих}^2(w)] = \text{Ln}[S_{вх}^2(w)] + \text{Ln}[H^2(w)] \quad (1.15)$$

Застосуємо щодо нього зворотне перетворення Фур'є, можна отримати вираз виду:

$$C(q) = C_s(q) + C_h(q)$$

з якого методами лінійної фільтрації можна виділити деякі характеристики $S_{ex}(t)$ і $h(t)$.

Також $C(q)$ може бути представлено у вигляді:

$$C(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln[S(w)]^2 e^{iwsq} dw \quad (1.17)$$

Дане перетворення прийнято називати "кепстр", дана назва отримана шляхом перестановки букв у слові "спектр". Аргумент q має розмірність часу, даний час є певною мірою особливим, так як $C(q)$ в будь-який момент q залежить від функції $s(t)$ вихідного сигналу зі спектром $S(w)$ заданою при $-\infty < t < \infty$. Іноді q називаються "сачтота" або "кьюфренсі" (анаграми від українського "частота" та англійського "фрекьюнсі")

Зважаючи на те, що в комп'ютерних та телекомунікаційних системах працюють з дискретним поданням мовного сигналу, доцільно навести запис кепстру в дискретній формі:

$$C(n) = \frac{1}{N} \sum_{k=0}^{N-1} \ln|X(k)|^2 e^{i\frac{2\pi}{N}kn}, 0 \leq n \leq N-1 \quad (1.18)$$

Однією з основних складових, що дозволяє ефективно вирішувати завдання розпізнавання та ідентифікації відрізків мовних сигналів, є вибір простору ознак, що найбільше відображає необхідні для розпізнавання відмінні властивості мовного сигналу. При цьому важливо зберігати прийнятний обсяг використовуваних ознак, щоб не підвищувати обчислювальну складність алгоритмів вище за необхідне значення.

1.4. Субсмуговий підхід до обробки мовних сигналів

Дослідження особливостей розподілу енергії за частотними інтервалами відрізків сигналів, відповідних звукам російської мови, проведених в [14,16,17], показали, що більша частина формант вокалізованої російської мови зосереджена в досить вузькому частотному діапазоні, а енергія шумових звуків розподілена по кількох інтервалах. високочастотної частини частотної осі

Оцифрований звуковий сигнал представлений у вигляді:

$$\vec{x} = (x_1, x_2 \dots x_{(m-1)}, x_m) \quad (1.19)$$

x – мовний сигнал, m – кількість відліків.

Далі необхідно вибрати довжину вікна аналізу - N і кількість частотних інтервалів - R , виходячи з необхідної роздільної здатності.

Мовний сигнал x піддається розподілу на вікна тривалістю N :

$$\vec{x}_n = (x_1, x_2 \dots x_{(n-1)}, x_n), N = 1 \dots n \quad (1.20)$$

Обчислення розподілу часток енергій мовного сигналу відбувається за такою формулою [14]:

$$P_r = \vec{x}_r A_r \vec{x}_r^T \quad (1.21)$$

де матриця $A_r = \{a_{ik}\}$ з елементами виду:

$$a_{ik} = \frac{\sin[v_{r+1}(i-k)] - \sin[v_r(i-k)]}{\pi(i-k)} \quad (1.22)$$

де $v = [0, \Delta, \Delta + 2\Delta, \dots, (n\Delta - 4\Delta) + 2\Delta, n\Delta]$ вектор, що задає межі частотних інтервалів (рисунок 1.16), ширина яких повинна задовольняти вираз:

$$\Delta = \frac{\pi}{2R + 1} \quad (1.23)$$



Рис. 1.16. Частотна вісь у субсмуговому аналізі, розбита на R інтервалів

Основними перевагами субсмугового підходу є:

- Забезпечує мінімальне просочування енергії через межі частотних інтервалів
- Забезпечує розрахунок розподілу часток енергій без переходу в частотну область
- Відсутність проблеми частотно-часового вирішення.

1.5. Висновки до розділу 1

У розділі описано сучасний стан напряму розпізнавання мовних сигналів, зокрема: особливості мовлення та сприйняття мови людиною, схеми реалізації системи розпізнавання мовлення, математичні моделі мовних сигналів, методи цифрової обробки сигналів у задачах розпізнавання мовних сигналів та методи аналізу мовних сигналів.

РОЗДІЛ 2

ОСНОВНА ЧАСТИНА

2.1. Акустико-фонетичний підхід до розпізнавання мовних сигналів

Будь-яку ділянку РС можна, певним чином, характеризувати через певний ознаку, що відповідає даному відрізку, або набір таких ознак. Описані в першому розділі методи цифрової обробки РС дозволяють певною мірою відбивати природу впливу, що породжує. Проте, якщо підходити з позиції, що мовний сигнал є лише окремою реалізацією переданої людиною інформації, то неможливо характеризувати весь можливий набір таких реалізацій лише одним ознакою. [25] Мовний сигнал залежить від безлічі факторів, як зовнішніх (шуми, навколишня обстановка і т.п.), так і залежати від джерела інформації - людини (темп мови, емоційне відображення, тембр, та ін.), що також говорити про те, що необхідно підібрати такий набір ознак, який дозволить виділяти тотожні (в інформаційному сенсі) реалізації мовного потоку [12].

Мовний сигнал, що є акустичну реалізацію переданої людиною інформації, містить у собі деяке повідомлення. Передане однією людиною повідомлення, наприклад, пропозиція "Я йду в університет" при акустичній реалізації може набувати різних форм, які можуть відрізнятися за безліччю параметрів: гучність, тривалість, емоційне відображення, висота тощо.

Одним з найпопулярніших підходів до розпізнавання мовних сигналів є акустико-фонетичний підхід, важлива схема якого представлена на рисунку 2.1 [50]. Мовний сигнал піддається процедурі виділення ознак, на яких формуються вектори ознак, які дозволяють класифікувати відрізки мовних сигналів.

Використання набору векторів ознак обумовлено необхідністю враховувати різні характеристики сигналу, які неможливо детектувати лише одним вектором ознак. Таким чином, для отримання найбільш повного опису мовного сигналу в кожний момент години, природно використовувати

сукупність ознак, які дозволяють здійснювати точну ідентифікацію відрізків.

Акустично - фонетичний підхід

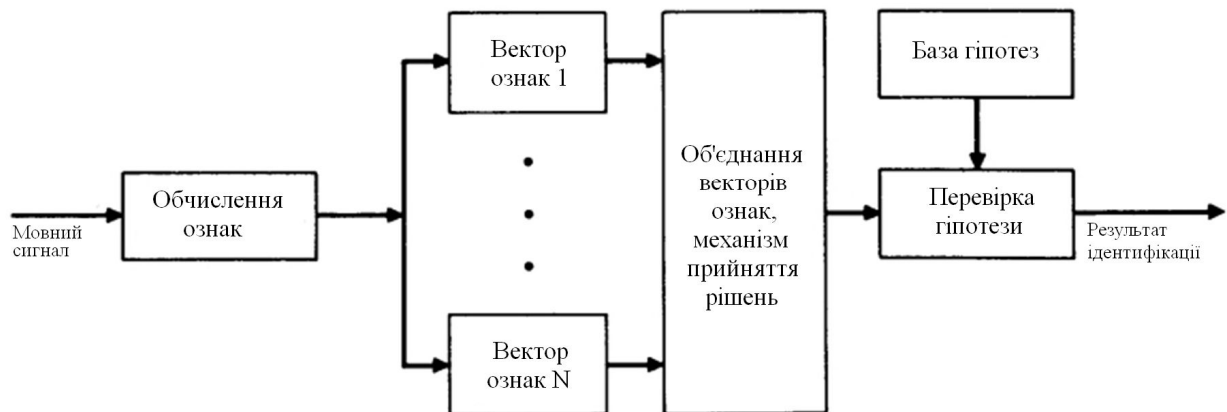


Рис. 2.1. Принципова схема ідентифікації відрізків мовних сигналів

Існує кілька підходів до отримання векторів ознак мовного сигналу. Всі вони задаються загальною метою зменшити надмірність сигналу та виділити найбільш релевантну інформацію, і, водночас, відкинути нерелевантну. Як правило, ознаки, що описують мовний сигнал з різних точок зору, комбінуються в один вектор ознак, на основі якого ідентифікується використанням обраної навченої моделі. Далі будуть представлені найпопулярніші ознаки, що виділяються з мовних сигналів. У цій роботі будуть досліджені такі вектори ознак:

- 1 – Декомпозиція сигналу банком фільтрів;
- 2- розподіл миттєвих енергій сигналу;
- 3- Розподіл часток енергії за частотними інтервалами;
- 4- розподіл інформаційних інтервалів;
- 5- Частота переходів через нуль;
- 6- Ширина частотної області, що займається сигналом;
- 7- Мел-кепстральні коефіцієнти.

Як було зазначено в першому розділі сам мовний сигнал у часовій області не має достатньої репрезентативністю для проведення розпізнавання через високу надмірність, отже, необхідно вибрати такий ознаковий простір, який би дозволив вирішувати поставлене завдання розпізнавання мовних сигналів із

заданою точністю.

Ознаки МС, відповідних різним фонемам, мають відмінності у вигляді відмінностей у природі впливів, що породжують їх. Необхідно знайти такі відмінності в ознаках, які б забезпечили можливість об'єднання фонем, мають одну природу в один клас, а також максимально розділити фонемі відповідні різним звукам.

Будь-яку ділянку МС можна, певним чином, характеризувати через певну ознаку, що відповідає даному відрізу, або ж набір таких ознак. Описані у першому розділі методи цифрової обробки МС дозволяють певною мірою відбивати природу впливу, що породжує. Однак, з огляду на те, що мовний сигнал є лише окремою реалізацією переданої людиною інформації, то неможливо характеризувати весь можливий набір таких реалізацій лише однією ознакою. [25] Мовний сигнал залежить від безлічі факторів, як зовнішніх (шуми, навколишня обстановка і т.п.), так і залежать від джерела інформації - людини (темп мови, емоційне забарвлення, тембр, та ін.), що також говорить про те, що необхідно підібрати такий набір ознак, який дозволить виділяти тотожні (в інформаційному сенсі) реалізації з мовного потоку [12].

Мовний сигнал, що є акустичну реалізацію переданої людиною інформації, містить у собі деяке повідомлення. Передане однією і тією самою людиною повідомлення, наприклад, пропозиція "Я йду в університет" при акустичній реалізації може приймати різні форми, які можуть відрізнятися за безліччю параметрів: гучність, тривалість, емоційне забарвлення, висота тощо.

Одним з найпопулярніших підходів до розпізнавання мовних сигналів є акустично-фонетичний підхід, важлива схема якого представлена на рисунку 2.1 [50]. Мовний сигнал піддається процедурі виділення ознак, у яких формуються вектори ознак, які дозволяють класифікувати відрізки мовних сигналів. Використання набору векторів ознак обумовлено необхідністю враховувати різні характеристики сигналу, які неможливо детектувати лише одним вектором ознак. Таким чином, для отримання найбільш повного опису мовного сигналу в кожний момент часу, природно використовувати сукупність ознак, які дозволять здійснювати точну ідентифікацію відрізків.

Існує кілька підходів до отримання векторів ознак мовного сигналу. Всі вони задаються загальною метою зменшити надмірність сигналу та виділити найбільш релевантну інформацію, і, водночас, відкинути не релевантну. Як правило, ознаки, що описують мовний сигнал з різних точок зору, комбінуються в один вектор ознак, на основі якого ідентифікується використанням обраної навченої моделі. Далі будуть представлені найпопулярніші ознаки, що виділяються з мовних сигналів. У цій роботі будуть досліджені такі вектори ознак:

Як було зазначено в першому розділі сам мовний сигнал у часовій області не має достатньої репрезентативністю для проведення розпізнавання через високу надмірність, отже, необхідно вибрати такий ознаковий простір, який би дозволив вирішувати поставлене завдання розпізнавання мовних сигналів із заданою точністю.

Ознаки МС, відповідних різним фонемам, мають відмінності у вигляді відмінностей у природі впливів, що породжують їх. Необхідно знайти такі відмінності в ознаках, які б забезпечили можливість об'єднання фонем, мають одну природу в один клас, а також максимально розділити фонемі відповідні різним звукам.

2.2. Обчислювальні аспекти субсмугового аналізу мовних сигналів у задачах розпізнавання мовних сигналів

При субсмуговому аналізі [14] розподілу енергій відрізків МС область осі нормованих частот $[0, \pi]$ розбивається на ряд інтервалів межі, що не перекриваються, яких визначає вектор наступного виду: $v = [0, \Delta, \Delta + 2\Delta, \dots, (n\Delta - 4\Delta) + 2\Delta, n\Delta]$, де ширина нульового частотного інтервалу $[0, \Delta)$ визначається як: $\Delta = \frac{\pi}{2R+1}$; а R – кількість інтервалів, на які розбивається вісь частот, ця змінна знаходиться у прямій залежності від тривалості вікна аналізу N : $R \leq \left\lfloor \frac{N}{4} \right\rfloor$;

Наступна ширина частотних інтервалів дорівнює подвоєної ширини

першого інтервалу.

Значення тривалості вікна при аналізі МС доцільно [50] брати рівним такій кількості звітів дискретного сигналу, яка б відповідала 10 - 20 мс аналогового сигналу.

Видається доцільним взяти значення вікна аналізу N рівним 256, яке відповідає відрізку часу $t = 16$ мс.

Запишемо вектор, що задає межі інтервалів у такому вигляді:

$$v = [v_r, v_{r+1}) \cup [v_r, v_{r+1}); 0 \leq v_r < v_{r+1} \leq \pi; \quad (2.1)$$

де $r = 0 \dots R$;

З описаного у розділі 1.6 випливає, що розподіл часток енергії відрізка сигналу в заданому частотному інтервалі може бути оцінено на основі виразу [14]:

$$P_r(\vec{x}_N) = \vec{x}_N^T A_r \vec{x}_N, \quad (2.2)$$

де \vec{x}_N – аналізований відрізок сигналу; $A_r = \{a_{ik}^r\}$, $i, k = 1, \dots, N$ субсмугова матриця з елементами виду [14]:

$$a_{ik} = \frac{\sin[v_{r+1}(i-k)] - \sin[v_r(i-k)]}{\pi(i-k)} \quad (2.3)$$

Відповідно до вектора розбиття частотної смуги, елементи матриці, що відповідають нульовому частотному інтервалу, можуть бути представлені у вигляді:

$$a_{ik}^0 = \frac{\sin[v_{r+1}(i-k)] - \sin[v_r(i-k)]}{\pi(i-k)} = \frac{\sin[v_{r+1}(i-k)]}{\pi(i-k)} \quad (2.4)$$

Тоді для інших частотних інтервалів елементи матриці можуть бути представлені у вигляді:

$$a_{ik}^r = \frac{2 \cdot \sin[(v_{r+1} - v_r) \cdot (i-k)/2] \cdot \cos[(v_{r+1} - v_r) \cdot (i-k)/2]}{\pi(i-k)} \quad (2.5)$$

З урахуванням вектора розбиття (2.1) та виразу (2.4) вираз (2.5) може бути переписаний у вигляді:

$$a_{ik}^r = 2 \cdot a_{ik}^0 \cos(\omega_r(i-k)), \quad (2.6)$$

де $r = \frac{4\pi(r+1)}{N}$ - центральна частота інтервалу, $r = 0 \dots R$

Субсмугові матриці є симетричними та невід'ємно визначеними. Тому вони мають повний набір ортонормальних власних векторів і відповідних негативних власних чисел, для яких виконуються співвідношення:

$$\lambda_{kr} \vec{q}_{kr} = A_r \vec{q}_{kr}, \quad (2.7)$$

$$(\vec{q}_{kr}, \vec{q}_{ir}) = \sum_{j=1}^N q_{jk}^r q_{ji}^r = \begin{cases} 1, i = k \\ 0, i \neq k \end{cases} \quad (2.8)$$

Причому можна уявити субсмугову матрицю виразом:

$$A_r = \sum_{k=1}^N \lambda_{kr} \vec{q}_{kr} \vec{q}_{kr}^T = Q L Q^T \quad (2.9)$$

де $Q = \{\vec{q}_{kr}\}, k = 1, \dots, N$; матриця власних векторів, а $L = \text{diag}(\lambda_1^r, \lambda_2^r, \dots, \lambda_N^r)$ - діагональна матриця власних чисел. Підставляючи (2.9) у вираз (2.2):

$$\begin{aligned}
 P_r(\vec{x}_N) &= \vec{x}_N^T A_r \vec{x}_N = \vec{x}_N^T \cdot \sum_{k=1}^N \lambda_{kr} \vec{q}_{kr} \vec{q}_{kr}^T \cdot \vec{x}_N = \\
 &= \sum_{k=1}^N \lambda_{kr} \vec{x}_N^T \vec{q}_{kr} \vec{q}_{kr}^T \vec{x}_N = \sum_{k=1}^N \lambda_{kr} \alpha_{kr}^2,
 \end{aligned}
 \tag{2.10}$$

де α_{kr} – скалярні проєкції:

$$\alpha_{kr} = (\vec{q}_{kr}, \vec{x}_N)
 \tag{2.11}$$

З виразу (2.10) видно, що властивості власних чисел субсмугової матриці такі, що кількість нульових значень відносно мало, таким чином можна скоротити кількість обчислень, за допомогою виключення з процесу розрахунку близьких до нуля значень власних чисел. Для цього власні числа суболосної матриці повинні бути відсортовані за спаданням.

$$\lambda_1^r > \lambda_2^r > \dots > \lambda_N^r \geq 0.
 \tag{2.12}$$

І вибираємо J власних чисел з відсортованої множини матриці L_r і відповідні їм власні вектора g .

$$J = 2[N \cdot \alpha_{ii}^r / 2\pi] + 5;
 \tag{2.13}$$

З вибраних власних чисел та векторів формуємо матрицю наступного виду:

$$AA = \begin{pmatrix} \sqrt{L_1} Q_1^T \\ \sqrt{L_2} Q_2^T \\ \dots \\ \sqrt{L_R} Q_R^T \end{pmatrix}
 \tag{2.14}$$

де елементи, що відповідають приватному інтервалу, можуть бути представлені

$$AA = \sqrt{L_r} Q_r^T$$

У цьому операція вилучення квадратного кореня обумовлена необхідністю дотримати вираз (2.9), так як матриця AA використовується двічі: для отримання наближених значень часток енергії та відновлення потрібної складової сигналу.

Для отримання наближених значень частотних складових необхідно використовуючи власні вектори матриці A_r отримати вектор:

$$\vec{y} = AA \vec{x} \quad (2.15)$$

Щоб отримати частотну складову сигналу, що рекомендує тимчасовому відрізьку скористаємося формулою:

$$\vec{y}_r = AA_r^T \vec{y}_k \quad (2.16)$$

де $k = [r \cdot j + 1] \dots [r + 1] \cdot j$; $r = 0 \dots R$

Словесний опис алгоритму попереднього розрахунку матриці:

1. Запровадити значення тривалості вікна аналізу N ;
2. Розрахунок значень алгоритму R, Δ ;
3. Побудова вектора, що задає межі частотних інтервалів (2.1);
4. Сформувати матрицю $A_r = \{a_{ik}^r\}$, $i, k = 1, \dots, N$;
5. Розрахунок кількості власних чисел J , що зберігаються;
6. Визначити власні вектори \vec{q}_k^r з числа λ_N^r субсмугової матриці A_r ;
7. Сформувати матрицю власних векторів $Q = \{\vec{q}_k^r\}$, $k = 1, \dots, N$;
8. Сформувати матрицю $L = \text{diag}(\lambda_1^r, \lambda_2^r, \dots, \lambda_N^r)$ із власних чисел, відсортованих за спаданням (2.12);
9. Сформувати матрицю AA (2.14);

10. Зберегти матрицю AA для подальших розрахунків.

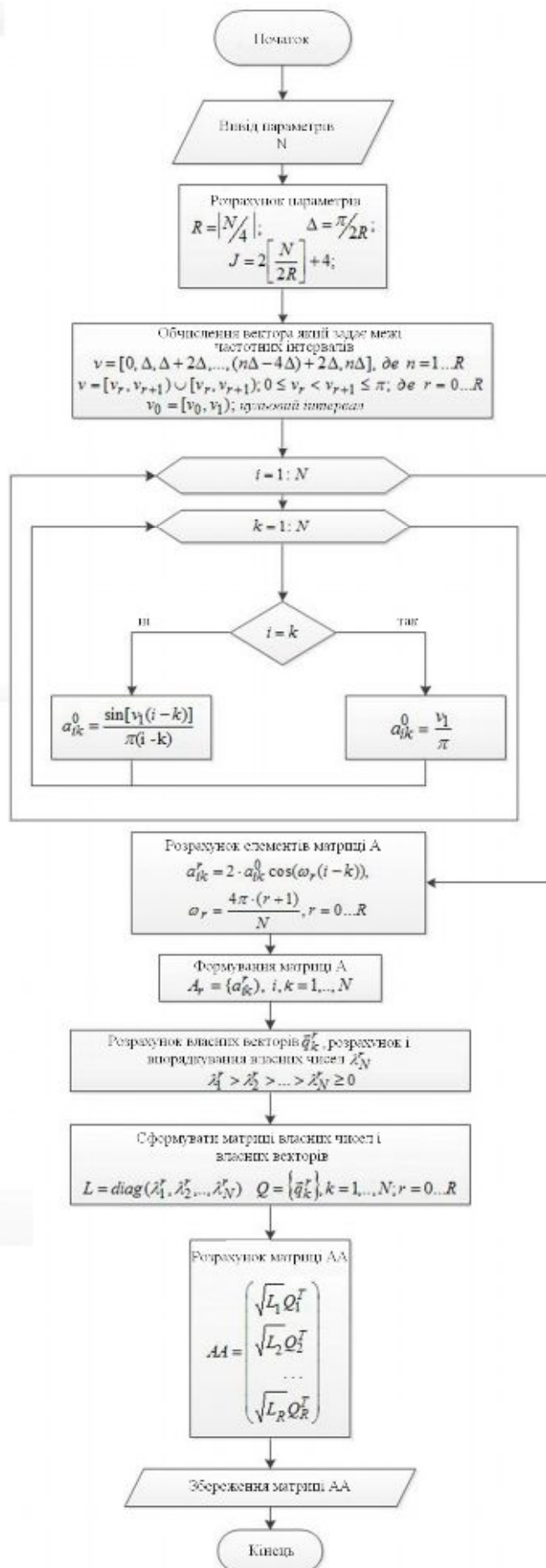


Рис. 2.2. Блок схема алгоритму обчислення матриці AA

Матриця AA надалі використовується як банк фільтрів, який дозволяє виділити необхідні ознаки, наприклад, розподіл енергії сигналу частотної осі. Попередній розрахунок матриці дозволяє суттєво знизити обчислювальне навантаження при обчисленні векторів ознак.

