Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _інформаційних радіотехнологій та технічного захисту інформації_
(повна назва)

Кафедра __медіаінженерії та інформаційних радіоелектронних систем_____
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти _____другий (магістерський)_____

_____Методи оцінювання параметрів людського голосу. _____
_____
(тема)

Виконав:
студент _2__ курсу, групи ___МІм-22-1____
_____Маслій О.О. _____
(прізвище, ініціали)

Спеціальність 172 Телекомунікації та_____
_____радіотехніка_____
(код і повна назва спеціальності)
Тип програми ___освітньо-професійна_____
Освітня програма _____Медіаінженерія_____
(повна назва освітньої програми)

Керівник _____доц. Посошенко В.О._____
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ ___Володимир КАРТАШОВ__
(підпис)

2023 р.

Харківський національний університет радіоелектроніки

Факультет інформаційних радіотехнологій та технічного захисту інформації
Кафедра медіаінженерії та інформаційних радіоелектронних систем

Рівень вищої освіти другий (магістерський)

Спеціальність 172 Телекомунікації та радіотехніка
(код і повна назва)

Тип програми освітньо-професійна

Освітня програма Медіаінженерія
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«____»_____ 20 ___ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту Маслію Олексію Олексійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Методи оцінювання параметрів людського голосу.

затверджена наказом по університету від "_20_"_____10_____ 2023 р. № ___1224 Ст____

2. Термін подання студентом роботи до екзаменаційної комісії _____10.01.2024 р._____

3. Вихідні дані до роботи Розглянути моделі формування мовлення і методи оцінювання параметрів мовного тракту по звуковому сигналу. В середовищі MATLAB розробити програми для обчислення кепстру, спектрограми, і параметричних спектрів. Виконати експериментальне визначення формантних частот із використанням спектрального розкладання. Порівняти точність визначення формантних частот порівняно з існуючими формант-трекерами.

4. Перелік питань, що потрібно опрацювати в роботі
Вступ
1. Моделі мовоутворення та їх параметри. Постановка завдання аналізу мовлення за параметрами.
2. Оцінювання параметрів мовного тракту по звуковому сигналу.
3. Експериментальне визначення формантних частот із використанням спектрального розкладання мовного сигналу.
Висновки
Перелік посилань
Додатки

5. Перелік графічного матеріалу із зазначенням обов'язкових креслеників, схем, плакатів, комп'ютерних ілюстрацій

1. Постановка задачі (1 аркуш А4).
2. Модель формування мовлення (1 аркуш А4).
3. Непараметричні спектральні оцінки (1 аркуш А4).
4. Структурна схема обробки звукового сигналу (1 аркуш А4).
5. Параметричний метод спектрального оцінювання (1 аркуш А4).
6. Кепстральний аналіз (1 аркуш А4).
7. Спектрограми сигналів (1 аркуш А4).
8. Автоматичне виділення формант (1 аркуш А4).
9. Алгоритм динамічного виділення формантних частот (1 аркуш А4).
10. Результати дослідження (1 аркуш А4).
11. Висновки (1 аркуш А4).

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналітичний огляд літератури	01.09.23–27.09.23	
2	Теоретичний аналіз методів оцінки	28.09.23–11.10.23	
3	Підготовка аудіофрагментів	12.10.23–10.11.23	
4	Експериментальна частина	11.11.23–03.12.23	
5	Обробка результатів	04.12.23–17.12.23	
6	Графічна частина роботи	18.12.23–17.12.23	
7	Перевірка керівником	18.12.23–30.12.23	
8	Перевірка на академічний плагіат	02.01.24–05.01.24	
9	Перевірка завідувачем кафедри, рецензування	06.01.24–09.01.24	

Дата видачі завдання 20.10.2023 р.

Студент _____ Олексій МАСЛІЙ
(підпис)

Керівник роботи _____ Віталій ПОСОШЕНКО
(підпис)

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 62 сторінки, 29 рисунків, 21 джерело.

КЕПСТР, ЛІНІЙНЕ ПЕРЕДБАЧЕННЯ, МОВНИЙ СИГНАЛ, НЕПАРАМЕТРИЧНІ ОЦІНКИ, ПАРАМЕТРИЧНІ ОЦІНКИ, РОЗПІЗНАВАННЯ, СПЕКТРАЛЬНЕ ОЦІНЮВАННЯ

Метою кваліфікаційної роботи є теоретичний аналіз і практичне дослідження деяких методів оцінки параметрів голосу з метою визначення їх переваг і недоліків для різних застосувань.

В роботі розглянуті моделі формування мовлення – акустична і модуляційна. Розглянуті методи оцінювання параметрів мовного тракту по звуковому сигналу. В середовищі MATLAB розроблено ряд програм, що дозволяють обчислювати кепстри, спектрограми, і параметричні спектри. Ці програми були використані в експериментальних дослідженнях. Виконано експериментальне визначення формантних частот із використанням спектрального розкладання мовного сигналу. Порівняльний аналіз показує достатньо високу точність визначення формантних частот порівняно з існуючими формант-трекерами. Поряд із цим, необхідно зазначити простоту реалізації, низьку обчислювальну складність, швидкість і відповідність методу наявним фізичним процесам. Результати, отримані в даній роботі зможуть бути застосованими в лабораторному практикумі з дисципліни «Методи обробки звукової інформації».

ABSTRACT

Explanatory note to the qualification work: 62 pages, 29 figures, 21 sources.

CAPSTR, LINEAR PREDICTION, SPEECH SIGNAL, NON-PARAMETRIC ESTIMATES, PARAMETRIC ESTIMATES, RECOGNITION, SPECTRAL ESTIMATION

The purpose of the qualification work is theoretical analysis and practical research of some methods of evaluating voice parameters in order to determine their advantages and disadvantages for various applications.

The paper considers models of speech formation - acoustic and modulation. The methods of evaluating the parameters of the speech tract based on the sound signal are considered. In the MATLAB environment, a number of programs have been developed that allow you to calculate cepstra, spectrograms, and parametric spectra. These programs were used in experimental studies. Experimental determination of formant frequencies using spectral decomposition of the speech signal was performed. Comparative analysis shows sufficiently high accuracy of formant frequency determination compared to existing formant trackers. Along with this, it is necessary to note the simplicity of implementation, low computational complexity, speed and correspondence of the method to existing physical processes. The results obtained in this work can be applied in the laboratory workshop on the discipline "Methods of sound information processing".

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

АРКС – авторегресія з ковзанням середнього;

АЧХ – амплітудно-частотна характеристика;

АЦП – аналого-цифровий перетворювач;

ВВ – власні вектори;

ДПФ – дискретне перетворення Фур'є;

ЗС – звуковий сигнал;

КВП – критерій відношення правдоподібності;

КМ – коваріаційна матриця;

КФ – кореляційна функція;

МП – максимальна правдоподібність;

СГП – спектральна густина потужності;

ШПФ – швидке перетворення Фур'є;

ФВЧ – фільтр верхніх частот;

ФНЧ – фільтр нижніх частот;

FFT – швидке перетворення Фур'є;

SNR – відношення сигнал-шум;

SPL (Sound Pressure Level) – рівень звукового тиску;

STFT – короткочасне перетворення Фур'є.

ВСТУП

У сучасному світі все більше значення приділяють інтерфейсам, які використовують мовне введення і виведення для взаємодії між користувачем і комп'ютером. Тому розробник систем мовної інформації має надавати увагу все більшій кількості налаштувань і параметрів у підсистемах, що реалізують акустичний інтерфейс.

Задача розпізнавання мови (у багатьох своїх проявах: від сегментації промови до верифікації та ідентифікації особи) нині є вкрай актуальною. Свідченням цього є зростаюча кількість публікацій і конференцій із цієї тематики, а також відкриття в транснаціональних корпораціях департаментів, що орієнтовані на дослідження в мовній інформації.