На рисунку 2.2 представлено блок-схему алгоритму попереднього розрахунку фільтруючої матриці AA . Надалі, дана матриця буде застосовуватися для аналізу мовного сигналу та виділення векторів ознак.

2.3. Дослідження просторів ознак у задачах розпізнавання мовних сигналів

2.3.1. Декомпозиція сигналу банком фільтрів.

Однією з ознак сигналу, може бути декомпозиція сигналу певну кількість частотних компонент. У цьому роботі декомпозиція сигналу виконуватиметься субполосним банком фільтрів, з допомогою розрахованої заздалегідь матриці AA .

Для отримання наближених значень частотних складових необхідно використовуючи власні вектори матриці A_r отримати вектор (2.15):

$$\vec{y}\vec{y} = AA \vec{x}$$

де $\vec{x} = (x_1 \dots x_N)$, мовний сигнал.

Щоб отримати потрібну частотну складову сигналу, що рекомендує тимчасовому відрізку скористаємося формулою (2.16):

$$\vec{y}_r = AA_r^T \vec{y}\vec{y}_k$$

де $k = [r \cdot j + 1] \dots [r + 1] \cdot j$; $r = 0 \dots R$.

Довжина вектора відповідає довжині аналізованого сигналу.

На рисунках 2.3-2.5 представлені відрізки сигналу, що відповідають

вимовленим словам “вісім” та “сім” та декомпозиції цих сигналів. Дані графіки дозволяють візуально оцінити рівень релевантності аналізованої ознаки.

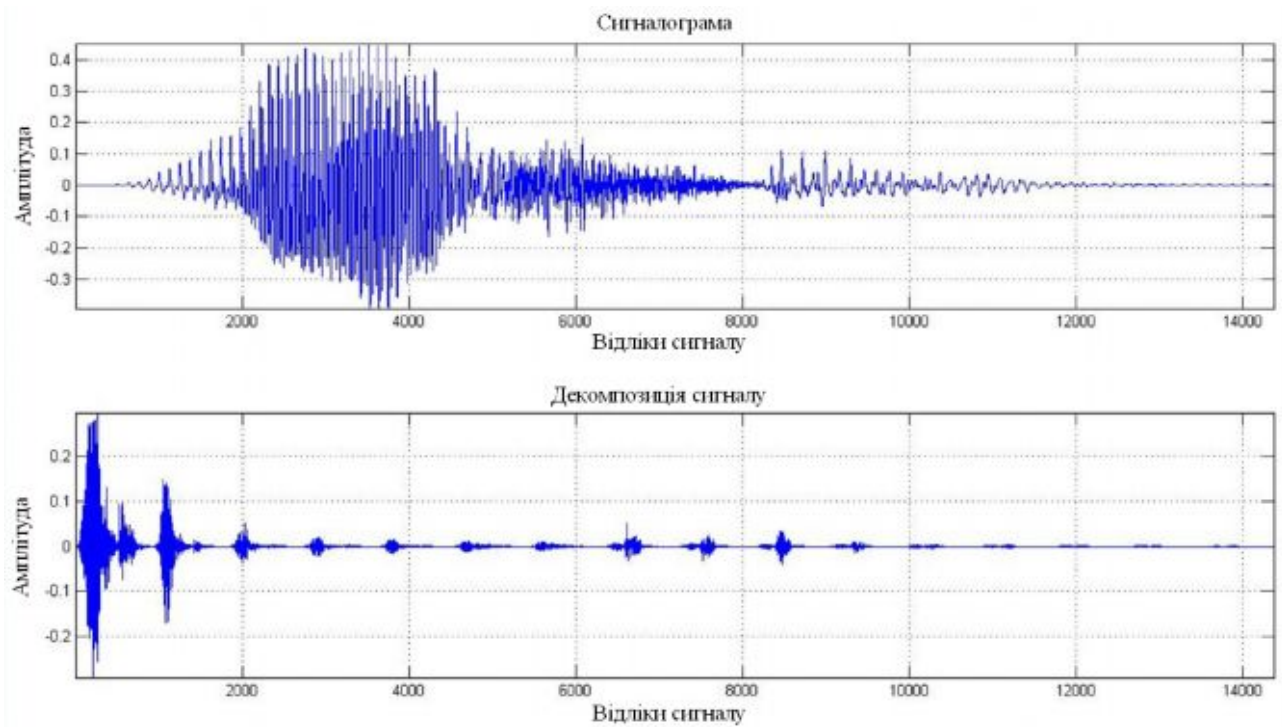


Рис. 2.3. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" та вектор ознак - декомпозиція сигналу банком фільтрів

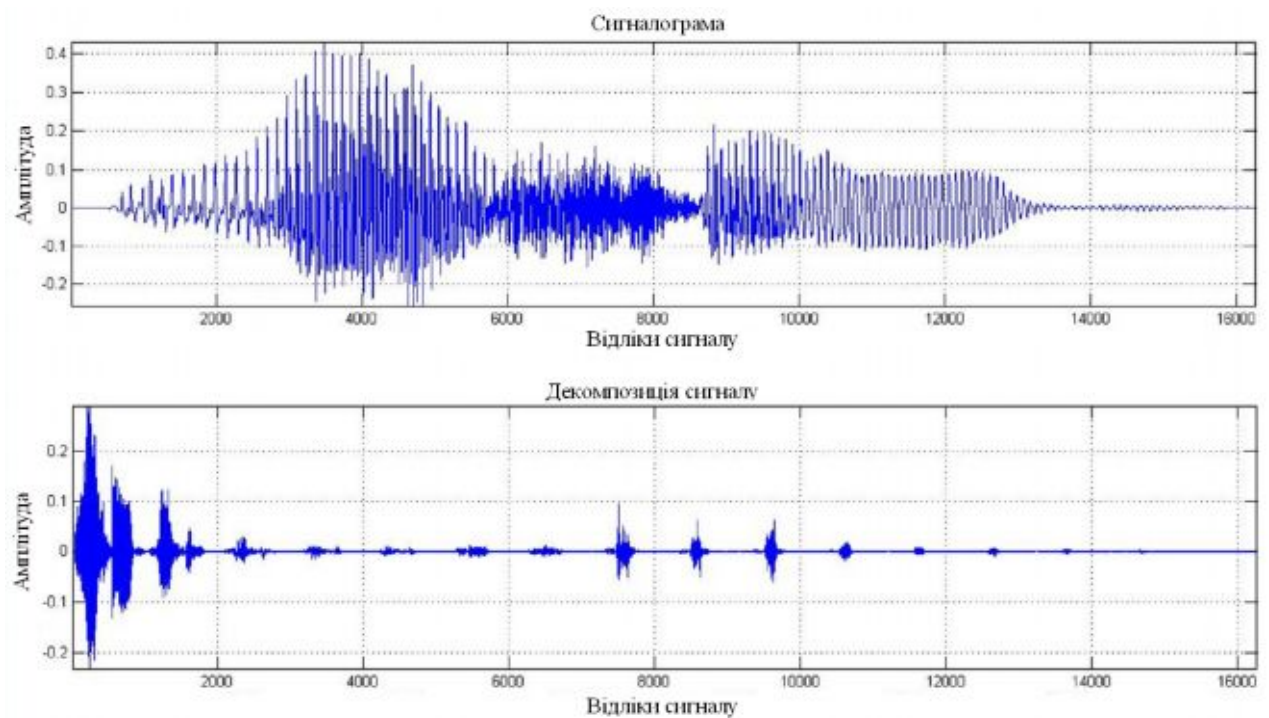


Рис. 2.4. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" та вектор ознак - декомпозиція сигналу банком

фільтрів

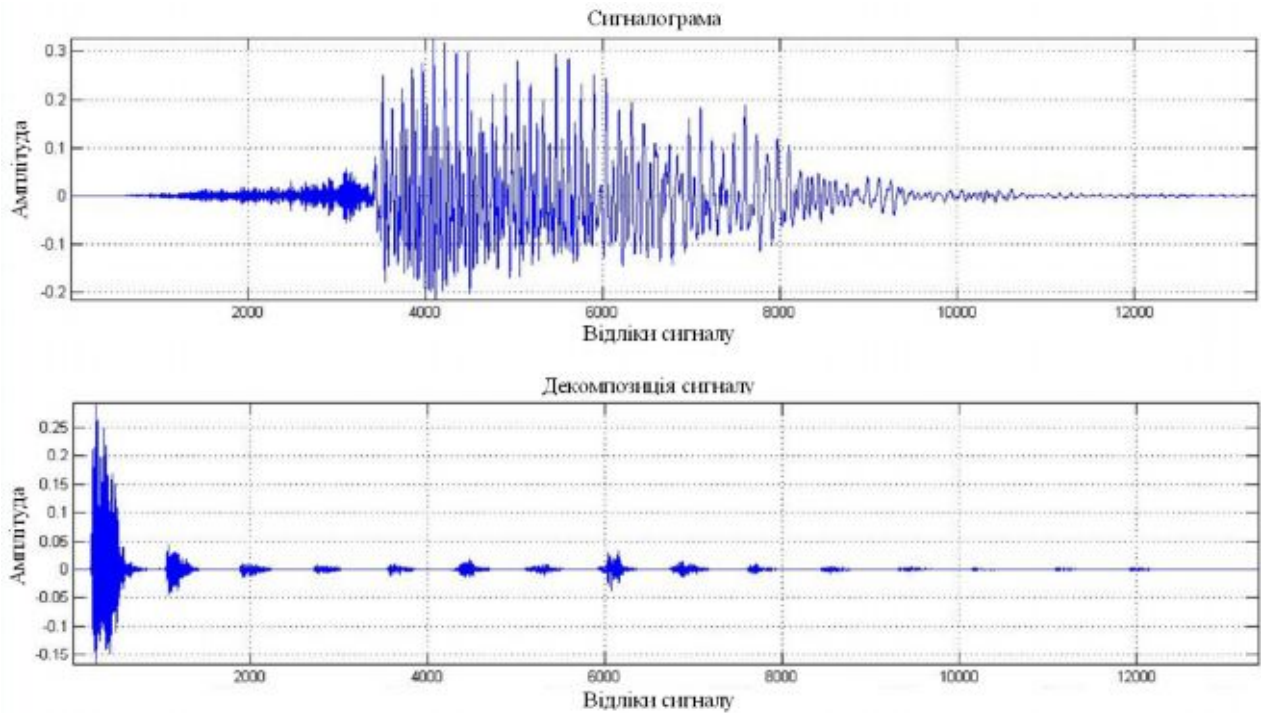


Рис. 2.5. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову “сім” та вектор ознак -декомпозиція сигналу банком фільтрів

2.3.2. Розподіл миттєвих енергій відрізка МС.

Іншою, важливою ознакою може бути вектор розподілу миттєвих енергій.

Для отримання миттєвих енергій для частотного інтервалу r скористаємося формулою:

$$\psi_r(i) = y_r^2(i) \quad (2.17)$$

Довжина вектора відповідає довжині аналізованого сигналу.

На рисунках 2.6-2.8 представлені відрізки сигналу, що відповідають вимовленим словам “вісім” та “сім” та розподіл миттєвих енергій цих сигналів. Дані графіки дозволяють візуально оцінити рівень релевантності аналізованої ознаки.

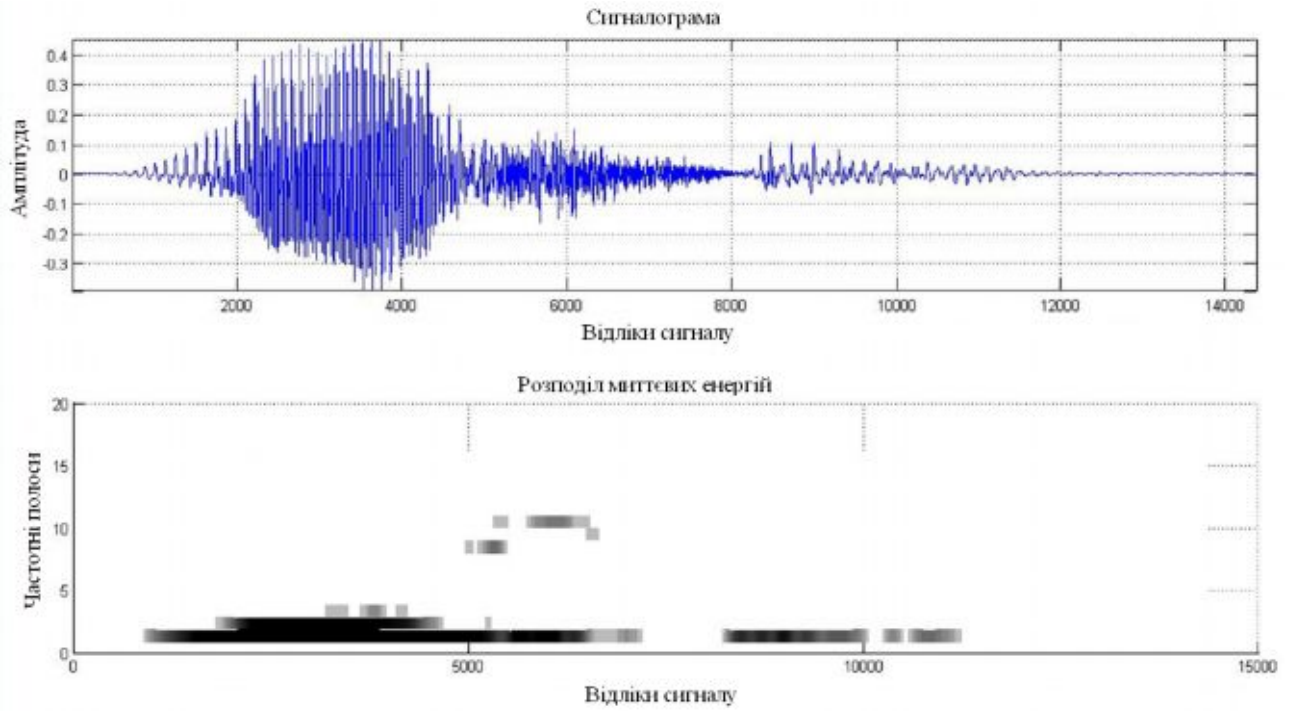


Рис. 2.6. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" та вектор ознак - розподіл миттєвих енергій МС.

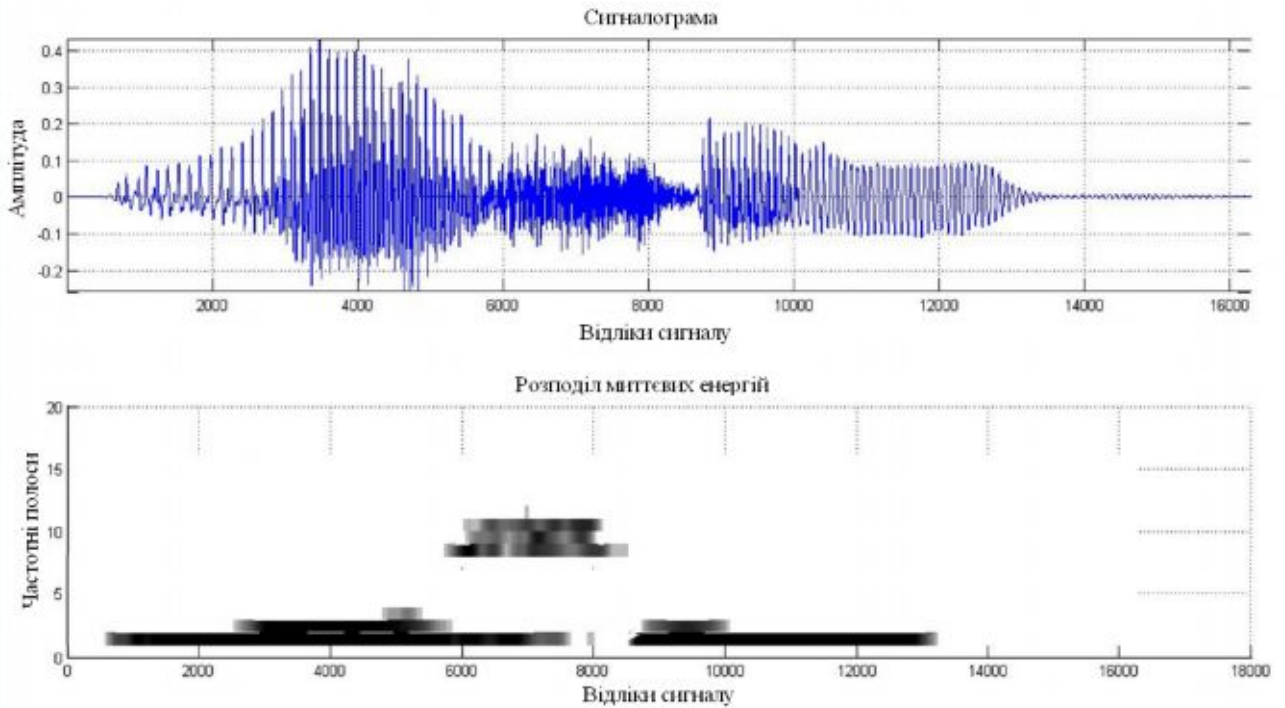


Рис. 2.7. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" та вектор ознак - розподіл миттєвих енергій МС.

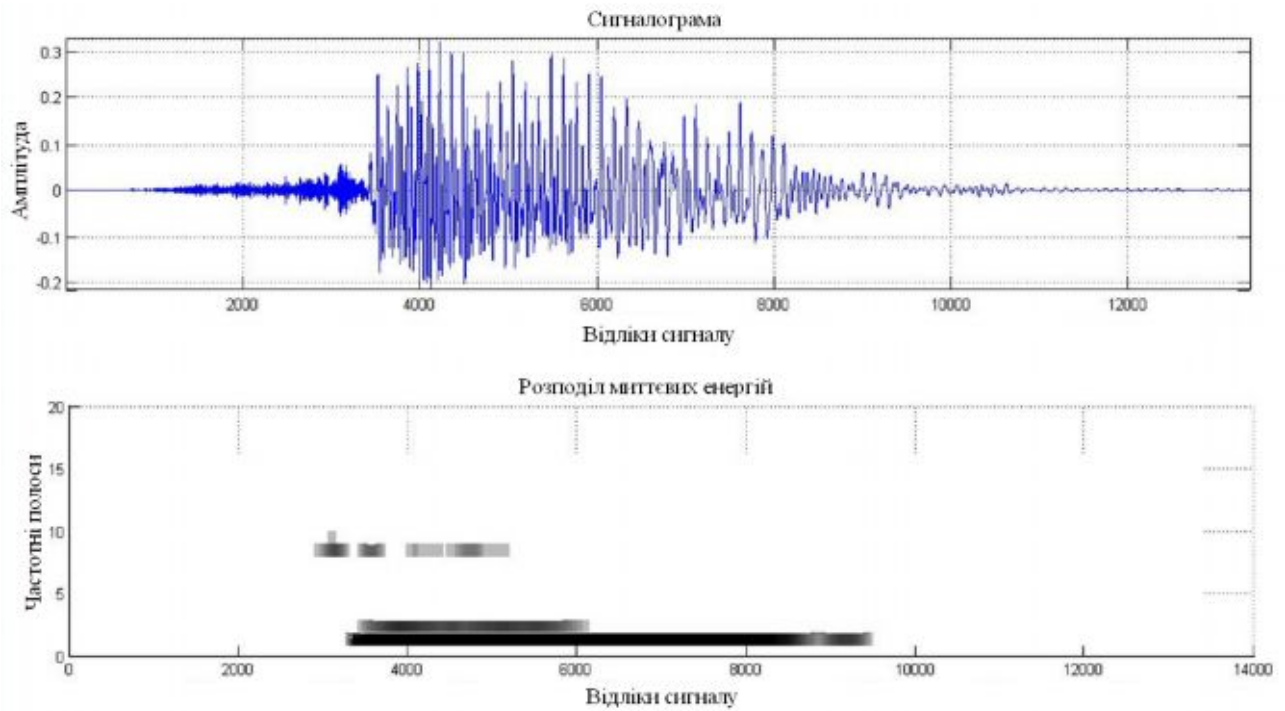


Рис. 2.8. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" та вектор ознак - розподіл миттєвих енергій МС.

2.3.3. Розподіл часток енергії відрізка МС.

Розподіл часток енергії по частотним інтервалам є основним і однією з найбільш інформативних ознак мовного сигналу.

Для отримання розподілу часток енергії скористаємося раніше обчисленою ознакою y_r і застосувавши формулу:

$$P_r = \sum_{i=1}^N y_{r_i}^2 \quad (2.18)$$

де $i=1 \dots N$, відліки сигналу $\vec{x} = (x_1 \dots x_N)$.

В результаті сформуємо вектор запропонованої ознаки у такому вигляді:

$$\vec{P} = (P_1 \dots P_R)$$

де R , кількість частотних смуг.

Отримаємо розподіл часток енергії відрізка шуканого сигналу. Довжина

вектора відповідає кількості частотних інтервалів R .

На рисунках 2.9-2.11 представлені відрізки сигналу, що відповідають вимовленим словам “вісім” та “сім” та розподіл часток енергії цих сигналів. Дані графіки дозволяють візуально оцінити рівень релевантності аналізованої ознаки.