Поліпшення існуючих систем розпізнавання мови дозволить істотно спростити взаємодію людини з комп'ютером у тому випадку, коли використання класичних інтерфейсів неможливо (наприклад, під час керування автомобілем або для людей з обмеженими фізичними можливостями), а також зробити подібну роботу більш комфортною та ефективною.

Необхідність досліджень із цієї тематики пояснюється незадовільними результатами існуючих систем при низькому співвідношенні сигнал/шум, залежностями результату від людини, а також невисокою швидкістю роботи подібного роду систем.

Дослідження методів визначення формант та розроблення нових точніших методів дозволить знизити похибку визначення та підвищити точність роботи систем ідентифікації мовної інформації.

Задача аналізу голосу людини виникає в багатьох застосуваннях. Найпоширеніша сучасна задача – це розпізнавання мови, тобто процес перетворення мовного сигналу у текстовий потік. Також на основі параметрів мови відбувається зворотна операція – синтез мовних сигналів.

Іншою задачею є аналіз голосового апарату людини. Це здійснюється в лікарняних цілях, або для виявлення і класифікації голосових патологій, або для аналізу стану вокального апарату співаків.

Також аналіз голосу застосовується в судовій криміналістиці для розпізнавання статі, віку, емоційного стану людини, а також особи розмовника.

Метою кваліфікаційної роботи є теоретичний аналіз і практичне дослідження деяких методів оцінки параметрів голосу з метою визначення їх переваг і недоліків для різних застосувань. Результати, отримані в даній роботі зможуть бути застосованими в лабораторному практикумі з дисципліни «Методи обробки звукової інформації».

1 МОДЕЛІ МОВОУТВОРЕННЯ ТА ЇХ ПАРАМЕТРИ. ПОСТАНОВКА ЗАВДАННЯ АНАЛІЗУ МОВЛЕННЯ ЗА ПАРАМЕТРАМИ

1.1 Поняття мови і мовоутворення

Звукова мова ґрунтується на звукових хвилях. Звуковими, або акустичними, хвилями називають слабкі механічні обурення, що розповсюджуються в пружному середовищі. Звукові хвилі, впливаючи на органи слуху, здатні викликати слухові відчуття [1].

При поширенні звуків у просторі слід враховувати такі особливості звукової хвилі:

– при віддаленні джерела звуку звукові коливання поступово загасають. Ослаблення звуку відбувається пропорційно квадрату відстані від джерела. Наприклад, до слухача, що знаходиться на відстані 5 м від того, хто говорить, доходить у 100 разів менше звукової енергії, ніж до слухача, що знаходиться на відстані 0,5 м;

– високочастотні звуки під час проходження через повітря поглинаються більшою мірою, ніж низькочастотні;

– при поширенні в повітрі звуків, що виходять із різних джерел одночасно (кілька розмовляючих знаходиться у різних частинах кімнати), відбувається накладання звукових хвиль;

– при розповсюдженні звуку в закритому приміщенні відбувається відображення звуку від стін та предметів, що знаходяться у цьому приміщенні. Це явище отримало назву реверберації. Необхідно враховувати можливість реверберації під час аудіозапису в закритих приміщеннях, особливо в домашніх умовах.

Основні характеристики звуку:

– частота;

– енергія;

– тривалість.

Від кількості енергії залежить інтенсивність звуку чи його сила. Основними характеристиками звуку є частота окремих складових та енергія. Частота коливальних рухів визначається їх числом в одиницю часу – Гц.

1.2 Особливості людського слуху

Людський слух сприймає частотний діапазон від 16 до 20 000 Гц при діапазоні енергії до 120 дБ. Але вухо людини найбільше чутливе до звуку, частота якого від 2000 Гц до 5000 Гц (рис. 1.1) [2].

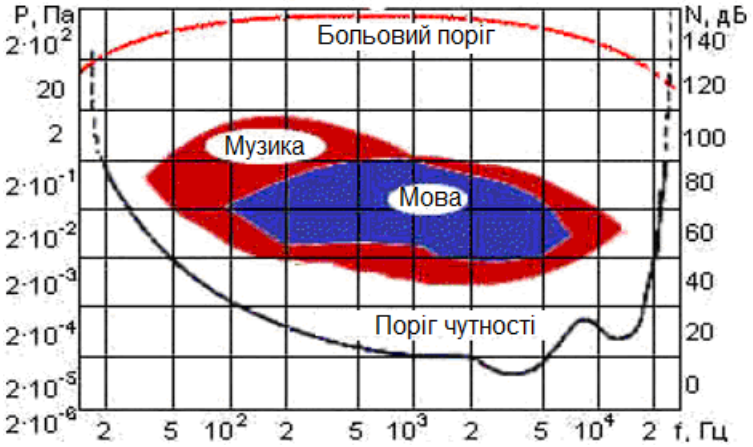


Рисунок 1.1 – Фізичні характеристики чутного діапазону

Нижня крива відповідає найслабшим звукам, які може чути людина; верхня – найгучнішим звукам, сприйняття яких викликає больове відчуття. Між цими кривими знаходиться діапазон чутних звуків. Виділені червоним кольором частини звукового діапазону представляють області, типові мови і музики.

Ефективність сприйняття мови залежить від її гучності. Гучність звуку є психоакустичним параметром і визначається здатністю людини оцінювати та визначати інтенсивність звуку суб'єктивними поняттями «тихо-голосно».

Одиниці вимірювання гучності, що використовуються як одиниця вимірювання в психоакустиці, отримали назву «фон».

Поріг чутності, який відповідає тону з частотою 1000 Гц, приймається рівним 0 дБ і називається стандартним або абсолютним порогом чутності.

Він відповідає звуковому тиску $p_0 = 2 \cdot 10^{-5}$ Па. Віжносно стандартного порога представляють інтенсивності всіх інших звуків діапазону, що сприймається.

Вухо вловлює звук, інтенсивність якого коливається від 0 дБ до 140 дБ. При цьому чутливість вуха до різних частот різна. Звуковий діапазон, значимий для промови, - 35-90 дБ.

У сприйнятті чутного мовлення важливу роль відіграють відділи слухової кори лівої півкулі головного мозку, що сприймають фонеми (тобто звуки). При поразці цих структур страждає фонематичний слух, що призводить до порушення розуміння слів. Фонема – мінімальна одиниця мовної системи. Фонемами називають звуки мови, заміщення яких змінює зміст слова. Фонетичні властивості мови – це властивості, що використовуються освіти звукової форми.

За допомогою фонем можна створити тисячі різних словоформ. Фонетична система сучасної української мови, включаючи літературні норми й деякі діалектні особливості, налічує 38 основних фонем: 6 голосних і 32 приголосних; додатково визначають 13 приголосних фонем у периферійній підсистемі. В англійській мові – від 40 до 45 фонем, залежно від діалекту [3].

Наприклад: слова «зов», «ров» відрізняються за першою фонемою; "бак", "бік" – по другій; "віл", "віз" – по третій.

Лише людина володіє системою звукових сигналів (фонемами), що у основі створення звукових одиниць, що мають мовне значення.

1.3 Процес мовоутворення

Мовоутворення – складний багатоступінчастий процес. У людському організмі немає спеціальних органів, призначених для вимови, вони виконують фізіологічні функції. Вимовними вони стали внаслідок багатовікової еволюції людини Умовно процес мовлення можна поділити на три фази.

Фаза моторного програмування – завдання вимовлення (вибирається спосіб досягнення необхідного стану мовних органів, координується дію м'язів, визначається тимчасова програма рухів).

Нейром'язова фаза – передача нервових імпульсів м'язовим волокнам, що веде до скорочення окремих м'язів чи його груп.

Двигуна фаза – виконання мовними органами певних рухів та прийняття положень, передбачених руховою програмою. Відбувається перехід від активності м'язів до активності органів (легкі, горло, язик тощо).

Мовний сигнал виникає внаслідок складної, координованої роботи апарату артикуляції.

Артикуляційний апарат або мовний тракт людини містить кілька основних компонентів, які забезпечують утворення мови.