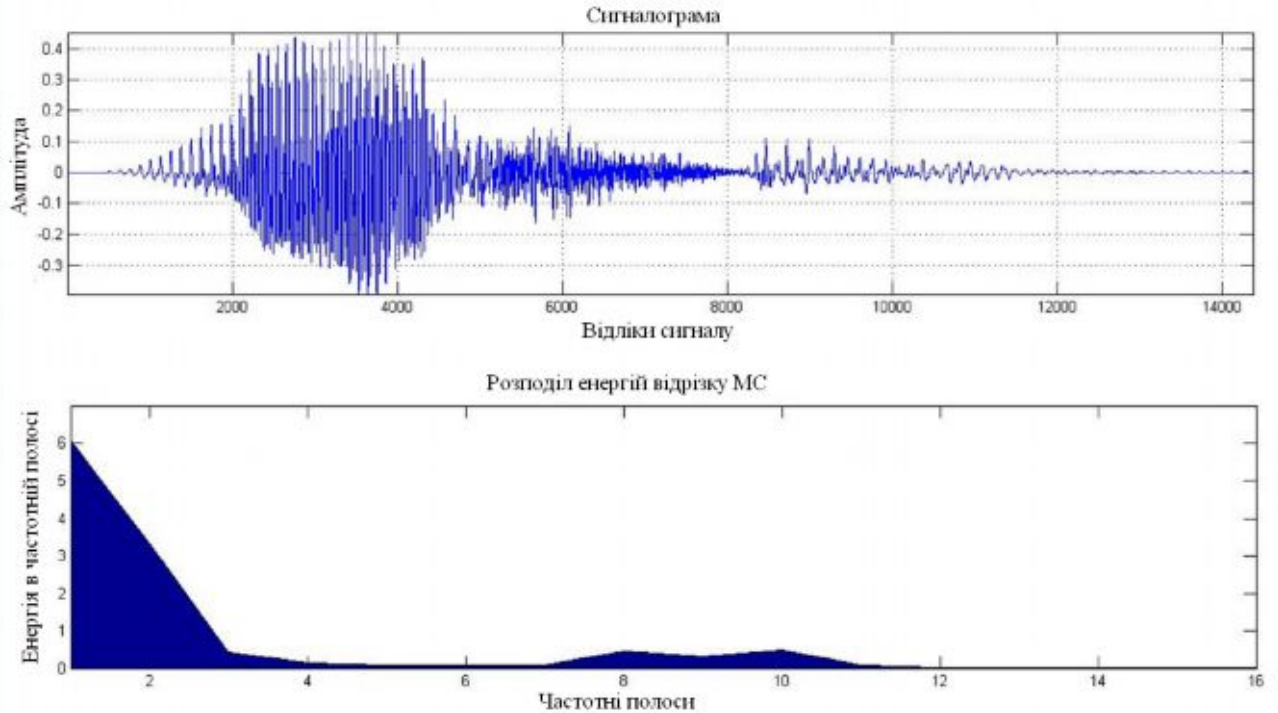


Рис. 2.9. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" і вектор ознак -розподіл часток енергії МС.

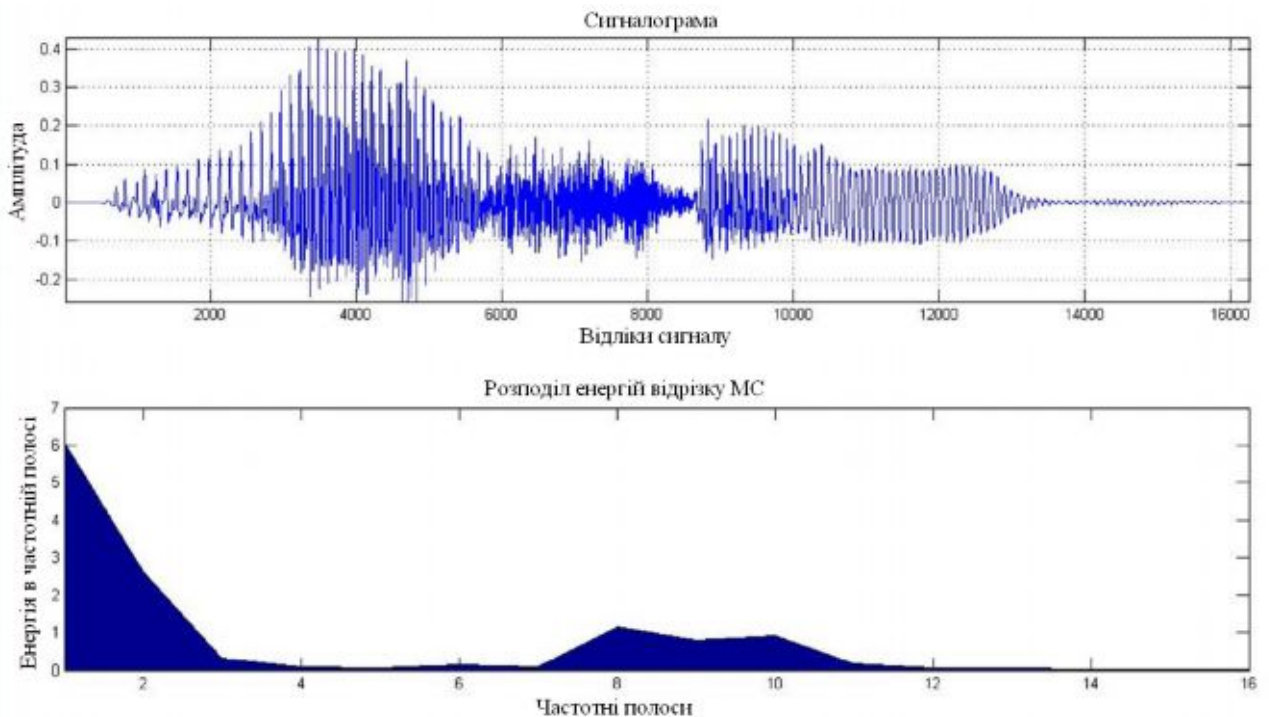


Рис. 2.10. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" і вектор ознак розподіл часток енергії МС

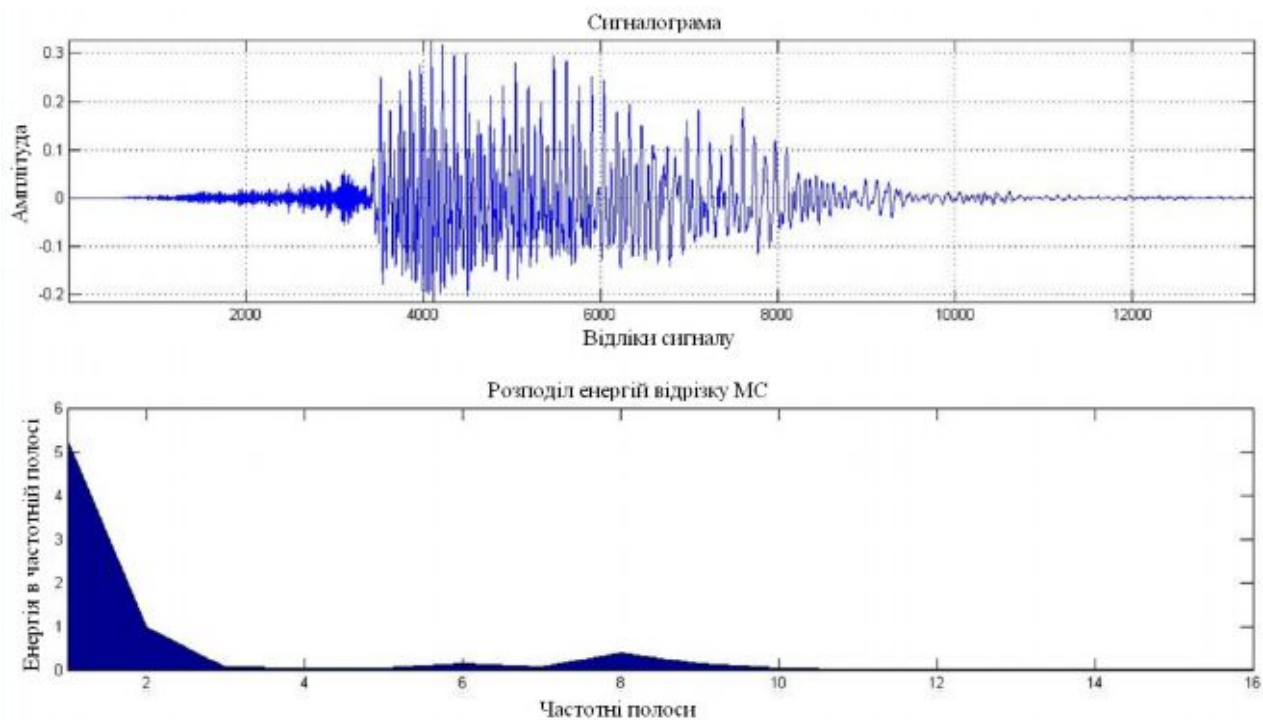


Рис. 2.11. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "сім" та вектор ознак - розподіл часток енергії МС.

2.3.4. Розподіл інформаційних інтервалів відрізка МС.

Для селекції найінформативніших ознак даного вектора зробимо розрахунок порога:

$$h_{\psi} = \frac{1}{N} \sum_{i=1}^N x_i^2 / R \quad (2.19)$$

де $i = 1 \dots N$, відліки сигналу $\vec{x} = (x_1 \dots x_N)$,

Порівнювання значення $\vec{\psi}_r$ з порогом виконується такий вираз:

$$\psi_r(i) = \begin{cases} \psi_r(i), & \psi_r(i) \geq h_{\psi}, \\ 0, & \psi_r(i) < h_{\psi} \end{cases} \quad (2.20)$$

Матриця значення для всіх частотних інтервалів:

$$\Psi = \begin{bmatrix} \tilde{\psi}_1 \\ \dots \\ \tilde{\psi}_R \end{bmatrix} \quad (2.21)$$

де R , кількість частотних смуг.

Для отримання нормованого розподілу інформаційних інтервалів необхідно підсумовувати значення, отримані в матриці (2.21) за стовпцями:

$$\xi_r = \frac{1}{N} \sum_{i=1}^N \psi_r(i) \quad (2.22)$$

Результат - вектор розподілу інформаційних інтервалів:

$$\bar{\xi} = (\xi_1 \dots \xi_R) \quad (2.23)$$

де R , кількість частотних смуг

На рисунках 2.12-2.14 представлені відрізки сигналу, що відповідають вимовленим словам “вісім” та “сім” та розподіл інформаційних інтервалів цих сигналів. Дані графіки дозволяють візуально оцінити рівень релевантності аналізованої ознаки.

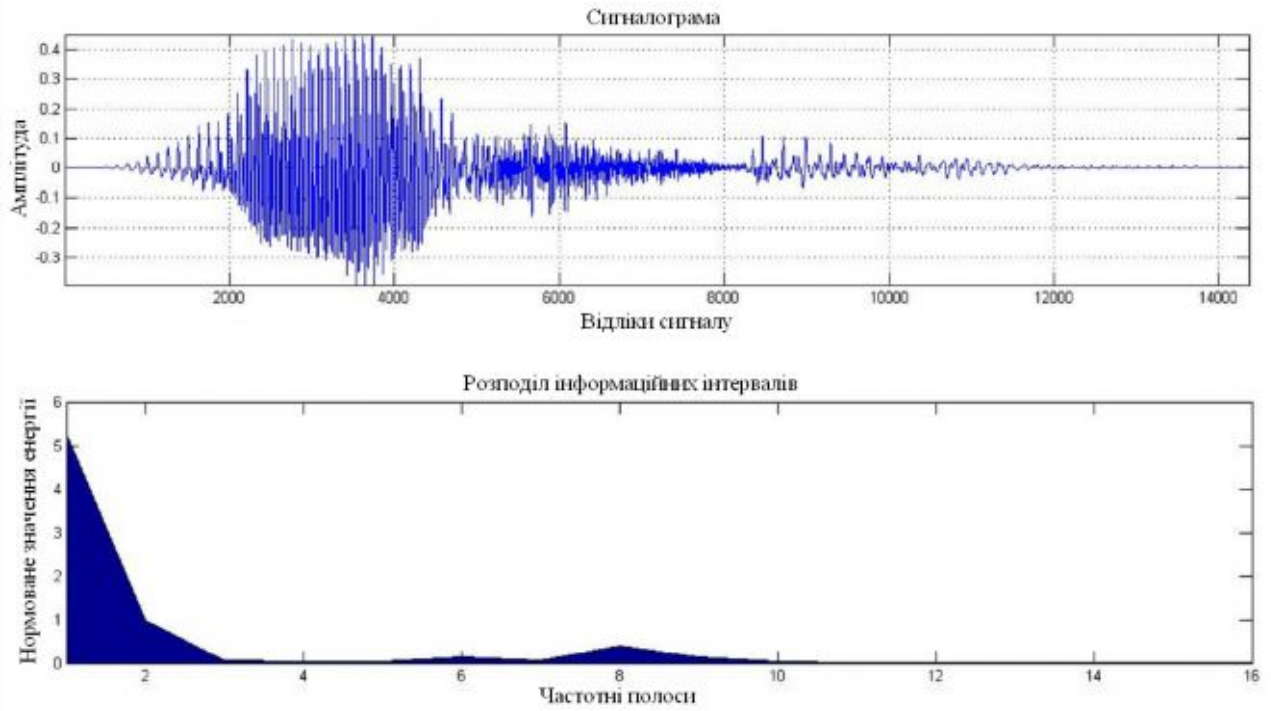


Рис. 2.12. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" та вектор ознак - розподіл інформаційних інтервалів.

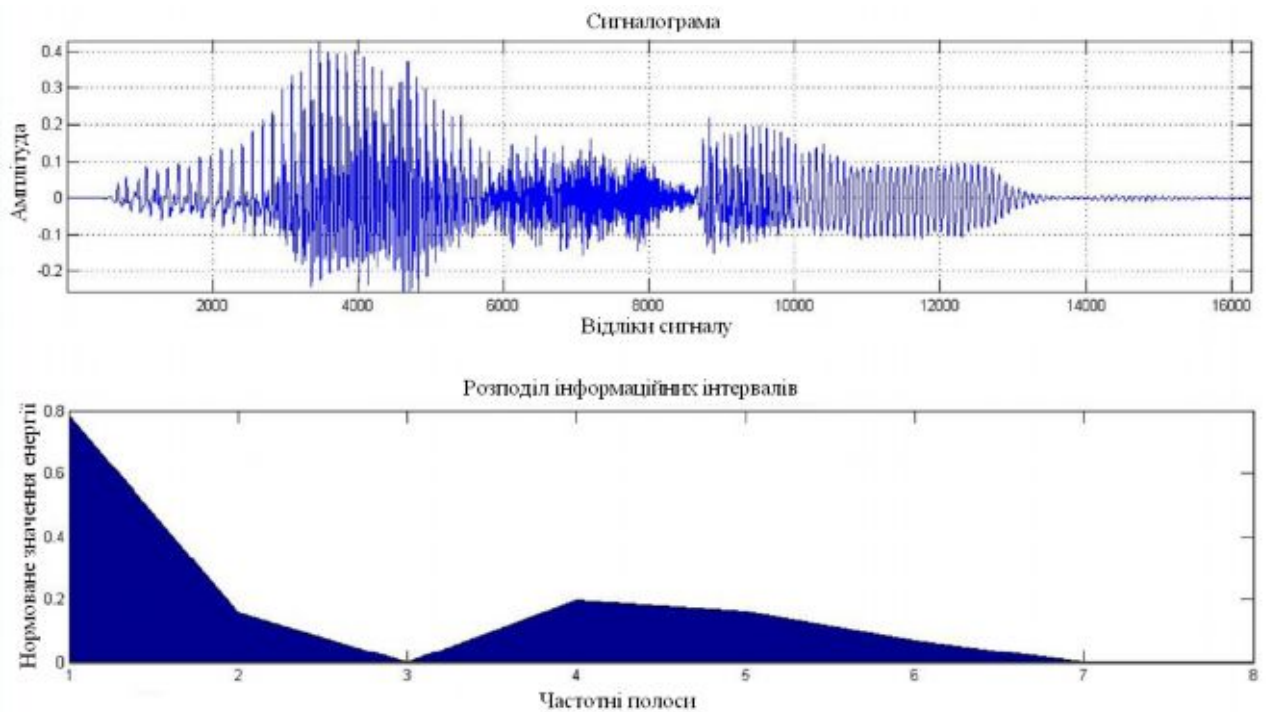


Рис. 2.13. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" та вектор ознак - розподіл інформаційних інтервалів.

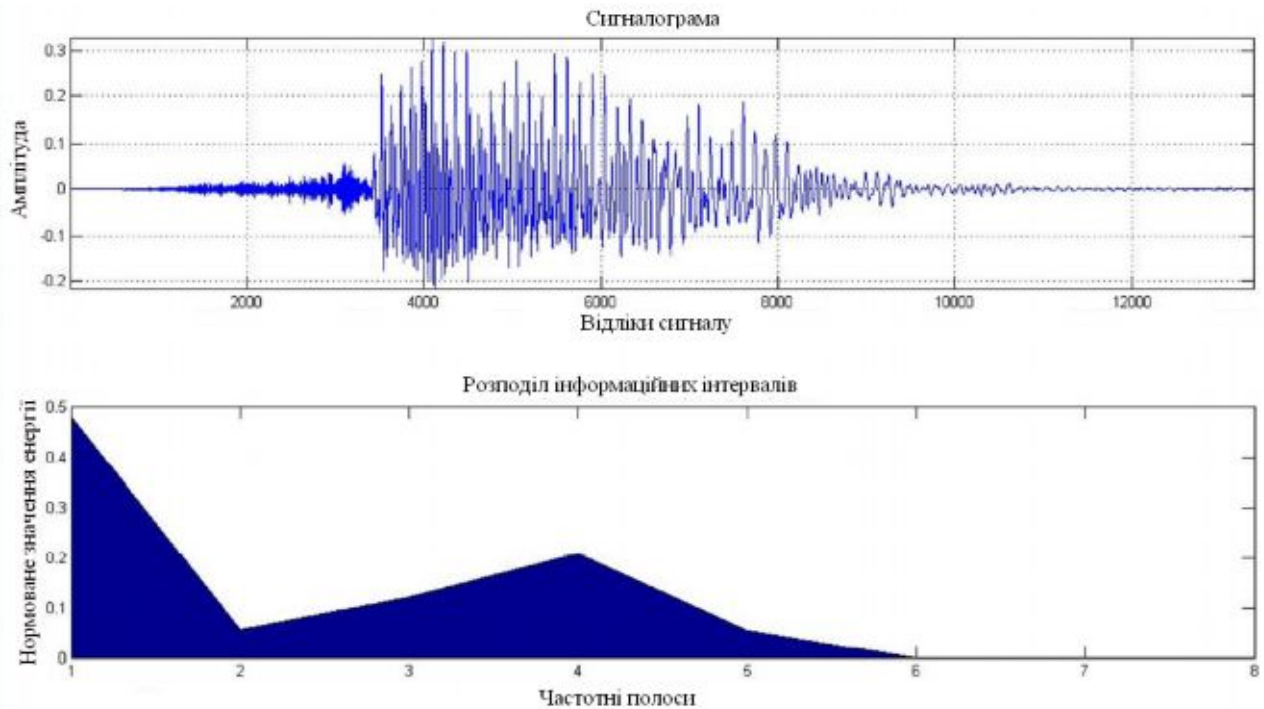


Рис. 2.14. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову “сім” та вектор ознак -розподіл інформаційних інтервалів.

2.3.5. Частота переходів через нуль.

При обробці сигналів у дискретному часі вважається, що якщо два послідовні звіти мають різні знаки, то відбувся перехід сигналу через нуль. Частота появи нулів у сигналі може бути однією з показників його спектральних якостей. Це найбільш справедливо для вузькосмугових синусоїдальних сигналів. [33]

При аналізі широкоплоских мовних сигналів частота переходів через нуль менш очевидно описує спектральні властивості сигналу, проте навіть грубі оцінки, отримані таким чином, є досить вагомим знаком.

Ознака, що показує частоту переходів сигналу через нуль, на заданій ділянці мовного сигналу може бути використаний виділення шумоподібних фонем, основною ознакою яких є висока частота подібних переходів. Результати роботи даної функції разом з такою ознакою як ширина частотної смуги, займана сигналом, дозволяють детектувати ділянки, які можна віднести до галасливих приголосних.

Визначимо середнє число переходів сигналу через нуль так:

$$\sigma_n = \sum_{i=1}^N |\operatorname{sgn}[x_i] - \operatorname{sgn}[x_{i-1}]| \cdot w(n-i), \quad (2.24)$$

де

$$\operatorname{sgn}[x_i] = \begin{cases} 1, & x_i \geq 0; \\ -1, & x_i < 0 \end{cases} \quad (2.25)$$

а також

$$w(n) = \begin{cases} 1/2N, & 0 \leq n \leq N-1, \\ 0, & \text{в іншому випадку} \end{cases} \quad (2.26)$$

Таким чином, можна сформувати вектор запропонованої ознаки:

$$\vec{\omega} = (w_1 \dots w_N); \quad (2.27)$$

де N - кількість відліків сигналу $\vec{x} = (x_1 \dots x_N)$.

Модель мовоутворення передбачає, що енергія вокалізованих сегментів мовного сигналу концентрується на частотах нижче 3 кГц, що зменшується спектром сигналу збудження, тоді як невокалізованих сегментів більшість енергії лежить у областях високих частот. Оскільки високі частоти призводять до великого числа переходів через нуль, а низькі до малого, існує жорсткий зв'язок між числом нульових перетинів і розподілом енергії по частотах. [33]

Можна припустити, що великому числу нульових перетинів відповідає невокалізовані сегменти, а малому числу – вокалізовані сегменти. Однак через невизначеність заходи в даному випадку виникає складність в однозначному визначенні. Для уточнення результатів, отриманої від розрахунку даної характеристики, можна використовувати коефіцієнт ширини частотної області займаної сигналом.

На рисунках 2.15-2.17 представлені відрізки сигналу, що відповідають словам “вісім” та “сім” та функція частоти переходів сигналу через нуль. Дані графіки дозволяють візуально оцінити рівень релевантності аналізованої

ознаки.

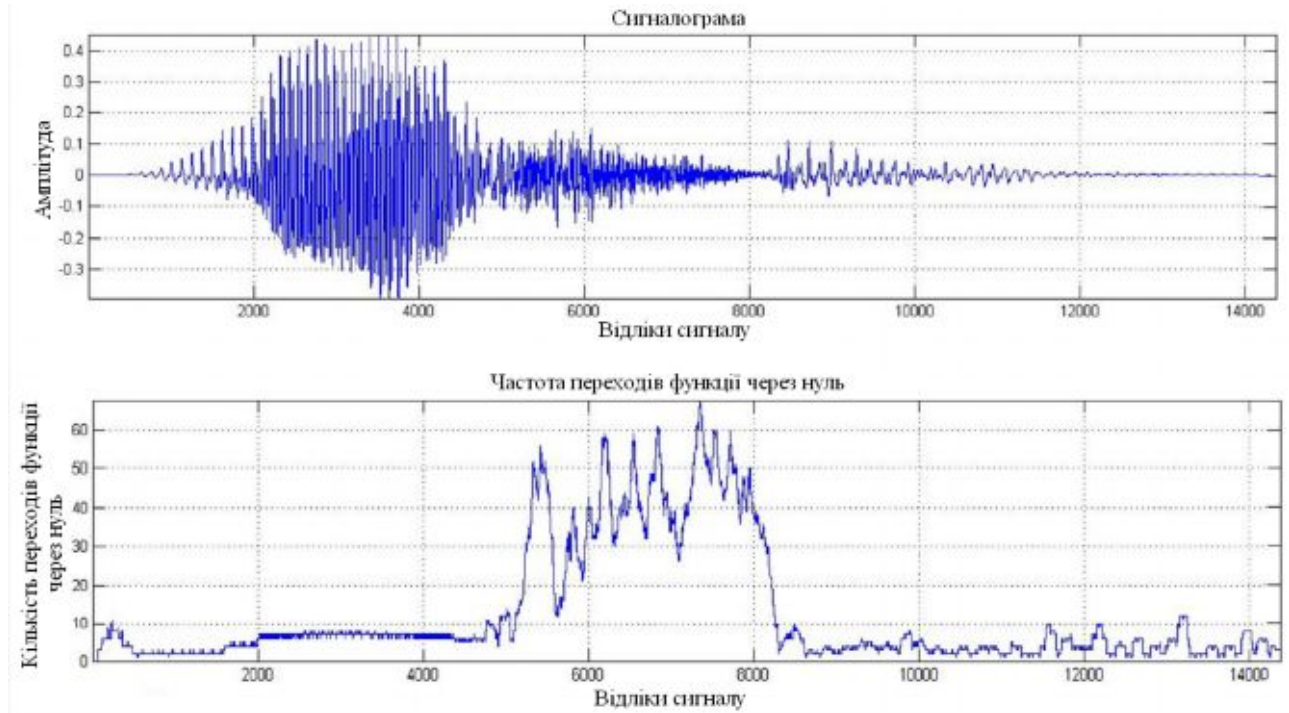


Рис. 2.15. Сигналограма відрізка мовного сигналу, що відповідає слову "вісім" та вектор ознак - Частота переходів функції через нуль.

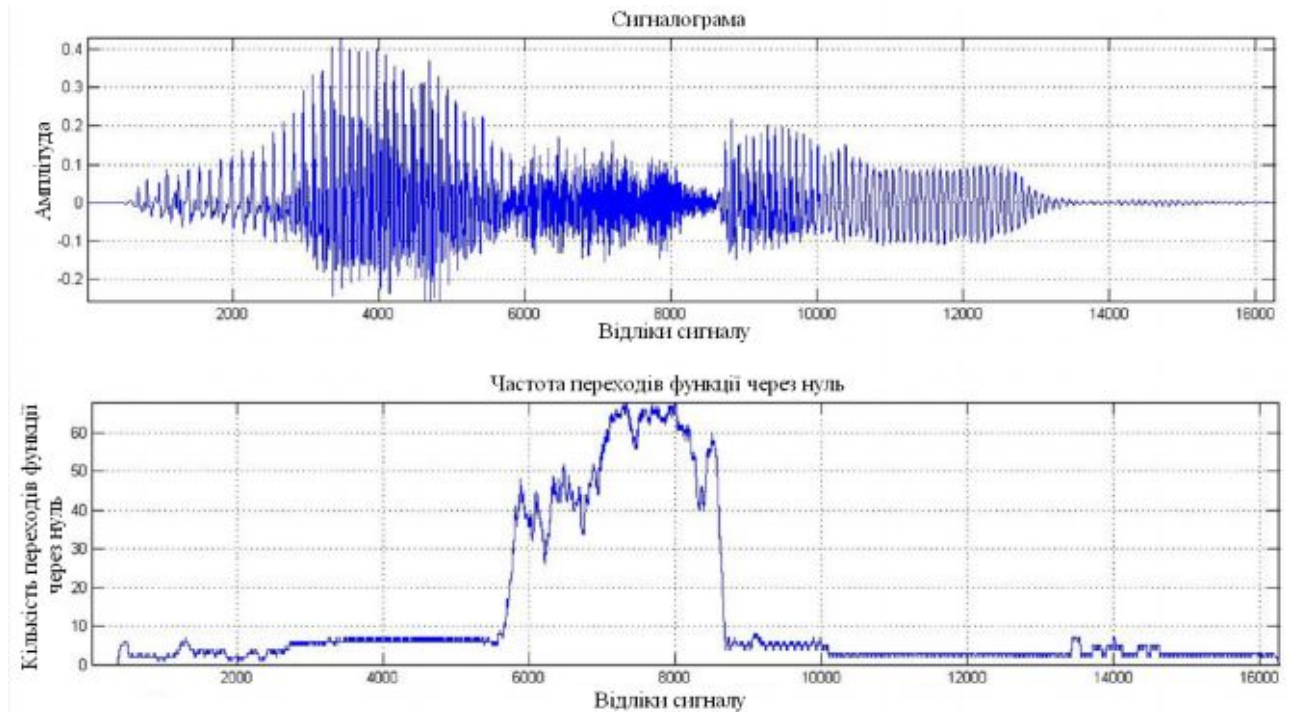


Рис. 2.16. Сигналограма відрізка мовного сигналу, що відповідає слову "вісім" та вектор ознак - Частота переходів функції через нуль.

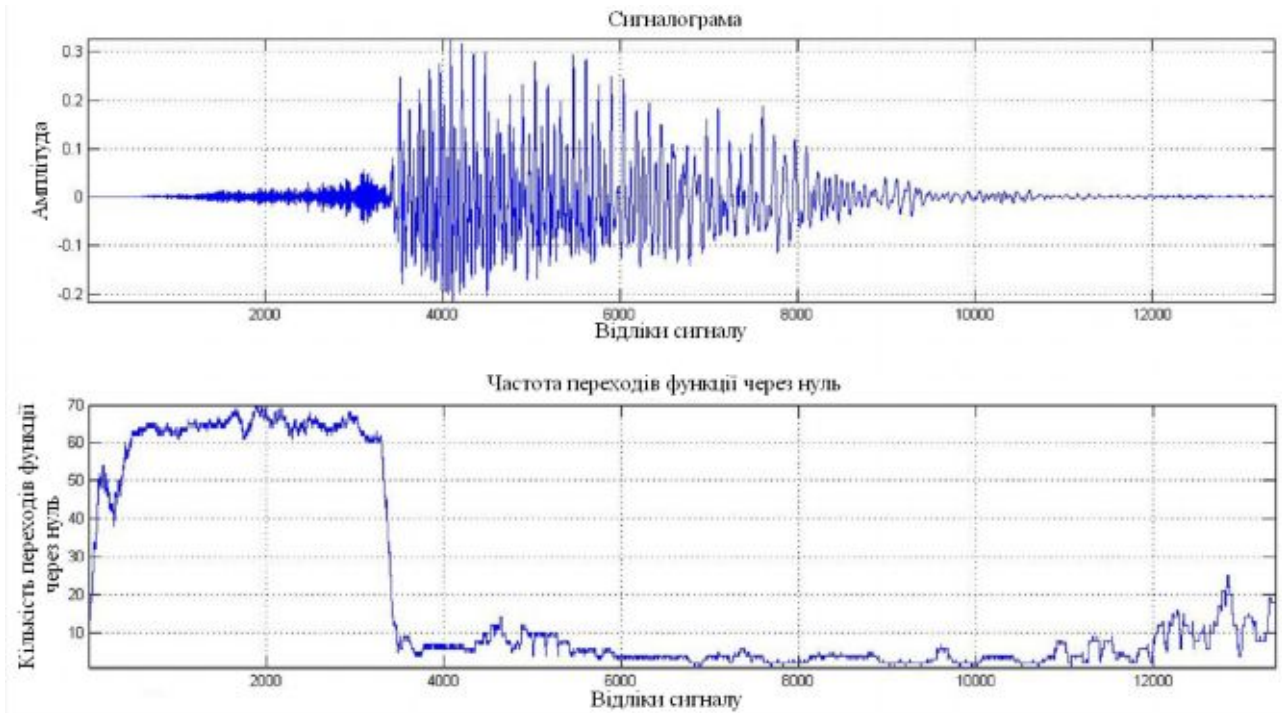


Рис. 2.17. Сигналограма відрізка мовного сигналу, що відповідає слову “сім” та вектор ознак - Частота переходів функції через нуль.

2.3.6. Ширина частотної області, що займає сигнал

З результатів дослідження розподілу інформаційних інтервалів, для звуків характерною особливістю може бути ширина частотної області, у якій сконцентровані інформаційні інтервали. Наприклад, вокалізовані звуки відрізняються досить щільною концентрацією інформаційних інтервалів (~10% від загальної смуги), тоді як невокалізовані звуки (Ж, З, Ф, Х, Ц, Т, Ч, С, Ш, Щ) займають близько ~45 -70% частотної смуги. Отже, можна дійти невтішного висновку, у тому, що це ознака може бути ідентифікації невокалізованих відрізків мовного сигналу. [25,26]

Для оцінки ширини частотної області, що займає сигнал, пропонується сформуванню наступний вектор: $\vec{v}=(v_1..v_N)$, N – тривалість аналізованого відрізка.

Для оцінки ширини частотної області необхідно піддати аналізу розподіл часток енергій сигналу отримані у виразі (2.20)

$$\Psi = \begin{bmatrix} \Psi_{11} & \dots & \dots & \Psi_{1N} \\ \Psi_{21} & \dots & \dots & \Psi_{2N} \\ \dots & \dots & \dots & \dots \\ \Psi_{R1} & \dots & \dots & \Psi_{RN} \end{bmatrix},$$

де R - кількість частотних інтервалів, а N – довжина аналізованого сигналу.

Необхідно здійснити сортування значень матриці Ψ у кожному стовпці N за спаданням:

$$\psi_1^n > \psi_2^n > \dots > \psi_r^n \geq 0 \quad (2.28)$$

Далі здійснимо обчислення суми енергії для кожного звіту сигналу по всіх інтервалах:

$$\Sigma_n = \sum_{r=1}^R \psi_r^n \quad (2.29)$$

В результаті сформуємо вектор Σ :

$$\vec{\Sigma} = (\Sigma_1 \dots \Sigma_N); \quad (2.30)$$

де N , кількість відліків сигналу $\vec{x} = (x_1 \dots x_N)$.

Виробляємо розрахунок ширини частотної області, займаної сигналом відповідно до встановленого коефіцієнта концентрації енергії K , за допомогою буферного вектора $\vec{\beta}_r$:

$$\vec{\beta}_r = \sum_{i=1}^r \vec{\psi}_i \quad (2.31)$$

Коефіцієнт концентрації енергії K несе у собі наступний сенс – відносне

кількість енергії, зосередженої у мінімальній кількості частотних інтервалів.

Далі робимо послідовне підсумовування вектора $\vec{\beta}_r$ для кожного рядка та порівняння з вектором Σ :

$$v(n) = \begin{cases} i, & \text{якщо } \beta_r(i) / \Sigma(i) \geq K \\ v(n), & \text{якщо } \beta_r(i) / \Sigma(i) < K \end{cases} \quad (2.32)$$

На виході отримуємо вектор значень ознаки у наступній формі:

$$\vec{v} = (v_1 \dots v_N) \quad (2.33)$$

На рисунку 2.18 представлена візуалізація алгоритму розрахунок ширини частотної області.

| matrix (P) | x1 | x2 | x3 | x4 | x5 | x6 |
|------------|-------|-------|-------|-------|-------|-------|
| r1 | 82,0 | 28,0 | 96,0 | 80,0 | 68,0 | 71,0 |
| r2 | 91,0 | 55,0 | 49,0 | 96,0 | 76,0 | 4,0 |
| r3 | 13,0 | 96,0 | 81,0 | 66,0 | 75,0 | 28,0 |
| r4 | 92,0 | 97,0 | 15,0 | 4,0 | 40,0 | 5,0 |
| r5 | 64,0 | 16,0 | 43,0 | 85,0 | 66,0 | 10,0 |
| r6 | 10,0 | 98,0 | 92,0 | 94,0 | 18,0 | 83,0 |
| summ | 352,0 | 390,0 | 376,0 | 425,0 | 343,0 | 201,0 |
| K-factor | 0,65 | | | | | |
| buffer | 0 | 0 | 0 | 0 | 0 | 0 |
| compare | 0 | 0 | 0 | 0 | 0 | 0 |
| x | 0 | 0 | 0 | 0 | 0 | 0 |

| matrix (P) | x1 | x2 | x3 | x4 | x5 | x6 |
|------------|-------|-------|-------|-------|-------|-------|
| r1 | 92,0 | 98,0 | 96,0 | 96,0 | 76,0 | 83,0 |
| r2 | 91,0 | 97,0 | 92,0 | 94,0 | 75,0 | 71,0 |
| r3 | 82,0 | 96,0 | 81,0 | 85,0 | 68,0 | 28,0 |
| r4 | 64,0 | 55,0 | 49,0 | 80,0 | 66,0 | 10,0 |
| r5 | 13,0 | 28,0 | 43,0 | 66,0 | 40,0 | 5,0 |
| r6 | 10,0 | 16,0 | 15,0 | 4,0 | 18,0 | 4,0 |
| summ | 352,0 | 390,0 | 376,0 | 425,0 | 343,0 | 201,0 |
| K-factor | 0,65 | | | | | |
| buffer | 0 | 0 | 0 | 0 | 0 | 0 |
| compare | 0 | 0 | 0 | 0 | 0 | 0 |
| x | 0 | 0 | 0 | 0 | 0 | 0 |

| matrix (P) | x1 | x2 | x3 | x4 | x5 | x6 |
|------------|----------|----------|----------|----------|----------|----------|
| r1 | 92,0 | 98,0 | 96,0 | 96,0 | 76,0 | 83,0 |
| r2 | 91,0 | 97,0 | 92,0 | 94,0 | 75,0 | 71,0 |
| r3 | 82,0 | 96,0 | 81,0 | 85,0 | 68,0 | 28,0 |
| r4 | 64,0 | 55,0 | 49,0 | 80,0 | 66,0 | 10,0 |
| r5 | 13,0 | 28,0 | 43,0 | 66,0 | 40,0 | 5,0 |
| r6 | 10,0 | 16,0 | 15,0 | 4,0 | 18,0 | 4,0 |
| summ | 352,0 | 390,0 | 376,0 | 425,0 | 343,0 | 201,0 |
| K-factor | 0,65 | | | | | |
| buffer | 92,0 | 98,0 | 96,0 | 96,0 | 76,0 | 83,0 |
| compare | 0,261364 | 0,251282 | 0,255319 | 0,225882 | 0,221574 | 0,412935 |
| x | 1 | 1 | 1 | 1 | 1 | 1 |

| matrix (P) | x1 | x2 | x3 | x4 | x5 | x6 |
|------------|----------|-------|-------|----------|----------|----------|
| r1 | 92,0 | 98,0 | 96,0 | 96,0 | 76,0 | 83,0 |
| r2 | 91,0 | 97,0 | 92,0 | 94,0 | 75,0 | 71,0 |
| r3 | 82,0 | 96,0 | 81,0 | 85,0 | 68,0 | 28,0 |
| r4 | 64,0 | 55,0 | 49,0 | 80,0 | 66,0 | 10,0 |
| r5 | 13,0 | 28,0 | 43,0 | 66,0 | 40,0 | 5,0 |
| r6 | 10,0 | 16,0 | 15,0 | 4,0 | 18,0 | 4,0 |
| summ | 352,0 | 390,0 | 376,0 | 425,0 | 343,0 | 201,0 |
| K-factor | 0,65 | | | | | |
| buffer | 183,0 | 195,0 | 188,0 | 190,0 | 151,0 | 154,0 |
| compare | 0,519886 | 0,5 | 0,5 | 0,447059 | 0,440233 | 0,766169 |
| x | 2 | 2 | 2 | 2 | 2 | 1 |

| matrix (P) | x1 | x2 | x3 | x4 | x5 | x6 |
|------------|----------|----------|----------|----------|----------|----------|
| r1 | 92,0 | 98,0 | 96,0 | 96,0 | 76,0 | 83,0 |
| r2 | 91,0 | 97,0 | 92,0 | 94,0 | 75,0 | 71,0 |
| r3 | 82,0 | 96,0 | 81,0 | 85,0 | 68,0 | 28,0 |
| r4 | 64,0 | 55,0 | 49,0 | 80,0 | 66,0 | 10,0 |
| r5 | 13,0 | 28,0 | 43,0 | 66,0 | 40,0 | 5,0 |
| r6 | 10,0 | 16,0 | 15,0 | 4,0 | 18,0 | 4,0 |
| summ | 352,0 | 390,0 | 376,0 | 425,0 | 343,0 | 201,0 |
| K-factor | 0,65 | | | | | |
| buffer | 265,0 | 291,0 | 269,0 | 275,0 | 219,0 | 182,0 |
| compare | 0,752841 | 0,746154 | 0,715426 | 0,647059 | 0,638484 | 0,905473 |
| x | 2 | 2 | 2 | 3 | 3 | 1 |

| matrix (P) | x1 | x2 | x3 | x4 | x5 | x6 |
|------------|----------|----------|----------|----------|----------|----------|
| r1 | 92,0 | 98,0 | 96,0 | 96,0 | 76,0 | 83,0 |
| r2 | 91,0 | 97,0 | 92,0 | 94,0 | 75,0 | 71,0 |
| r3 | 82,0 | 96,0 | 81,0 | 85,0 | 68,0 | 28,0 |
| r4 | 64,0 | 55,0 | 49,0 | 80,0 | 66,0 | 10,0 |
| r5 | 13,0 | 28,0 | 43,0 | 66,0 | 40,0 | 5,0 |
| r6 | 10,0 | 16,0 | 15,0 | 4,0 | 18,0 | 4,0 |
| summ | 352,0 | 390,0 | 376,0 | 425,0 | 343,0 | 201,0 |
| K-factor | 0,65 | | | | | |
| buffer | 329,0 | 346,0 | 318,0 | 355,0 | 285,0 | 192,0 |
| compare | 0,934659 | 0,887179 | 0,845745 | 0,835294 | 0,830904 | 0,955224 |
| x | 2 | 2 | 2 | 3 | 3 | 1 |

Рис. 2.18. Візуалізація алгоритму визначення ширини частотної області, що займається сигналом.

На рисунках 2.19-2.21 представлені відрізки сигналу, що відповідають вимовленим словам "вісім" та "сім" та функція ширини частотної області, що займає сигнал. Дані графіки дозволяють візуально оцінити рівень релевантності аналізованої ознаки.

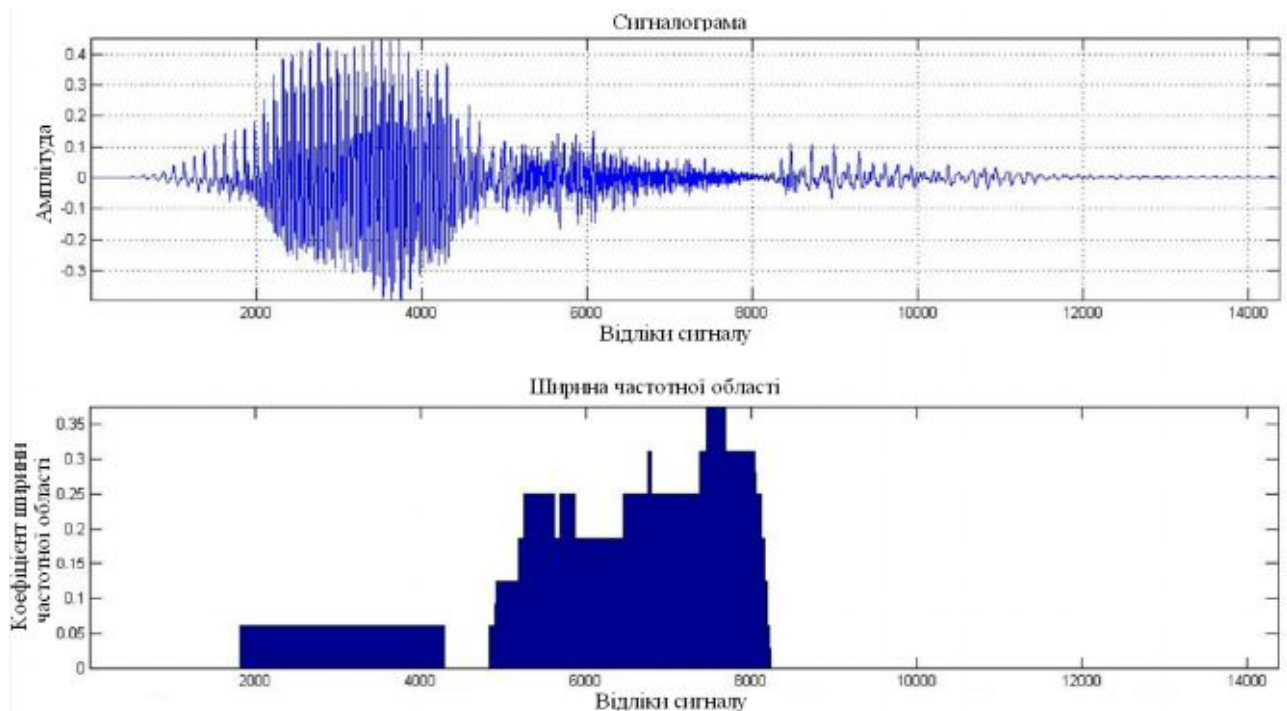


Рис. 2.19. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" та вектор ознак - Ширина частотної області, займана сигналом.