Схематичне зображення мовленнєвого апарату людини [4] показано на рис. 1.1.

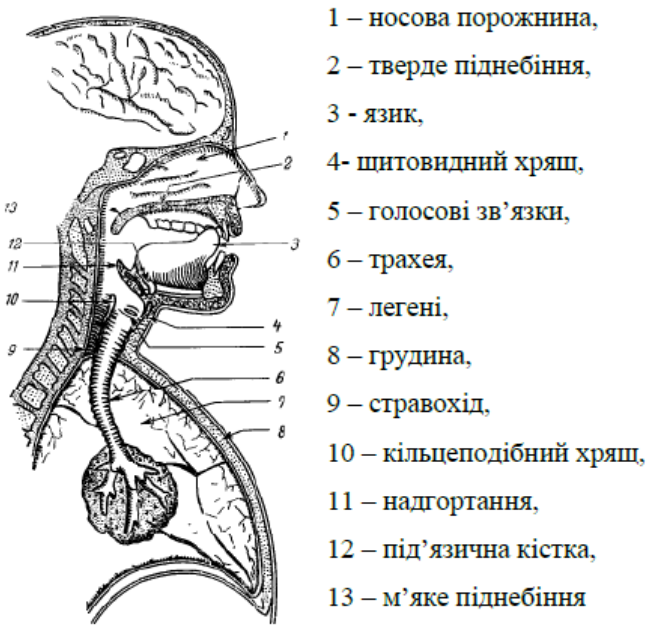


Рисунок 1.1 – Схематичне зображення мовленнєвого апарату людини

Мовленнєвий апарат умовно можна поділити на 3 частини:

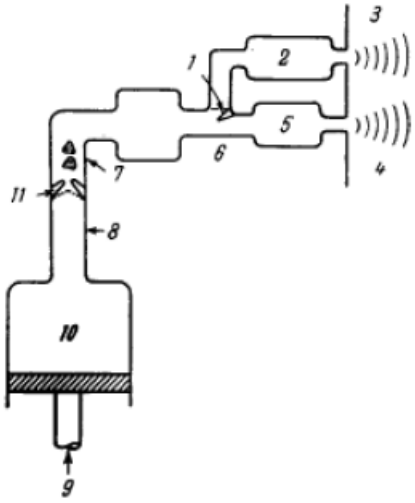
– енергетичний апарат (трахея, бронхи, легені, система м'язів) – забезпечує доставку повітряного струменя до місця фонації, тобто в порожнину рота;

– генератор звуку (горло) – поведінка гортані і голосових складок визначає характер звуку, що вимовляється;

– резонаторна частина (ковтка, порожнина носа, гайморові пазухи) – відповідно до теорії резонансного співу, виділяють сім функцій резонаторів: енергетична (посилення людського голосу та забезпечення здатності голосу поширюватися на довгі дистанції з меншими втратами сили звуку), генераторна (забезпечення звучання голосу з різними тембровими) фонетична (звукорозрізнення), естетична (надання голосу приємного забарвлення), захисна (посилення сили голосу), індикаторна (резонуючи в порожнинах дихального тракту, звук викликає у співака різні відчуття), активізуюча (функція рефлекторного підстроювання резонаторів);

– периферичний артикуляційний апарат (ротова порожнина, зуби, губи, тверде та м'яке піднебіння) – у процесі генерації звуку конфігурація та розмір повітроносних порожнин голосового тракту постійно змінюються, завдяки чому досягається відмінність у звучанні голосних звуків.

Схематичне зображення функціональних частин голосового тракту показано на рис. 1.2 [4].



1 – м'яке піднебіння,
2 - носова порожнина,
3 – випромінювання носа,
4 – випромінювання рота,
5 – ротова порожнина,
6 – піднята частина язика,
7 – гортанна трубка,
8 – трахея та бронхи,
9 - м'язова сила,
10 – обсяг легень,
11 – голосові зв'язки

Рисунок 1.2 – Схематичне зображення функціональних частин голосового тракту

Матеріалом, який забезпечує звукоутворення, є повітря. При голосоутворенні промовець управляє положенням хрящів, яких залежить

форма голосової щілини, і напруженістю голосових зв'язок. Різні зміни голосової щілини пропонують різні види коливань голосових складок, що зумовлює поява різних імпульсів.