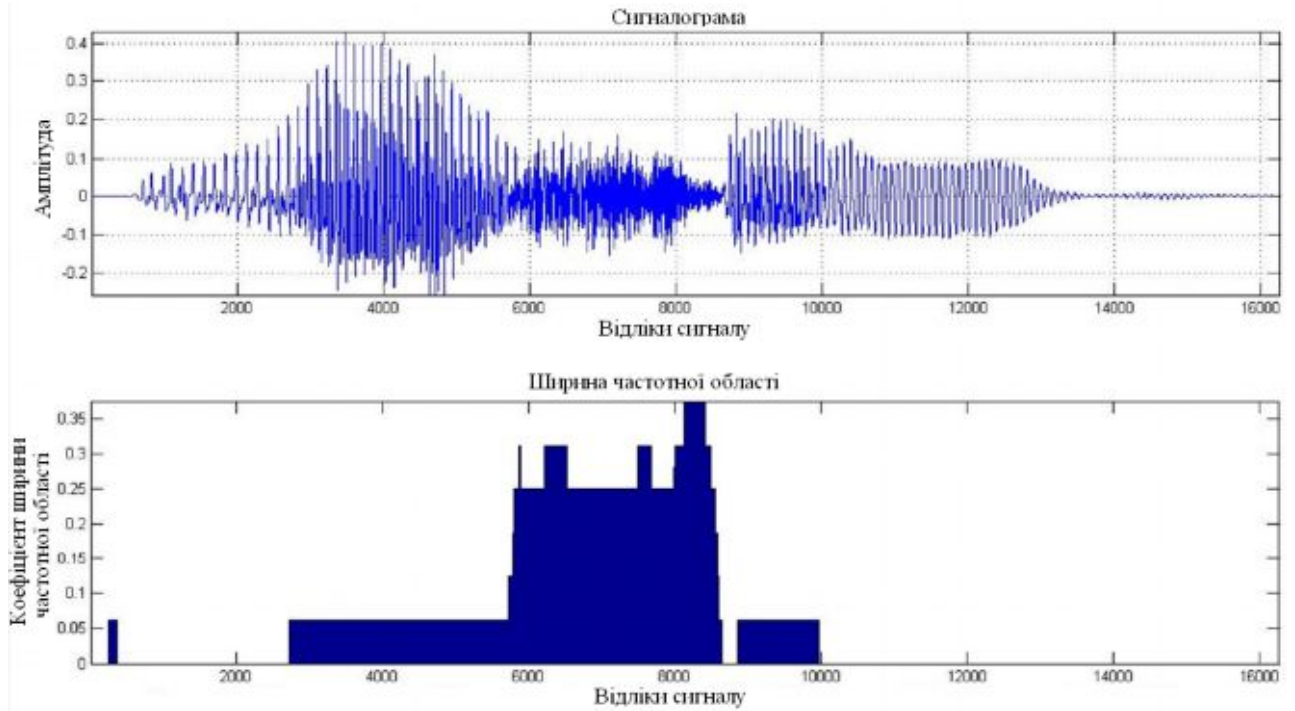


Рис. 2.20. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "вісім" та вектор ознак - Шириня частотної області, займана СИГНАЛОМ.

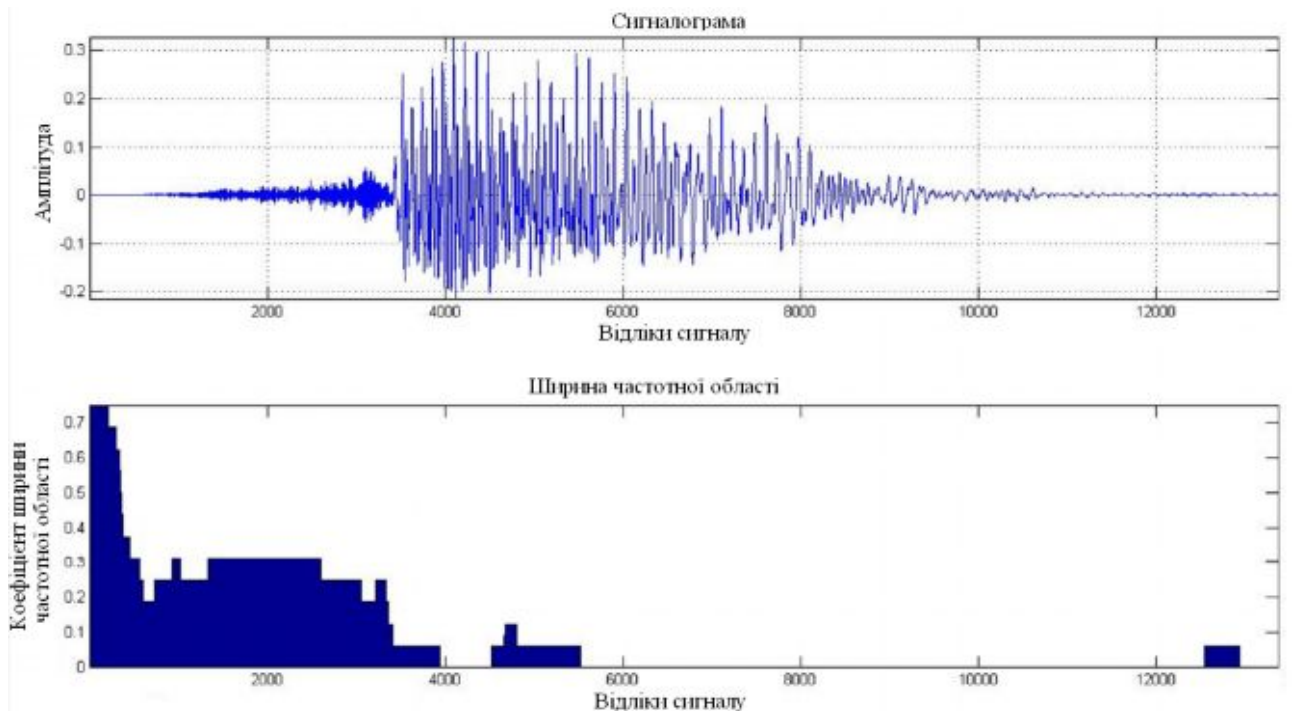


Рис. 2.21. Сигналограма відрізка мовного сигналу, що відповідає вимовленому слову "сім" та вектор ознак - Шириня частотної області, що займається СИГНАЛОМ.

2.3.7. Мел-кепстральні коефіцієнти мовного сигналу

Мел-кепстральні коефіцієнти є таке уявлення сигналу, в якому частотні смуги розподілені відповідно до крейди шкалою (рисунок 1.9). Це призводить до більшої щільності фільтрів в області низьких частот та меншої щільності в області високих частот, що відображає чутливість сприйняття звукових сигналів людським вухом. Таким чином, основна інформація розраховується з аудіо сигналу в області низьких частот, що є найбільш репрезентативним ознакою задач розпізнавання мовних сигналів [25,26,49,50].

Щоб розрахувати крейдиані коефіцієнти потрібно зробити такі дії:

Здійснити розрахунок ДПФ для сигналу (1.9):

$$X(k) = \sum_{i=1}^N x(i) e^{-j \frac{2\pi}{N} (i-1)(k-1)}$$

де $k=1,2,\dots,N$, j - комплексна одиниця, $j = \sqrt{-1}$.

Отриманий спектр сигналу $X(k)$ подається на вхід банку фільтрів. Банк фільтрів є трикутними фільтрами, що перекривають один одного, розподілені за крейдою-шкалою (рисунок 1.9).

$$H_m = \begin{cases} 0 & \\ \frac{(k - f[m-1])}{(f[m] - f[m-1])} & k < f[m-1] \\ \frac{(f[m+1] - k)}{(f[m+1] - f[m])} & f[m-1] \leq k < f[m] \\ \frac{(f[m+1] - k)}{(f[m+1] - f[m])} & f[m] \leq k \leq f[m+1] \\ 0 & k > f[m+1] \end{cases} \quad (2.34)$$

Частоти отримуємо з наступної рівності:

$$f[m] = \left(\frac{N}{F_s} \right) \cdot B^{-1} \left(B(f_1) + m \frac{B(f_h) - B(f_1)}{M+1} \right) \quad (2.35)$$

де

$$B^{-1}(b) = 700(\exp(b/1125) - 1)$$

Далі знайдемо натуральний логарифм від суми часток енергії сигналу за інтервалами, після віконного перетворення Фур'є:

$$S(m) = \ln \left(\sum_{k=0}^{N-1} |X(k)|^2 \cdot H_m(k) \right), \quad 0 \leq m < M$$

Застосовуючи дискретне косинусне перетворення Фур'є (ДКП) отримаємо крейдянні коефіцієнти для відрізка МС

$$\mu(n) = \sum_{m=0}^{M-1} S(m) \cos(\pi n(m+1/2)/M), \quad 0 \leq n < M \quad (2.36)$$

На рисунках 2.22-2.24 представлені відрізки сигналу, що відповідають вимовленим словам “вісім” та “сім” та крейдянні коефіцієнти цих сигналів. Дані графіки дозволяють візуально оцінити рівень релевантності аналізованої ознаки.

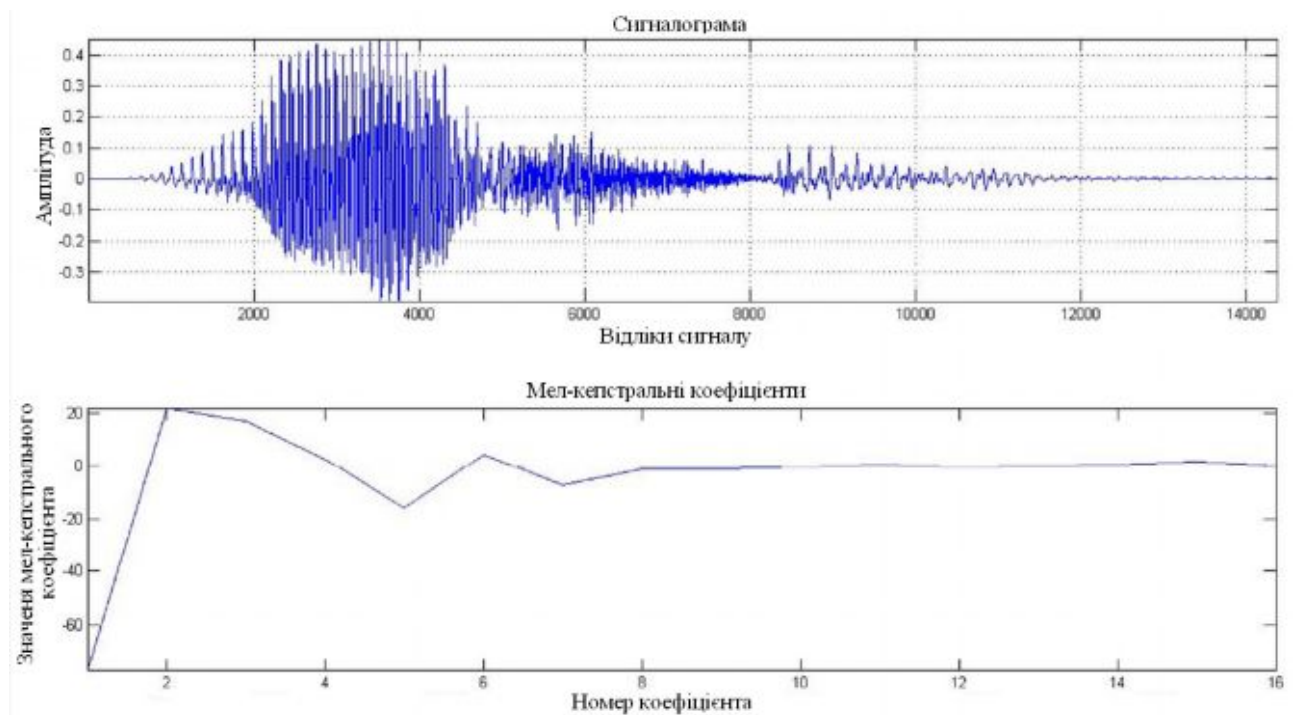


Рис. 2.22. Сигналограма відрізка мовного сигналу, що відповідає

вимовленому слову "вісім" і вектор ознак - Мел-кепстральні коефіцієнти.

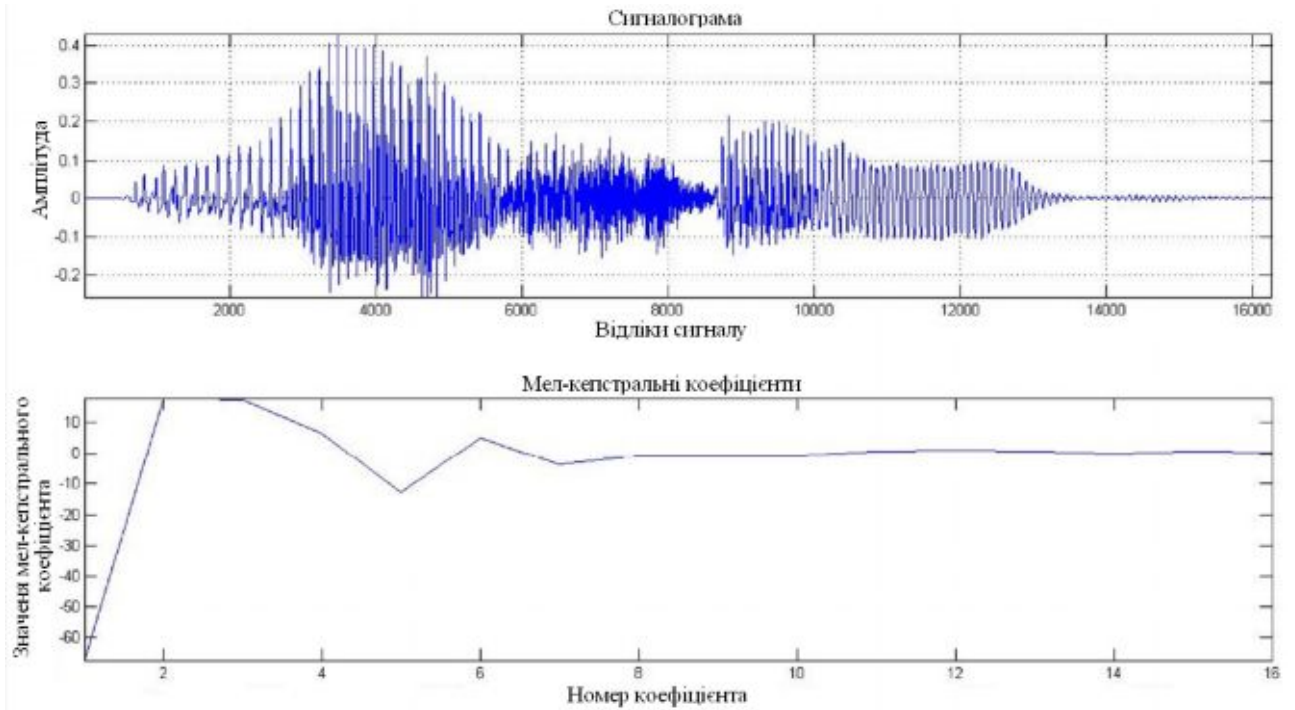


Рис. 2.23. Сигналограма відрізка мовного сигналу, що відповідає слову "вісім" та вектор ознак - Мел-кепстральні коефіцієнти

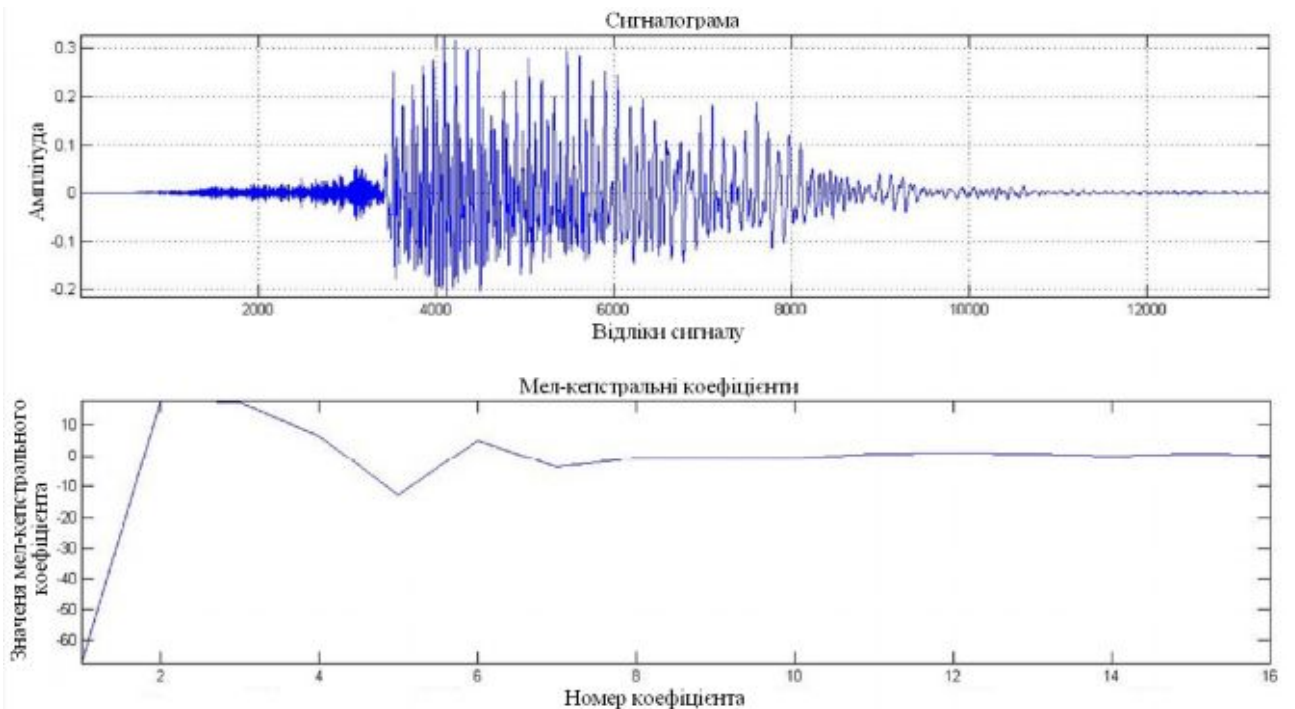


Рис. 2.24. Сигналограма відрізка мовного сигналу, що відповідає слову "сім" та вектор ознак - Мел-кепстральні коефіцієнти.

2.4 Заходи близькості у задачах розпізнавання мовних сигналів

При визначенні гіпотези приналежності порівнюваного об'єкта до того чи іншого класу використовуються різні міри близькості: відстань Евклідова, відстань Махаланобіса, кореляція, динамічна трансформація тимчасової шкали, середньоквадратичне відхилення та інші.

2.4.1. Евклідова відстань

Евклідова відстань є одним з найпопулярніших заходів близькості, яка дозволяє обчислювати відстань ε між двома i -ми точками послідовностей x і y в евклідовому просторі [32], обчислюється виразом (2.37), та являє собою геометричну відстань в багатовимірному просторі:

$$\varepsilon = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (2.37)$$

де x_i - i -ий елемент вектора $\vec{x} = (x_1, x_2, \dots, x_{(i-1)}, x_i)$, $i = 1 \dots N$; y_i - i -ий елемент вектора $\vec{y} = (y_1, y_2, \dots, y_{(i-1)}, y_i)$, $i = 1 \dots N$.

2.4.2. Середньоквадратичне відхилення

Середньоквадратичне відхилення є показником розсіювання значень однієї випадкової величини від іншої та обчислюється таким чином:

$$\delta = \sqrt{\frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N x_i^2}} \quad (2.38)$$

де x_i - i -ий елемент вектора $\vec{x} = (x_1, x_2, \dots, x_{(i-1)}, x_i)$, $i = 1 \dots N$; y_i - i -ий елемент вектора $\vec{y} = (y_1, y_2, \dots, y_{(i-1)}, y_i)$, $i = 1 \dots N$.

2.4.3. Відстань Махаланобіса.

Відстань Махаланобіса [51] є міра відстані між векторами випадкових величин узагальнює розуміння евклідової відстані. За допомогою відстані Махаланобіса можна визначати схожість невідомої та відомої вибірки. Воно відрізняється від відстані Евкліда тим, що враховує кореляції між змінними та інваріантними до масштабу. Розрахунок відстані Махаланобіса можна виконати таким чином:

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})^T S^{-1} (\vec{x} - \vec{y})} \quad (2.39)$$

де \vec{x}, \vec{y} - досліджувані вектори, S -матриця коваріації.

$$S = (\sigma_{ij}), i, j = 1 \dots n \quad (2.40)$$

$$\sigma_{ij} = \text{cov}(x_i, x_j) = M(x_i - Mx_i)(x_j - Mx_j) \quad (2.41)$$

де M – математичне очікування випадкової величини

2.4.4 Кореляція послідовностей

Коефіцієнт кореляції двох послідовностей може бути розрахований наступним чином: дані дві послідовності \vec{x} і \vec{y} довжиною N : $\vec{x} = (x_1, x_2, \dots, x_N)$; $\vec{y} = (y_1, y_2, \dots, y_N)$.

Необхідно провести розрахунок середніх значень кожної послідовності:

$$\bar{X} = \sum_{i=1}^N x_i; \bar{Y} = \sum_{i=1}^N y_i \quad (2.42)$$

Коефіцієнт кореляції можна розрахувати так:

$$\eta = \frac{\sum_{i=1}^N (x_i - \bar{X}) \cdot (y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{Y})^2}} \quad (2.43)$$

Коефіцієнт кореляції η змінюється в межах від -1 до 1. У даному випадку це лінійний коефіцієнт кореляції, він показує лінійний взаємозв'язок між \vec{x} і \vec{y} : η дорівнює 1, якщо зв'язок лінійний (або -1, якщо зв'язок не лінійний) [34].

2.4.5. Динамічна трансформація тимчасової шкали.

Алгоритм динамічної трансформації часової шкали дає змогу знайти мінімальну відстань між двома послідовностями різної довжини. Вперше застосований у розпізнаванні мови, де використаний визначення того, як два мовних сигналу представляють одну й ту саму вихідну промовлену фразу. Згодом було знайдено застосування і в інших областях [45,48].

Вимірювання будь-якої фізичної величини у часі широко поширений тип даних, що зустрічається, фактично, у будь-якій науковій галузі, і порівняння двох послідовностей є стандартним завданням. Для обчислення відхилення буває досить простого виміру відстані між компонентами двох послідовностей (Евклідова відстань). Однак часто дві послідовності мають різну тривалість. Щоб визначити подібність між такими послідовностями, ми повинні «деформувати» вісь часу однієї (або обох) послідовностей, щоб досягти кращого вирівнювання. [45,48]

Розглянемо алгоритм розрахунку цієї міри близькості.

Нехай дані дві послідовності \vec{x} і \vec{y} довжиною N і M відповідно:
 $\vec{x} = (x_1, x_2, \dots, x_N)$; $\vec{y} = (y_1, y_2, \dots, y_M)$.

Далі необхідно зробити розрахунок шляху найменшої збіжності. Визначимо матрицю відстаней $\Omega^{N \times M}$ так, щоб її елемент відповідав відстані між i -им і j -им елементами послідовностей \vec{x} і \vec{y} , тобто відповідав вирівнюванню між x_i і y_j . Як міру близькості використовується Евклідова відстань:

$$\varepsilon_{i,j} = \sqrt{(x_i - y_j)^2} \quad (2.44)$$

Далі використовуючи значення матриці відстаней $\Omega^{N \times M}$ сформуємо матрицю деформації $D^{N \times M}$ за значеннями якої буде розрахований шлях

трансформації послідовностей [94]. Причому елементи матриці деформації визначаються таким чином:

$$d_{i,j} = \begin{cases} \varepsilon_{i,j} + \min(d_{i,j-1}, d_{i-1,j}, d_{i-1,j-1}), & \text{если } i > 1, j > 1 \\ \varepsilon_{i,j}, & \text{если } i = 1, j = 1 \end{cases} \quad (2.45)$$

Трансформаційний шлях W визначається за формулою:

$$W = \{w_1, \dots, w_K\}, \quad (2.46)$$

де K – довжина шляху.

Визначимо k -ий елемент шляху трансформації W як елемент матриці $w_k = d_{i,j}$. Причому шлях трансформації повинен задовольняти низку умов:

Гранична умова: початок шляху W – перший елемент матриці деформації $w_1 = d_{1,1}$, а кінець шляху – останній, $w_K = d_{N,M}$. Таке обмеження гарантує, що трансформаційний шлях містить точки часових рядів.

Будь-які два суміжні елементи шляху, як умова безперервності W , $w_k = \{w_i, w_j\}$ и $w_{k+1} = \{w_i, w_j\}$ задовольняють наступним нерівностям $w_i - w_{i+1} \leq 1$, $w_j - w_{j+1} \leq 1$. Ця умова забезпечує обмеження один крок при виборі наступного елемента шляху.

Умова монотонності: будь-які два суміжні елементи шляху, $w_k = \{w_i, w_j\}$ и $w_{k-1} = \{w_{i-1}, w_{j-1}\}$ задовольняють наступним нерівностям $w_i - w_{i-1} \leq 0$, $w_j - w_{j-1} \leq 0$.

Існує велика кількість шляхів трансформації, що задовольняють всі вищевказані умови, проте нас цікавлять тільки той шлях, який мінімізує DTW відстань [94]. Шуканий шлях повинен задовольняти умові мінімальної вартості колії:

$$DTW(\vec{x}, \vec{y}) = \min \left\{ \frac{1}{K} \sqrt{\sum_{k=1}^K w_k} \right\} \quad (2.47)$$

Слід зазначити, що останнє розраховане значення d_{ij} матриці деформації є шуканою відстанню (заходом близькості) між послідовностями \vec{x} і \vec{y} .

Наведені заходи близькості використовуються для визначення відстані між ознаками об'єктів, що розпізнаються, надалі отримані значення відстані використовуються для прийняття рішення про належність порівнюваного об'єкта тому чи іншому класу

2.5. Висновки до розділу 2

У розділі розглянуто підходи до розпізнавання мовних сигналів, та методи їх аналізу які дають змогу відслідкувати як спектральні так і енергетичні характеристики сигналу.

РОЗДІ 3 НАУКОВО-ДОСЛІДНА ЧАСТИНА

3.1. Методика оцінки методів розпізнавання мовних сигналів

Оцінка методів розпізнавання мовних сигналів проводиться за допомогою наступної методики.

Дано безліч об'єктів X - сегменти мовних сигналів, набір дійсних відповідей Y , також є цільова функція $y^*: X \rightarrow Y$, в якій $y_i = y^*(x_i)$ на кінцевій підмножині об'єктів відомі значення $\{x_1, \dots, x_\ell\} \subset X$. Прецедентами називаються пари «об'єкт-відповідь» (x_i, y_i) . Навчальною вибіркою називається набір пар $X^\ell = (x_i, y_i)_{i=1}^\ell$. Завданням прецедентної ідентифікації є відновлення залежності від вибірки, тобто побудова вирішальної функції $\alpha: X \rightarrow Y$, яка б апроксимувала цільову функцію на всій множині X .

В якості навчальної вибірки будемо використовувати базу мовних даних, що складається з мовних сигналів записаних при промовленні: "нуль", "один", "два", "три", "чотири", "п'ять", "шість", "сім", "вісім", "дев'ять". Кожен відрізок МС є об'єктом $x_i \in X^\ell$ і є записом окремого слова-відповіді $y_i \in Y^\ell$. Таким чином, прецедентом (x_i, y_i) вважатимемо слово та його акустичну реалізацію.

Ознака f об'єкта x — результат вимірювання певної характеристики об'єкта. Формально ознакою називається відображення $f: X \rightarrow D_f$, де D_f - безліч допустимих значень ознаки. Нехай є набір ознак f_1, \dots, f_n . Вектор $f_1(x), \dots, f_n(x)$ називають ознаковим описом об'єкта $x \in X$.

Для кожного об'єкта в навчальній вибірці $x_i \in X^\ell$ проводилося обчислення певного вектора ознак $f_n(x)$ з різними параметрами (довжина вікна аналізу кількість частотних інтервалів). При цьому використовувалася наступна вибірка ознак із встановленими параметрами – таблиця 3.1.

Таблиця 3.1

Таблиця вибірки ознак

| Ознака | Параметри розрахунку | f_n |
|--|---|-------|
| Декомпозиція мовного сигналу | N=128, R= 8 N=256, R=16 N=512, R= 32 | f_1 |
| Розподіл миттєвих енергій мовного сигналу | N=128, R= 8 N=256, R=16 N=512, R= 32 | f_2 |
| Розподіл часток енергії мовного сигналу | N=128, R= 8 N=256, R=16 N=512, R= 32 | f_3 |
| Розподіл інформаційних інтервалів мовного сигналу | N=128, R= 8 N=256, R=16 N=512, R= 32 | f_4 |
| Мел-кепстральні коефіцієнти мовного сигналу | N=128, R= 8 N=256, R=16 N=512, R= 32 | f_5 |
| Частота переходів через нуль мовного сигналу | N=128 N=256 N=512 | f_6 |
| Ширина частотної області, яку займає мовний сигнал | N=128 N=256 N=512 | f_7 |

Надалі ми будемо розрізняти об'єкти з X та його ознакові описи, вважаючи $X = D_{f_1} \times \dots \times D_{f_n}$. При цьому значення m параметрів містяться у безлічі $P = \{p_1, \dots, p_m\}$.

Сукупність ознакових описів всіх об'єктів вибірки X_ℓ , записану у вигляді таблиці розміру $\ell \times m$, називають матрицею об'єктів - ознак:

$$F = \|f_n(x_i)\|_{\ell \times m} = \begin{pmatrix} f_1(x_1) & \dots & f_m(x_1) \\ \dots & \dots & \dots \\ f_1(x_\ell) & \dots & f_m(x_\ell) \end{pmatrix} \quad (3.1)$$

Для кожної розрахованої ознаки запишемо матрицю об'єктів – ознак.

$$F_n = \begin{pmatrix} f_1(x_1) & \dots & f_m(x_1) \\ \dots & \dots & \dots \\ f_1(x_\ell) & \dots & f_m(x_\ell) \end{pmatrix}$$

Кожен стовпець матриці містить значення ознаки, розраховані з певними

параметрами, наприклад, третій стовпець матриці F_1 містить значення ознаки: f_1 ; розраховані з параметрами: $N = 256$; $R=16$.

У разі розв'язання задачі ідентифікації кількість класів обмежується двома: $Y=\{0,1\}$ і алгоритм $\alpha(x)$ повинен відповідати на запитання «чи тотожні порівнювані об'єкти x_q і x_s ?». У цьому об'єкти x_q і $x_s \in X$. Отже, ми можемо визначити таку гіпотезу:

H_0 : порівнювані об'єкти x_q і x_s тотожно рівні.

Альтернативна гіпотеза визначається так:

H_1 : порівнювані об'єкти x_q та x_s не рівні.

Імовірнісна постановка завдання вважається більш загальною, тому що функціональну залежність $y^*(x)$ можна подати у вигляді імовірнісного розподілу $p(x, y) = p(x)p(y|x)$ [7], поклавши $p(y|x) = \delta(y - y^*(x))$, де $\delta(z)$ - Дельта-функція.

Для оцінки ефективності досліджуваних заходів близькості пропонується використовувати функціонал якості, який визначається через функцію втрат (ФВ).

ФВ – це інтегральна функція від $L(a,x)$, яка показує величину помилки алгоритму a на об'єкті x . Якщо $L(a,x)=0$, відповідь $a(x)$ називається коректним. [7]

Функціонал якості алгоритму a у вибірці X^ℓ (3.2):

$$Q(a, X^\ell) = \frac{1}{\ell} \sum_{i=1}^{\ell} L(a, x_i) \quad (3.2)$$

Функціонал Q називають також функціоналом середніх втрат або емпіричним ризиком [7], оскільки він обчислюється за емпіричними даними $(x_i, y_i)_{i=1}^{\ell}$.

У цьому дослідженні ми зазвичай використовуємо двійкову функцію втрат, яка приймає значення 0 і 1. Де $L(a,x)=1$ означатиме, що алгоритм a має помилку на об'єкті x . Помилкою швидкості алгоритму a на вибірці X^ℓ називається функція Q .

У дослідженні будемо використовувати наступний ідентифікатор функції втрат:

$$L(a, x) = [a(x) \neq y^*(x)] = 1 \quad (3.3)$$

Відповідно до прийнятої нульової гіпотезою можна розділити помилки, що виникають під час роботи алгоритму a на два типи - таблиця 3.2.

Таблиця 3.2

Таблиця типів помилок

| | | Нульова гіпотеза Порівнювані МС тотожні | |
|---------------------------------|------------------|--|-----------------------------------|
| | | <i>Вірна</i> | <i>Не вірна</i> |
| Результат застосування критерію | <i>Прийнята</i> | ідентичність підтверджено | ідентичність невірно підтверджена |
| | <i>Відкинута</i> | ідентичність невірно відкинута | ідентичність відкинута |

Дотримуючись інформації, викладеної в таблиці, можна описати помилки таким чином:

I) Помилками першого роду вважатимемо неправильно відкинуту тотожність.

II) Помилками другого роду вважатимемо неправильне підтвердження гіпотези.

Для підрахунку кількості помилок першого та другого роду необхідно розділити досліджувану навчальну вибірку на два непересічні підмножини:

$X^\ell \subset X^{\ell_1} \cup X^{\ell_2}$. Елементи підмножин можна записати так:

$X^{\ell_1} = \{x_1^{\ell_1}, \dots, x_\mu^{\ell_1}\}$ - об'єкти, у яких виникають помилки 1 роду.

$X^{\ell_2} = \{x_1^{\ell_2}, \dots, x_\eta^{\ell_2}\}$ - об'єкти, у яких виникають помилки 2 роду.

Застосування функції втрат (3.3) на кожне підмножина дасть кількість помилок, що дає алгоритм:

$$L(a(h), x^{\ell_1}) = \sum_{i=1}^{\mu} L_i[a(x_i^{\ell_1}) \neq y^*(x_i^{\ell_1})]$$

$$L(a(h), x^{\ell_2}) = \sum_{i=1}^{\eta} L_i[a(x_i^{\ell_2}) \neq y^*(x_i^{\ell_2})]$$

де h значення порога вирішальної функції.

Дані параметри необхідно використовувати при пошуку оптимального, з точки зору рівня значущості, порога h , для вирішального правила гіпотези, попередньо поставивши межу відносної кількості помилок одному з підмножин. Як алгоритми α будемо використовувати міри близькості, описані в першому розділі: 1) Евклідова відстань; 2) Середньоквадратичне відхилення; 3) Відстань Махаланобіса; 4) Кореляція; 5) Динамічна трансформація тимчасової шкали. Таким чином, ми можемо визначити найбільш адекватні стосовно вирішуваної задачі ознаки об'єктів. Варіації порівнюваних ознак та заходів близькості представлені у таблиці 3.3.

Таблиця 3.3

Варіації параметрів експерименту

| Ознака, F_n | Міра близькост, a |
|---|--|
| Декомпозиція мовного сигналу | Відстань Евклідова, Середньоквадратичне відхилення, Відстань Махаланобіса, Кореляція, Динамічна трансформація часової шкали |
| Розподіл миттєвих енергій мовного сигналу | |
| Розподіл часток енергії мовного сигналу | |
| Розподіл інформаційних інтервалів мовного сигналу | |
| Мел-кепстральні коефіцієнти | |
| Частота переходів через нуль | Відстань Махаланобіса, Динамічна трансформація тимчасової шкали |
| Ширина частотної області, що займає сигнал | |

3.2. Дослідження підходів до розпізнавання мовних сигналів

Для дослідження підходів до розпізнавання мовних сигналів запишемо

мовні сигнали при проголошенні наступних чисельних диктором чоловіком: "нуль", "один", "два", "три", "чотири", "п'ять", "шість", "сім", "вісім дев'ять". У кількості по 10 штук кожного чисельного. Параметри запису: частота дискретизації 16 кГц, глибина дискретизації 16 біт, моноканал.

Потрібно зробити експерименти, з метою оцінки ймовірності помилок I та II роду при використанні різних ознак та заходів близькості для їх розпізнавання.

Результати розрахунку мір близькості для наявної вибірки ознак можна як наступних двовимірних графіків (рисунок 3.1 – 3.5 графіки для мел-кепстральних коефіцієнтів с параметрами $R=16$, $N=256$). Порівняння об'єктів навчального підмножини відбувається за принципом "кожний з кожним", отже, елементи, що розташовуються на головній діагоналі, є результатом порівняння одного і того ж об'єкта x_s (самого з собою). За допомогою даних графіків можна візуально оцінити застосування того чи іншого ознакового опису об'єктів X.

Після розрахунку заходів близькості можна приступати до визначення порога вирішальної функції та оцінки ймовірностей помилок I та II роду.

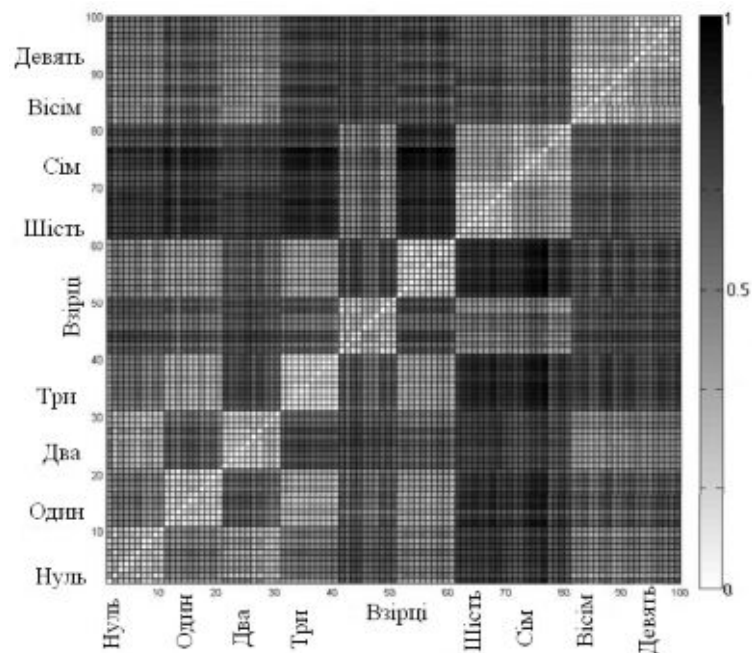


Рис. 3.1. Результат розрахунку динамічної трансформації тимчасової шкали для Мел-кепстральних коефіцієнтів мовних сигналів

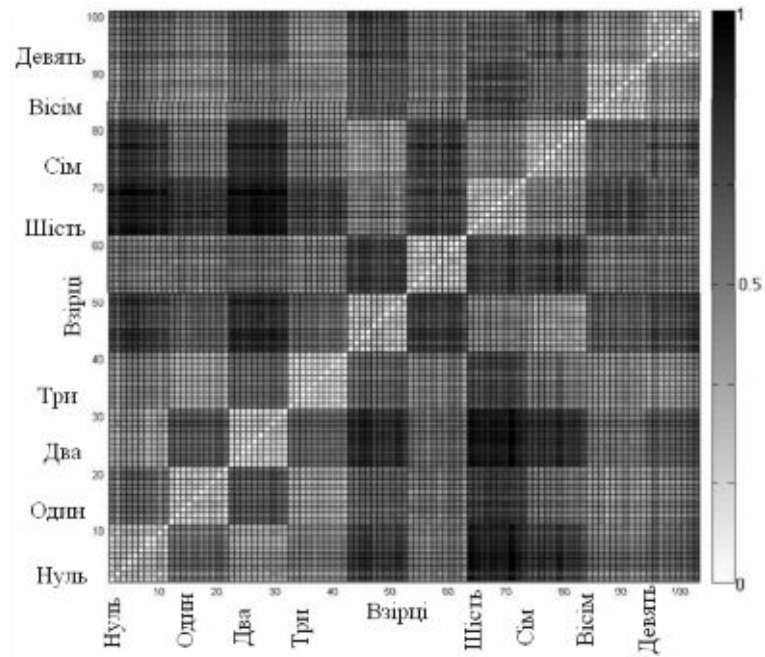


Рис. 3.2. Результат розрахунку динамічної трансформації тимчасової шкалі для Мел-кепстральних коефіцієнтів мовних сигналів

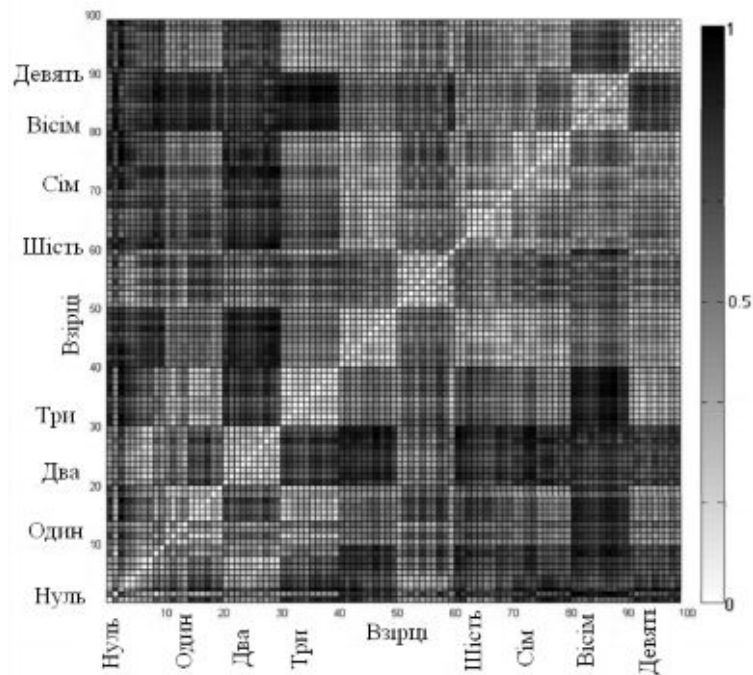


Рис. 3.3. Результат розрахунку відстані Махаланобісу для Мел-кепстральних коефіцієнтів мовних сигналів

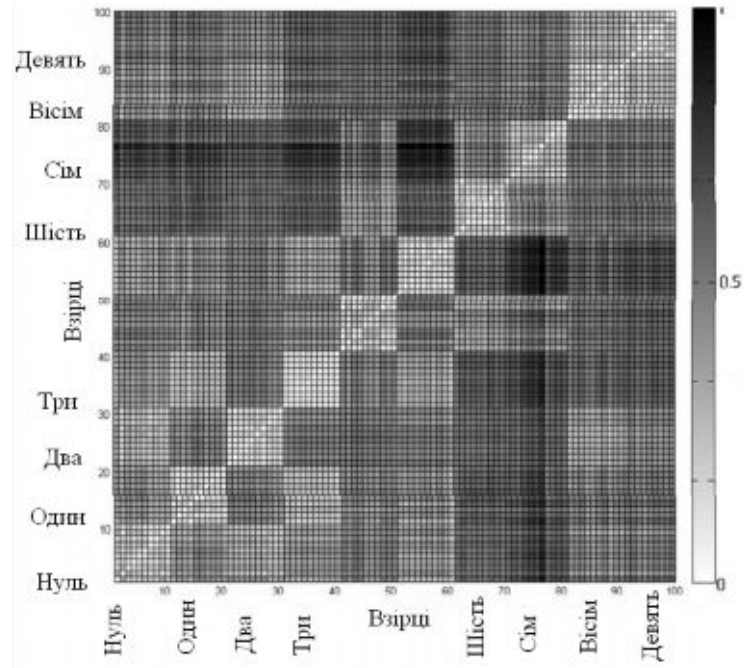


Рис. 3.4. Результат розрахунку Евклідова відстані для Мел-кепстральних коефіцієнтів мовних сигналів

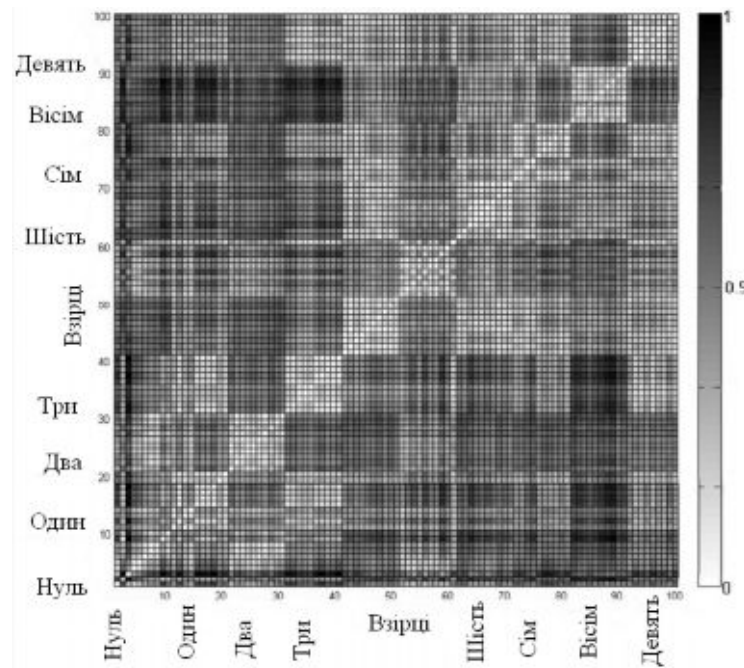


Рис. 3.5. Результат розрахунку кореляції для Мел-кепстральних коефіцієнтів мовних сигналів

Далі потрібно зробити підбір значень порога для вирішального критерію гіпотези H_0 .