Результатом вібрації є періодичний комплексний звук, що складається з основної частоти або частоти основного тону та кількох десятків (до 40) гармонік основної частоти (які називаються обертонами).

Обертони – ряд тонів, що виникають під час звучання основного тону та надають голосу певний відтінок або тембр.

Частота основного тону (ЧОТ) визначається Гц і на слух визначається як висота голосу. Значення ЧОТ у чоловіків – близько 80-150 Гц; у жінок – 200-400 Гц.

Зміни частоти основного тону в часі визначають інтонацію голосу – наголос, питання, оповідання, вигук і т.д.

1.4 Акустична модель мовоутворення

Одна з перших акустичних моделей організації звуків мови була запропонована німецьким лікарем, механіком та фізиком Християном Готтлібом Кратценштейном у 1779 р. Наукові основи акустичної теорії мовлення були закладені у працях німецького фізика, фізіолога та психолога Германа Людвіга Фердинанда фон Гельмгольця. У XX ст. основні положення акустичної теорії мовоутворення були сформульовані в роботах Г. Фанта, К. Н. Стевенса та Дж. Фланагана. Монографія Фанта «Акустична теорія мовоутворення» визнана класичним працею з теорії мовоутворення.

Відповідно до класичної (акустичної) теорії мовоутворення роль мовних органів у тому, щоб створити у мовному тракті аеродинамічні умови, необхідні освіти акустичних коливань. Умовно мовний тракт розглядають як акустичну трубу. Зовнішнім джерелом є потік повітря, що рухається. Акустичне обурення потоку здійснюється голосовими складками. Потік повітря, що коливається, проходячи мовним трактом, піддається

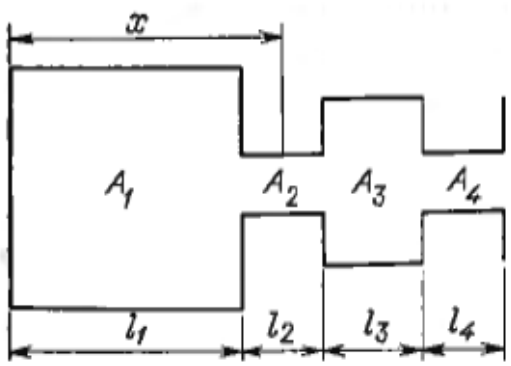
перетворенням, при цьому деякі частотні складові в спектрі джерела посилюються, а інші або залишаються незмінними, або пригнічуються. Тобто мовний тракт діє як акустичний резонатор.

Основні положення акустичної теорії мовоутворення [4]:

– процес мовлення складається з двох незалежних компонентів: із збудження звуку та формування фонетичної якості звуку за рахунок збудження резонансних частот артикуляційного тракту або фільтрації (у сучасній інтерпретації).

– фонетична якість звуку визначається формантами, які визначаються як резонансні частоти тракту артикуляції (полюса передавальної функції артикуляційного фільтра) або як максимуми спектра мовного сигналу.

Акустичну модель системи резонаторів голосового тракту представлено на рис. 1.3 [4].



A1 - гортань та задня ротова порожнина (до язика)
A2 – ділянка звуження між язиком та твердим піднебінням
A3 – передня ротова порожнина
A4 – прохід між губами

Рисунок 1.3 – Акустична модель системи резонаторів голосового тракту

Основне положення акустичної теорії мовлення: мовний сигнал виникає в результаті впливу одного або декількох джерел звуку на систему резонаторів, що утворюються повітряними порожнинами мовного тракту.

Властивості джерела звуку та резонаторної системи не є незмінними. Класична (акустична) теорія мовотворення передбачає незалежність роботи джерела та фільтра, що формує формантну структуру. Вимова звуків і звукових послідовностей є складним динамічним процесом, що змінюється в часі.