Підбір порога для гіпотези H_0 здійснюватимемо наступним чином (3.4):

$$h = \min(L(a(h), x^{t_1}) + L(a(h), x^{t_2})) + \begin{cases} 99999, \mu / L(a(h), x^{t_1}) > \Delta \\ 0, \mu / L(a(h), x^{t_1}) \leq \Delta \end{cases} \quad (3.4)$$

де Δ - задана припустима можливість помилок I роду. Перебір порогових значень здійснюється у таких межах: $h = \{0, 0.01, 0.02, \dots, 1\}$.

При доборі порога (3.4) здійснюється обмеження ймовірності появи помилок I роду значенням $\Delta=0,05$. Модель встановлення порога з урахуванням оцінки ймовірностей помилок I та II роду проілюстровано малюнку 3.6. Поріг вирішальної функції виставляється з урахування допустимого значення ймовірності помилок I роду.

У разі неможливості встановити поріг при заданій ймовірності помилок I роду $\Delta=0,05$, перерахуємо поріг для $\Delta=0,1$. Якщо не вдається встановити поріг при значенні $\Delta=0,1$, перерахуємо поріг для $\Delta=0,25$. Таким чином, допустимі граничні значення ймовірності помилок I роду: $\Delta=\{0,05;0,1;0,25\}$

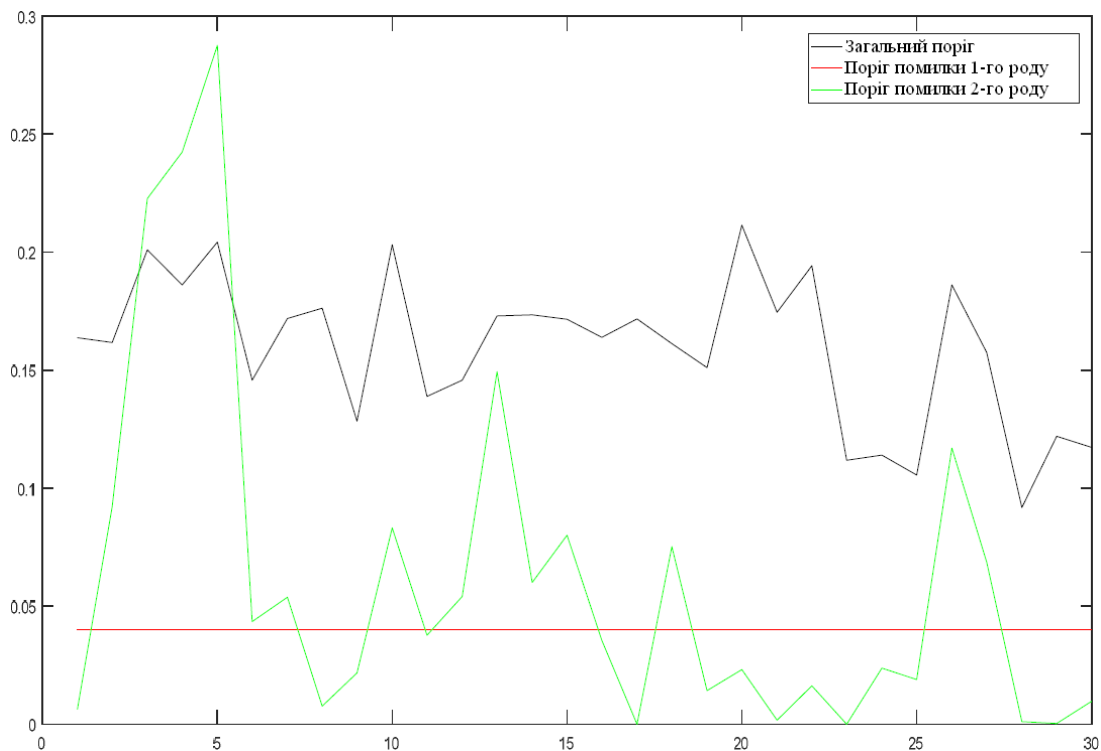


Рис. 3.6. Модель встановлення порога з урахуванням оцінки ймовірності помилок I та II роду

Згідно з описаним вище алгоритмом, проводився розрахунок функціоналу якості представлених алгоритмів (мір близькості) для різних ознак просторів. Основні результати даного експерименту представлені у таблицях 3.4 – 3.6. У таблицях представлені результати оцінки помилок I та II роду для різних заходів близькості (Евклідова відстань, середньоквадратичне відхилення, відстань Махаланобіса, кореляція, динамічне стиснення тимчасової послідовності) при їх використанні з ознаками: декомпозиція МС банком фільтрів; розподіл миттєвих енергій МС, розподіл часток енергії МС, розподіл інформаційних інтервалів МС, крейдянні коефіцієнти МС, частота переходів через нуль, ширина частотної області, що займається сигналом.

Таблиця 3.4

Результати оцінки помилок I та II роду, R=8, N=128

| Вид ознаки | Тип помилок | Заходи близькості, оцінка ймовірності помилок | | | | |
|--|-----------------|---|---------------------------|-----------------------|-----------|------|
| | | Євклідова відстань | Середньоквадр. відхилення | Відстань Махаланобіса | Кореляція | DTW |
| Декомпозиція мовного сигналу | Помилки I роду | 0,25 | 0,24 | 0,25 | 0,25 | 0,23 |
| | Помилки II роду | 0,48 | 0,40 | 0,46 | 0,42 | 0,39 |
| Розподіл миттєвих енергій МС | Помилки I роду | 0,24 | 0,22 | 0,24 | 0,24 | 0,19 |
| | Помилки II роду | 0,46 | 0,41 | 0,44 | 0,47 | 0,38 |
| Розподіл часткою енергії МС | Помилки I роду | 0,09 | 0,10 | 0,24 | 0,24 | 0,08 |
| | Помилки II роду | 0,40 | 0,41 | 0,39 | 0,41 | 0,36 |
| Розподіл інформаційних інтервалів МС | Помилки I роду | 0,05 | 0,05 | 0,10 | 0,10 | 0,04 |
| | Помилки II роду | 0,27 | 0,28 | 0,34 | 0,39 | 0,22 |
| Мел-кепстр коеф. МС | Помилки I роду | 0,05 | 0,04 | 0,10 | 0,10 | 0,04 |
| | Помилки II роду | 0,25 | 0,19 | 0,39 | 0,34 | 0,16 |
| Частота переходів через нуль | Помилки I роду | 0,25 | 0,24 | 0,25 | 0,25 | 0,21 |
| | Помилки II роду | 0,47 | 0,44 | 0,52 | 0,49 | 0,45 |
| Ширина частотної області, що займається сигналом | Помилки I роду | 0,24 | 0,24 | 0,25 | 0,25 | 0,24 |
| | Помилки II роду | 0,42 | 0,41 | 0,48 | 0,46 | 0,36 |

Результати оцінки помилок I та II роду, R=16, N=256

| Вид ознаки | Тип помилок | Заходи близькості, оцінка ймовірності помилок | | | | |
|--|-----------------|---|---------------------------|-----------------------|-----------|------|
| | | Євклідова відстань | Середньоквадр. відхилення | Відстань Махаланобіса | Кореляція | DTW |
| Декомпозиція мовного сигналу | Помилки I роду | 0,24 | 0,24 | 0,25 | 0,25 | 0,18 |
| | Помилки II роду | 0,46 | 0,39 | 0,42 | 0,37 | 0,36 |
| Розповсюджені миттєві енергії МС | Помилки I роду | 0,22 | 0,22 | 0,22 | 0,24 | 0,17 |
| | Помилки II роду | 0,41 | 0,44 | 0,45 | 0,46 | 0,32 |
| Розповсюджені частотою енергії МС | Помилки I роду | 0,08 | 0,07 | 0,21 | 0,22 | 0,08 |
| | Помилки II роду | 0,29 | 0,30 | 0,29 | 0,31 | 0,26 |
| Розповсюджені інформаційні інтервали МС | Помилки I роду | 0,05 | 0,04 | 0,09 | 0,10 | 0,05 |
| | Помилки II роду | 0,12 | 0,10 | 0,27 | 0,30 | 0,09 |
| Мел-кепстр коеф. МС | Помилки I роду | 0,05 | 0,04 | 0,08 | 0,08 | 0,04 |
| | Помилки II роду | 0,08 | 0,07 | 0,28 | 0,26 | 0,07 |
| Частота переходів через нуль | Помилки I роду | 0,22 | 0,23 | 0,24 | 0,25 | 0,20 |
| | Помилки II роду | 0,40 | 0,39 | 0,48 | 0,46 | 0,38 |
| Ширина частотної області зайнятої сигналом | Помилки I роду | 0,21 | 0,22 | 0,23 | 0,24 | 0,19 |
| | Помилки II роду | 0,39 | 0,36 | 0,41 | 0,42 | 0,37 |

Таблиця 3.6

Результати оцінки помилок I та II роду, R=32, N=512

| Вид ознаки | Тип помилок | Заходи близькості, оцінка ймовірності помилок | | | | |
|----------------------------------|-----------------|---|---------------------------|-----------------------|-----------|------|
| | | Євклідова відстань | Середньоквадр. відхилення | Відстань Махаланобіса | Кореляція | DTW |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Декомпозиція мовного сигналу | Помилки I роду | 0,25 | 0,24 | 0,24 | 0,24 | 0,24 |
| | Помилки II роду | 0,49 | 0,45 | 0,42 | 0,47 | 0,44 |
| Розповсюджені миттєві енергії МС | Помилки I роду | 0,25 | 0,23 | 0,22 | 0,23 | 0,21 |
| | Помилки II роду | 0,47 | 0,40 | 0,44 | 0,45 | 0,39 |

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|-----------------|------|------|------|------|------|
| Розпов. часткою енергії МС | Помилки I роду | 0,08 | 0,08 | 0,22 | 0,23 | 0,09 |
| | Помилки II роду | 0,35 | 0,34 | 0,38 | 0,37 | 0,32 |
| Розпов. інформаційне інтерв. МС | Помилки I роду | 0,05 | 0,05 | 0,09 | 0,10 | 0,04 |
| | Помилки II роду | 0,22 | 0,18 | 0,31 | 0,35 | 0,15 |
| Мел-кепстр коэф. МС | Помилки I роду | 0,05 | 0,04 | 0,10 | 0,10 | 0,04 |
| | Помилки II роду | 0,13 | 0,10 | 0,27 | 0,29 | 0,09 |
| Частота переходів через нуль | Помилки I роду | 0,23 | 0,21 | 0,22 | 0,24 | 0,20 |
| | Помилки II роду | 0,42 | 0,41 | 0,47 | 0,42 | 0,37 |
| Ширина частотної області займаемо. сигналом | Помилки I роду | 0,23 | 0,24 | 0,25 | 0,25 | 0,21 |
| | Помилки II роду | 0,39 | 0,35 | 0,41 | 0,40 | 0,32 |

Проаналізувавши отримані під час експериментів результати, можна зробити такі висновки:

I) Найбільш придатними для завдань розпізнавання векторами ознак можна вважати:

1. Мел-кепстральні коефіцієнти;
2. Розподіл інформаційних інтервалів мовного сигналу;

II) При оцінці даних векторів ознак найбільш придатними для вирішуваного завдання заходами близькості можна вважати:

1. Динамічна трансформація тимчасової шкали;
2. Середньоквадратичне відхилення;
3. Євклідова відстань.

Найменша ймовірність помилок I та II роду досягається при застосуванні як ознак Мел-кепстральні коефіцієнти. Як міра близькості найкращий результат показує DTW (динамічна трансформація тимчасової шкали).

Можна також стверджувати, що як вікно аналізу для отримання ознаки доцільно вибрати вікно тривалістю 256 відліків.

Для отримання найкращих результатів можна використовувати порівняння об'єкта з деяким усередненим значенням ознаки класу, т.к. відстань між усередненим значенням ознаки та окремими векторами ознаками буде меншою, ніж у разі порівняння строго кажучи унікальних об'єктів.

3.3. Висновки до розділу 3

В розділі проведено порівняльний аналіз застосування різних методів та їх інформативних ознак для мовних сигналів, зокрема, найбільш придатними для завдань розпізнавання ознак можна вважати мел-кепстральні коефіцієнти та розподіл інформаційних інтервалів мовного сигналу.

РОЗДІЛ 4

ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1. Охорона праці

Під час роботи з радіотехнічними системами враховано всі небезпечні фактори ризику (перевищений рівень шуму та вібрацій, електротравматизм, негативний вплив освітлення освітлення та інші), які би негативно впливали на рівень безпеки обслуговуючого персоналу в процесі експлуатації системи.

Оскільки система, живиться безпосередньо від електромережі, тому необхідно максимізувати рівень електробезпеки обслуговуючого персоналу шляхом адекватного дотримання правил роботи з електроприладами, зокрема системою, які прописані в стандарті ГОСТ 12.1.030-81 «ССБТ. Електробезпека. Захисне заземлення. Занулення» [38].

Із врахуванням вище сформульованого припущення, встановлена необхідність розроблення рекомендації по питанням охорони праці при роботі з радіотехнічною системою шляхом аналізу негативного впливу електричного струму на обслуговуючий персонал при роботі із системою, способів нормування та захисту від його дії.

Внаслідок дії електричного струму на організм обслуговуючого персоналу під час експлуатації блоку може виникнути загальна (електричний удар) або місцева електротравма (опіки, електричні знаки, електрометалізація шкіри, механічні пошкодження). Розрізняють три ступені впливу струму при проходженні через організм людини (змінний струм) [38]:

- відчутний струм – початок болісних відчуттів (до 0-1,5 мА);
- невідпускний струм – судоми і біль, важке дихання (10-15 мА);
- фібриляційний струм – фібриляція серця при тривалості дії струму 2-3с, параліч дихання (90-100 мА).

На рисунку 4.1 зображено основні фактори, які впливають на організм людини при ураженні електричним струмом.



Рис. 4.1. Фактори впливу електричного струму на людину

Правильне визначення необхідних засобів та заходів обслуговуючого персоналу від ураження електричним струмом необхідно враховувати гранично допустимі значення напруг дотику та струмів, що проходять через тіло людини по шляху "рука - рука" чи "рука - ноги" (таблиця 4.1) (регламентується ГОСТом 12.1.038-82).

Таблиця 4.1

**Гранично допустимі значення напруги дотику та сили струму,
що проходить через тіло людини**

| Вид струму | Нормоване значення | Тривалість струму, сек | | | | | |
|----------------|----------------------------------|------------------------|-----|-----|-----|----|----------|
| | | 0,1 | 0,2 | 0,5 | 0,7 | 1 | Більше 1 |
| Змінний, 50 Гц | Напруги дотику, В (не більше) | 500 | 250 | 100 | 70 | 50 | 36 |
| | Сила струму, мА (не більше) | 500 | 250 | 100 | 70 | 50 | 6 |

Основне завдання електробезпеки - мінімізувати можливість негативного впливу електричного струму на людину. Досягти цієї мети можна за допомогою таких заходів і засобів: 1) безпечною і надійною конструкцією елементів системи; 2) організаційними та технічними заходами щодо безпечної експлуатації системи та використання електричної енергії; 3) технічними засобами захисту.

У даному випадку це досягнуто шляхом конструктивного виконання складових системи класу I, який відповідає вимогам технічних умов і стандарту ГОСТ 12.1.030-81. Згідно класу I складові системи мають робочу ізоляцію і виконаний таким чином, що підключити його до електричної мережі можна лише після під'єднання корпусу до заземлювача (нульового захисного провідника), а при від'єднанні від мережі - корпус відключається від заземлювача (нульового захисного провідника) в останню чергу.

Стан ізоляції струмопровідних частин відповідає правилам використання системи. Цими правилами передбачене періодичне випробування ізоляції 2 рази на рік у приміщеннях зі складними умовами, підвищеною вологістю і 1 раз на рік у приміщеннях з нормальним середовищем. Ізоляція створює великий опір, який перешкоджає протіканню через неї струму. Опір ізоляції складових системи становить не меншим 0,5 МОм (згідно вимог ГОСТ 12.1.030-81). Якщо опір ізоляції знижується на 50% від початкового, мережу або ізоляцію необхідно замінити.

При роботі в приміщеннях без підвищеної небезпеки напруга складових системи повинна бути не більше 220 В. При роботі в приміщеннях з підвищеною небезпекою і за межами приміщень напруга складових системи повинна бути не більше 36 В. В особливих умовах дозволяється використовувати блок з напругою до 220 В, але при наявності захисного відключення або надійного заземлення корпусу з використанням захисних засобів (діелектричні рукавички, килимки, калоші).

Захисне заземлення - навмисне електричне з'єднання із землею металевих струмопровідних неструмоведучих частин, на яких може з'явитися напруга. Заземлення - це сукупність заземлювача і заземлювальних провідників. Заземлювачі можуть бути штучні (створені спеціально для заземлення блоку) і природні. Для штучних заземлювачів застосовують вертикальні і горизонтальні електроди. Вертикальні - зі сталевих прутів діаметром 10-12мм, кутової сталі розміром 40x40 мм або сталевих труб діаметром 30-50мм, довжиною 2,5-3 м. Вертикальні електроди з'єднують сталюю штабою розміром 4x12 мм або круглим дротом діаметром не менше 6 мм. Опір заземлюючого пристрою не

повинен перевищувати 4-10 Ом (перевіряється щорічно).

Таким чином врахувавши вище сформульовані рекомендації по питанням охорони праці при експлуатації радіотехнічної системи буде забезпечено небезпечні умови праці обслуговуючого персоналу.

4.2 Безпека в надзвичайних ситуаціях

Підприємство з випуску радіотехнічних систем (ДРС) є пожежонебезпечним, тому актуальним є забезпечення протипожежного захисту робітників та службовців, які на них працюють. Заходи протипожежного захисту здійснюються з дотриманням вимог глави 13 Кодексу цивільного захисту України від 02.10.2012 р. №5403-VI.

Всі заходи організаційно-технічного характеру протипожежного захисту на виробництві на об'єкті можна підрозділити на організаційні, технічні, режимні та експлуатаційні [36].

Забезпечення пожежної безпеки є складовою частиною виробничої або іншої діяльності посадових осіб, працівників підприємств та підприємців. Це повинно бути відображено у трудових договорах (контрактах) та статутах підприємств.

Керівник підприємства з випуску ДРС повинен визначити обов'язки посадових осіб щодо забезпечення пожежної безпеки, призначити відповідальних за пожежну безпеку окремих будівель, споруд, приміщень, діляниць, технологічного та інженерного устаткування, а також за утримання і експлуатацію технічних засобів протипожежного захисту. Обов'язки щодо забезпечення пожежної безпеки, утримання та експлуатації засобів протипожежного захисту мають бути відображені у відповідних посадових документах (функціональних обов'язках, інструкціях, положеннях тощо).

На кожному підприємстві з урахуванням його пожежної небезпеки наказом (інструкцією) повинен бути встановлений відповідний протипожежний режим, в тому числі визначені:

- можливість (місце) паління, застосування відкритого вогню та

побутових нагрівальних приладів;

- порядок проведення тимчасових пожежо небезпечних (в тому числі зварювальних) робіт;
- правила проїзду та стоянки транспортних засобів;
- місця для зберігання і допустима кількість сировини, напівфабрикатів та готової продукції, які можуть одночасно знаходитися у виробничих приміщеннях і на території (у місцях зберігання);
- порядок прибирання горючого пилю та відходів, зберігання промасленого спецодягу і шмаття, очищення повітроводів вентиляційних систем від горючих відкладень;
- порядок відключення від мережі електрообладнання у разі пожежі;
- порядок огляду і зачинення приміщень після закінчення роботи;
- порядок проходження посадовими особами навчання та перевірки знань з пожежної безпеки, а також проведення з працівниками протипожежних інструктажів та занять з пожежно-технічного мінімуму з призначенням відповідальних за їх проведення;
- порядок організації експлуатації і обслуговування наявних технічних засобів протипожежного захисту (протипожежного водопроводу, насосних станцій, вогнегасників тощо);
- дії працівників у разі виявлення пожежі.

Для об'єктів з перебуванням людей вночі інструкції мають передбачати два варіанти дій відповідно у денний та нічний час.

Усі працівники при прийнятті на роботу і за місцем здійснення професійної діяльності повинні проходити інструктаж з питань пожежної безпеки (вступний, первинний, повторний на робочому місці, позаплановий та цільовий). Посадові особи до початку виконання своїх обов'язків і періодично один раз на 3 роки мають проходити навчання і перевірку знань з питань пожежної безпеки.

Отже, організаційні заходи пожежної безпеки передбачають: організацію пожежної охорони на об'єкті, проведення навчань з питань пожежної безпеки (включаючи інструктажі та пожежно-технічні мінімуми), застосування наочних

засобів протипожежної пропаганди та агітації, проведення перевірок, оглядів стану пожежної безпеки приміщень, будівель, об'єкта в цілому та ін.

До технічних заходів належать: суворе дотримання правил і норм, визначених чинними нормативними документами при реконструкції приміщень, будівель та об'єктів, технічному переоснащенні виробництва, експлуатації чи можливому переобладнанні електромереж, опалення, вентиляції, освітлення і т. п.

Заходи режимного характеру передбачають заборону куріння та застосування відкритого вогню в недозволених місцях, недопущення появи сторонніх осіб у вибухонебезпечних приміщеннях чи об'єктах, регламентацію пожежної безпеки при проведенні вогневих робіт тощо.

Експлуатаційні заходи охоплюють своєчасне проведення профілактичних оглядів, випробувань, ремонтів технологічного та допоміжного устаткування, а також інженерного господарства (електромереж, електроустановок, опалення, вентиляції).

4.3 Висновки до розділу 4

У підрозділі з охорони праці розроблено рекомендації по питанням охорони праці при роботі з радіотехнічною системою шляхом аналізу негативного впливу електричного струму на обслуговуючий персонал при роботі із системою, способів нормування та захисту від його дії.

У підрозділі з безпеки в надзвичайних ситуаціях проаналізовано заходи організаційно-технічного характеру протипожежного захисту на виробництві радіотехнічної системи.

ВИСНОВКИ

У ході виконання випускної кваліфікаційної роботи було виявлено, що найбільш підходящими для задач розпізнавання ознаками є такі ознаки, що відображають властивості концентрації енергії та враховують особливості сприйняття слуху людини для отримання адекватної природі мовного впливу частотного розбиття. Даними ознаками є: крейдянні коефіцієнти та розподіл інформаційних інтервалів мовного сигналу.

Проведено аналіз особливостей обробки мовних сигналів у задачах розпізнавання мовлення: наведено концептуальні схеми розпізнавання, надано відомості про сприйняття та відтворення звуку людиною. Вивчено існуючі методи представлення мовних сигналів у завданнях розпізнавання: частотні (наприклад, частотний розподіл) та часові (наприклад, частота переходу сигналу через нуль) уявлення, проведено їх порівняльний аналіз. Для отримання деяких ознак (декомпозиція сигналу банку фільтрів, розподіл миттєвих енергій, розподіл часток енергії, розподіл інформаційних інтервалів) використовувався субсмуговий підхід, який дозволяє точно виділяти розподіл часток енергій по частотних інтервалах з мінімальним просочуванням.

Вивчено заходи близькості, що застосовуються для порівняння ознак: евклідова відстань, середньоквадратичне відхилення, відстань Махаланобісу, кореляція, динамічна трансформація тимчасової шкали. Найбільш підходящими для задач розпізнавання мовних сигналів є: динамічна трансформація часової шкали; середньоквадратичне відхилення; евклідова відстань.

Проведено порівняльний аналіз застосування різних заходів близькості та векторів ознак, який показав, що найбільш придатними для завдань розпізнавання векторами ознак можна вважати: Мел-кепстральні коефіцієнти; розподіл інформаційних інтервалів мовного сигналу, а їх порівняння використовувати такі заходи близькості: динамічну трансформацію часової шкали; середньоквадратичне відхилення; евклідова відстань.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аграновский, А.В. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов [Текст]/ А.В. Аграновский, Д.А. Леднов – М.: Радио и связь, 2004. – 164 с.
2. Алдошина, И.А. Слуховые модели восприятия линейных и нелинейных искажений в музыке и речи. Часть1 [Текст] / И.А. Алдошина // Информационно-технический журнал «Звукорежиссер». - 2006. - №3. – С.38-44.
3. Ананьев, Б. Г. Теория ощущений. — Л., 1961. — С. 579. — 928 с.
4. Вапник, В. Н. Теория распознавания образов [Текст] / Вапник В. Н., Червоненкис А. Я. / М.: Наука, 1974
5. Винцюк, Т.К., Анализ, распознавание и интерпретация речевых сигналов [Текст]/ Винцюк Т.К. - Киев: Наук.думка, 1987. - 264с.
6. Воеводин, В.В. Матрицы и вычисления [Текст] / В.В. Воеводин, Ю.А. Кузнецов. – М.: Наука, 1984. – 318 с.
7. Воронцов К.В. Математические методы обучения по прецедентам [Текст]/ материалы лекций МФТИ – М., 2008
8. Герасимов, А.В. Применение метода модифицированного линейного предсказания к задачам выделения акустических признаков речевых сигналов [Текст] / А.В.Герасимов, О.А. Морозов, В.Р. Фидельман // Радиотехника и Электроника. – 2005. – том 50. №10. – С. 1287-1292.
9. Гребнов, С.В. Аналитический обзор методов распознавания речи в системах голосового управления [Текст]/ С.В. Гребнов // Вестник ИГЭУ. – 2009. – Вып.3. – С.83-85.
10. Гривен, В.Г. Введение в Вейвлет преобразование / АВТЭКС, Санкт-Петербург, 2009, С. 302
11. Губочкин, И.В. Разработка алгоритмов анализа и распознавания речи на основе адаптивной кластерной модели и критерия минимального информационного рассогласования [Текст]: автореф. дис. канд. техн наук / И.В. Губочкин – Нижний Новгород: НГЛУ, 2011. – 22с.

12. Гудонавичюс, Р.В. Распознавание речевых сигналов по их структурным свойствам [Текст]/Р.В. Гудонавичюс, П.П. Кемешис, А.Б. Читавичюс – Л.: «Энергия», 1977. – 64 с.

13. Деркач, М.Ф. Динамические спектры речевых сигналов [Текст]/ М.Ф. Деркач, Р.Я. Гумецкий, Б.М. Гура, М.Е. Чабан – Львов: Виша школа. Изд-во при Львов. ун-те, 1983. – 168 с.

14. Жиляков, Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. [Текст] / Е.Г. Жиляков. – Белгород: Изд-во БелГУ, 2007. – 160 с.

15. Жиляков, Е.Г. Методы обработки речевых данных в информационно-телекоммуникационных системах на основе частотных представлений [Текст]/ Е.Г. Жиляков, С.П. Белов, Е.И. Прохоренко. – Белгород: Изд-во БелГУ, 2007. - 136 с.

16. Жиляков, Е.Г. Модели распределения энергии звуков русской речи на основе частотных представлений [Текст] / Е.Г. Жиляков, А.В. Болдышев, А.А. Фирсова// XXIII Международной научной конференции Математические методы в технике и технологиях – Саратов. – 2010. – С.236-239.

17. Жиляков, Е.Г. Частотный анализ речевых сигналов [Текст] / Е.Г. Жиляков, Е.И. Прохоренко // Научные ведомости Белгородского государственного университета. Сер. Информатика и прикладная математика – 2006. – №2(31), выпуск 3. – С.201-208.

18. Засыпкин, А.В. О дикторонезависимой системе голосового телефонного номеронабирателя [Текст] / А.В. Засыпкин, А.Т. Мицевич, М.В. Овецкий, В.Ю. Шелепов// Труды международной конференции “Знание-Диалог-Решение”. – Ялта. – 1995. – С.427-430.

19. Кавальчук, А.Н. (2011), "Формула для перехода из области частот к шкале барков и обратно," А.Н. Кавальчук, Ал.А. Петровский // Информатика, 2011, 4(32), стр. 71-81

20. Каганов, А.Ш. Криминалистическая экспертиза звукозаписей. – М.: "Юрлитинформ", 2005. - 272с.

21. Кипяткова И.С. Автоматическая обработка разговорной русской

речи: монография / И.С. Кипяткова, А.Л. Ронжин, А.А. Карпов. СПИИРАН – СПб.: ГУАП, 2013. – 314 с.

22. Колерс, П.А. Распознавание образов. Исследование живых и автоматических распознающих систем [Текст]/ П.А. Колерс, Е.Д. Мюррей, пер. Л.И. Титомира – М.: «Мир», 1970. – 288 с.

23. Ле, Н.В. Распознавание речи на основе искусственных нейронных сетей [Текст] / Н.В. Ле, Д.П. Панченко // Технические науки в России и за рубежом: материалы междунар. заоч. науч. конф.– Москва. – 2011. – С.8-11.

24. Леонович, А.А. Современные технологии распознавания речи [Текст] /А.А. Леонович // Материалы конференции «Диалог: Компьютерная лингвистика и интеллектуальные технологии». – Звенигород. – 2005.

25. Ли, У.А. Методы автоматического распознавания речи. [Текст] В 2-х книгах. Кн.1. / Пер. с англ./Под ред. У.Ли. – М.; Мир, 1983. –328 с.

26. Ли, У.А., Методы автоматического распознавания речи. [Текст] В 2-х книгах. Кн.2. /Пер. с англ. Под ред. У.Ли. – М.; Мир, 1983. – 392 с.

27. Мазуренко, И.Л. Компьютерные системы распознавания речи [Текст] / И.Л. Мазуренко // Интеллектуальные системы. – Москва. – 1998. – т.3. вып. 1-2. – С.117-134.

28. Мазуренко, И.Л. Одна модель распознавания речи [Текст] / И.Л. Мазуренко // Компьютерные аспекты в научных исследованиях и учебном процессе. – Москва – 1996 – С.107-112.

29. Малла, С. Вэйвлеты в обработке сигналов [Текст] / М.: Мир, 2005. — 672 с.

30. Ниценко, А.В. Алгоритмы пофонемного распознавания слов наперед заданного словаря [Текст] / А.В. Ниценко, В.Ю. Шелепов // Искусственный интеллект. – 2004. – С.633-639.

31. Оппенгейм А. В., Шафер Р. В. Цифровая обработка сигналов: Пер. с англ./Под ред. С. Я. Шаца. — М.: Связь, 1979. 416 с., ил.

32. Рабинер, Л. Теория и применение цифровой обработки сигналов [Текст] / Л.Рабинер, Б.Гоулд – М.: Мир, 1978. – 848с.

33. Рабинер, Л.Р. Цифровая обработка речевых сигналов [Текст] / Л.Р.

Рабинер, Р.Ф. Шафер – М.: Радио и связь, 1981. – 496 с.

34. Сергиенко А.Б. Цифровая обработка сигналов. 2 – изд. – СПб.: Питер, 2006. – 608 с.

35. Сорокин, В.Н. Артикуляторно-ориентированная система распознавания речи [текст] / В.Н. Сорокин, А.Н. Ижнин, А.И. Цыплихин, Д.Н. Чепелев // Труды Международного семинара «Диалог - 2003». – 2003. С.657-662.

36. Смоленцев, Н. К. Введение в теорию вейвлетов [Текст] /Ижевск: РХД, 2010. — 292 с.

37. Сорокин, В.Н. Сегментация и распознавание гласных [Текст] / В.Н. Сорокин, А.И. Цыплихин // Информационные процессы. – 2004. – Т.4, №2. – С. 202-220.

38. Сорокин, В.Н. Теория речеобразования [Текст] / В.Н. Сорокин – М.: Радио и связь, 1985. – 312 с.

39. Фланаган, Дж. Л. Анализ, синтез и восприятие речи [Текст]/ пер.с англ. А.А. Пирогова. – М.:Связь, 1968. – 397с.

40. Чистович, Л.А. Физиология речи. Восприятие речи человеком [Текст] /Л.А. Чистович, А.И. Венцов, М.П. Гранстрем и др. – М.: Наука, 1976. – 388 с.

41. Шелепов, В.Ю. К проблеме пофонемного распознавания [Текст] / В.Ю. Шелепов, А.В. Ниценко // Искусственный интеллект. – 2005. – №4. – С.662-668.

42. Шелухин, О.И. Цифровая обработка и передача речи [Текст] / О.И. Шелухин, Н.Ф.Лукиянецв; под ред. О.И. Шелухина. – М.: Радио и связь, 2000. – 456с.

43. Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. — Киев: Наукова думка, 2004.

44. Allen, J.B., "How Do Humans Process and Recognize Speech?," IEEE Trans. On Speech and Audio Processing, 1994, 2(4), pp. 567-577.

45. Al-Naymat Ghazi, Chawla Sanjay, Taheri Javid “Sparse DTW: A novel approach to speed up Dynamic Time Warping” Proc. of the 8th Australasian Data

Mining Conference (AusDM'09) p. 117-127; (2009)

46. Bishop, C. M. *Pattern Recognition and Machine Learning*. — Springer, Series: Information Science and Statistics, 2006. — 740 pp

47. Dong Yu *Automatic Speech Recognition: a deep learning approach (Signals and Communication Technology)* / Springer; 2015 edition (November 11, 2014), p. 321

48. Eamonn J. Keogh, Michael J. Pazzani *Derivative Dynamic Time Warping*, Section 1 *Proceedings of the sixth ACM SIGKDD*, 2010

49. Giannakopoulos T. *Introduction to Audio Analysis: A Matlab Approach* 1st Edition / Theodoros Giannakopoulos, Aggelos Pikrakis / Academic Press; 1 edition (April 21, 2014), p. 288

50. Huang X.D. *Spoken Language Processing: A Guide to Theory, Algorithm and System Development [Text]*/ Xuedong Huang, Alex Acero, Hsiao-Wuen Hon/ Prentice Hall PTC, New Jersey, 2001

51. Mahalanobis, Prasanta Chandra (1936). «On the generalised distance in statistics». *Proceedings of the National Institute of Sciences of India* 2 (1): 49–55.

52. Stevens, Stanley Smith; Volkman; John; & Newman, Edwin B. (1937). "A scale for the measurement of the psychological magnitude pitch". *Journal of the Acoustical Society of America* 8 (3): 185–190.

ДОДАТОК А
Копія тези конференції

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

МАТЕРІАЛИ

ІХ НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ
«ІНФОРМАЦІЙНІ МОДЕЛІ,
СИСТЕМИ ТА ТЕХНОЛОГІЇ»



8–9 грудня 2021 року

ТЕРНОПІЛЬ
2021

ЗМІСТ

СЕКЦІЯ 1. МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ

| | |
|---|----|
| Р.А. Бутій, С.А. Лупенко ПРИНЦИП КЕРУВАННЯ РОБОТИЗОВАНОЮ РУКОЮ ЗІ ЗВОРОТНИМ ЗВ'ЯЗКОМ ЗА ДОПОМОГОЮ НЕЙРОІНТЕРФЕЙСУ R.A. Butsiy, S.A. Lupenko THE PRINCIPLE OF CONTROLLING A ROBOTIC ARM WITH FEEDBACK VIA A NEUROINTERFACE | 3 |
| С.В. Венгер, М.І. Яворська ДОСЛІДЖЕННЯ ОСОБЛИВОСТЕЙ ПРОЦЕСУ ПЕРЕТВОРЕННЯ СИГНАЛУ В МІКРОСХЕМІ AD598 S.V. Venger, M.I. Yavorska INVESTIGATION OF THE FEATURES OF THE SIGNAL CONVERSION PROCESS IN THE AD598 CHIP | 4 |
| Н. Гашчин, Н. Круча, Г. Семенюшин ІНЖЕНЕРНА МЕТОДИКА РОЗРАХУНКУ НАГРІВУ ДИСКА N. Gashchyn, N. Krucha, H. Semenyushyn ENGINEERING METHOD OF CALCULATION OF DISC HEATING | 6 |
| А.Т. Гефко, М.В. Пшенничий, Т.С. Дубняк ОЦІНКИ МЕЖ ДЕФОРМАЦІЇ БАЛКИ ПРИ ВАРІАЦІЇ ЇЇ ПОПЕРЕЧНИХ РОЗМІРІВ МАТЕРІАЛУ І ПРИКЛАДЕНИХ ЗУСИЛЬ A.T. Hefko, M.V. Pshenychnyi, T.S. Dubyniak ESTIMATES OF LIMITS OF DEFORMATION OF A BEAM AT VARIATION OF ITS CROSS SIZES OF MATERIALS AND APPLIED EFFORTS | 8 |
| В. Дунець, Ю. Кутс, Н. Трач МОДЕЛЮВАННЯ РАДІОСИГНАЛУ ІЗ ФАЗОВОЮ МОДУЛЯЦІЄЮ ДЛЯ ОЦІНЮВАННЯ ЗАВАДОСТІЙКОСТІ ЗВ'ЯЗКУ V. Dunets, Yu. Kuts, N. Trach MODELING OF RADIO SIGNAL WITH PHASE MODULATION FOR ASSESSMENT OF COMMUNICATION DIFFICULTY | 10 |
| Р.М. Карабін, І.В. Литвинович ВИБІР АДЕКВАТНОЇ МОДЕЛІ НА ОСНОВІ ЗАСТОСУВАННЯ МЕТОДІВ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ R.M. Karabin, I.V. Lytvynenko CHOOSING AN ADEQUATE MODEL BASED ON APPLICATION OF DECISION-MAKING SUPPORT METHODS | 11 |
| А.Б. Кашчин, В.А. Невожай, М.І. Яворська ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ПРИСТРОЮ ДОСЛІДЖЕННЯ ПАРАМЕТРІВ ШОРСТКОСТІ ДЕТАЛЕЙ З ПЛОСКОЮ ПОВЕРХНЕЮ A.B. Kashchyn, V.A. Nevozhai, M. I. Yavorska INFORMATION SUPPORT FOR THE DEVICE USING THE PARAMETERS OF THE ROUGHNESS OF PARTS WITH A FLAT SURFACE | 12 |
| А. Марценюк, В. Андрійчук, А. Шипський МОДЕЛЮВАННЯ РАДІОСИГНАЛУ ІЗ АМПЛІТУДНОЮ МОДУЛЯЦІЄЮ ДЛЯ ОЦІНЮВАННЯ ЗАВАДОСТІЙКОСТІ ЗВ'ЯЗКУ A. Martsenyuk, B. Andriychuk, A. Shchipsky MODELING OF RADIO SIGNAL WITH AMPLITUDE MODULATION FOR ASSESSMENT OF COMMUNICATION DIFFICULTY | 14 |

УДК 621.376

А. Марценюк, Б. Андрейчук, А. Щіпський

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

МОДЕЛЮВАННЯ РАДІОСИГНАЛУ ІЗ АМПЛІТУДНОЮ МОДУЛЯЦІЄЮ ДЛЯ ОЦІНЮВАННЯ ЗАВАДОСТІЙКОСТІ ЗВ'ЯЗКУ

UDC 621.376

A. Martsenyuk, B. Andriychuk, A. Shchipsky

MODELING OF RADIO SIGNAL WITH AMPLITUDE MODULATION FOR ASSESSMENT OF COMMUNICATION DIFFICULTY

При створенні систем передачі даних в більшості випадків виявляється, що спектр початкового передаючого сигналу зосереджений зовсім не на тих частотах, які ефективно пропускає наявний канал зв'язку. Вирішення вказаної проблеми досягається шляхом використання модуляції. В процесі передачі даних по каналах із певним типом модуляції відбувається вплив на нього різного роду завад, що негативно впливає на його завадозахищеність і відповідно спотворення передаючих даних. Тому визначення завадостійкості каналу зв'язку як показника якості функціонування за призначенням без погіршення робочих характеристик під впливом різного роду завад є актуальною задачею.

В роботі представлено результати оцінювання каналу зв'язку із амплітудною модуляцією за блок-схемою дослідження, яку зображена на рисунку.



Застосовуючи програмне середовище Matlab здійснено процес амплітудної модуляції над синусоїдальним сигналом із різними значеннями несучої частоти. Для організації процесу оцінювання завадостійкості каналу зв'язку з амплітудною модуляцією при різних несучих частотах здійснено процес впливу завади типу білого шуму на амплітудно-модульовані реалізації шляхом її адитивного додавання.

Основним параметром завадостійкості каналу зв'язку є відношення сигнал-шум (SNR), яке обчислюється виразом:

$$q = 10 \log \left(\frac{E_{\text{сигнал}}}{A_{\text{бб вбб}}} \right), \text{ де } E_{\text{сигнал}} - \text{енергія}$$

сигналу, $E_{\text{вих шуму}}$ – енергія вихідного шуму.

Із отриманої залежності (відношення сигнал-шум (SNR)) встановлено, що при

збільшенні середньоквадратичного відхилення білого шуму відношення сигнал/шум як показник завадостійкості каналу зменшується із різними значенням частот несучого коливання.

Література.

1. Дунєць В. Л. Метод оптимального виявлення сигналів в каналах зв'язку / В.Л. Дунєць, Г.І. Цимбала, Р.В. Ракуш // Збірник тез доповідей V Міжнародної науково-технічної конференції молодих учених та студентів „Актуальні задачі сучасних технологій“, 17–18 листопада 2016 року. – Т. : ТНТУ, 2016. – Том II. – С. 37–38.