Електричний еквівалент акустичної моделі голосового тракту показано на рис. 1.4 [4].

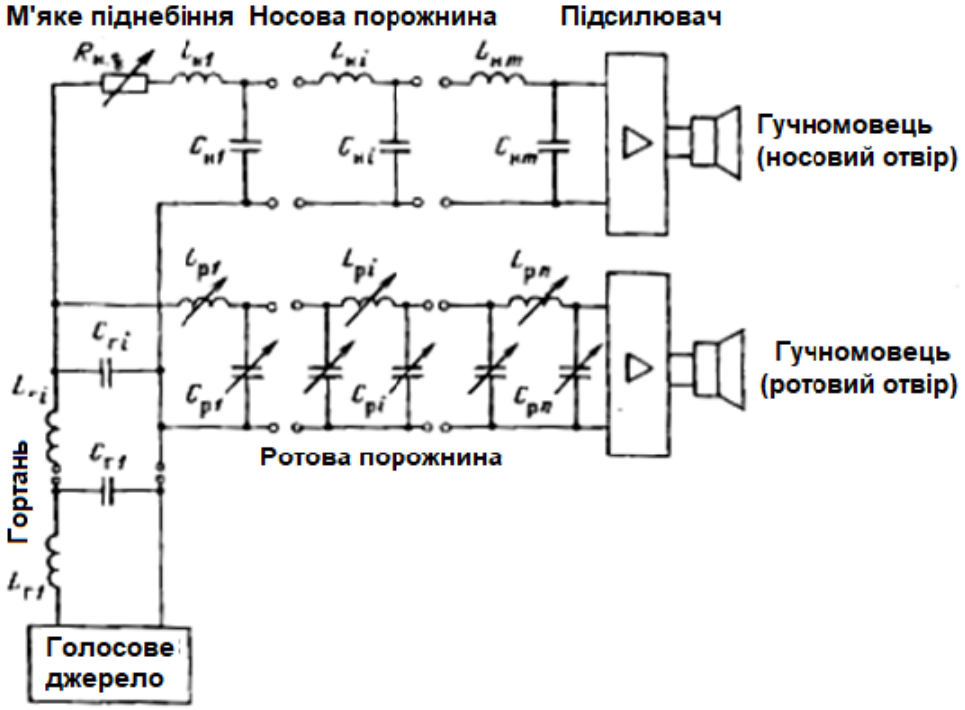


Рисунок 1.4 – Електричний еквівалент акустичної моделі голосового тракту

Основні особливості моделі Г. Л. Ф. Гельмгольця:

– положення про незалежність джерела звуку та артикуляційного фільтра. Гельмгольц показав, що фонетична якість голосних значною мірою сформована вже в гортані за винятком впливу артикуляційного фільтра;

– положення про визначальне значення формант (максимумів у спектрі звуків) визначення фонетичного якості звуків промови.

АЧХ однотрубної моделі голосового апарату для труби перетином 5 см^2 та довжиною 17,5 см показано на рис. 1.5, а розподіл ймовірностей формантних частот реальних чоловічих голосів показано на рис. 1.6 [4].

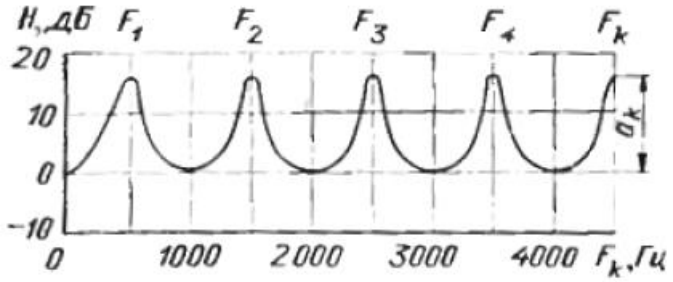


Рисунок 1.5 – АЧХ однотрубної моделі голосового апарату для труби перетином 5 см^2 та довжиною 17,5 см

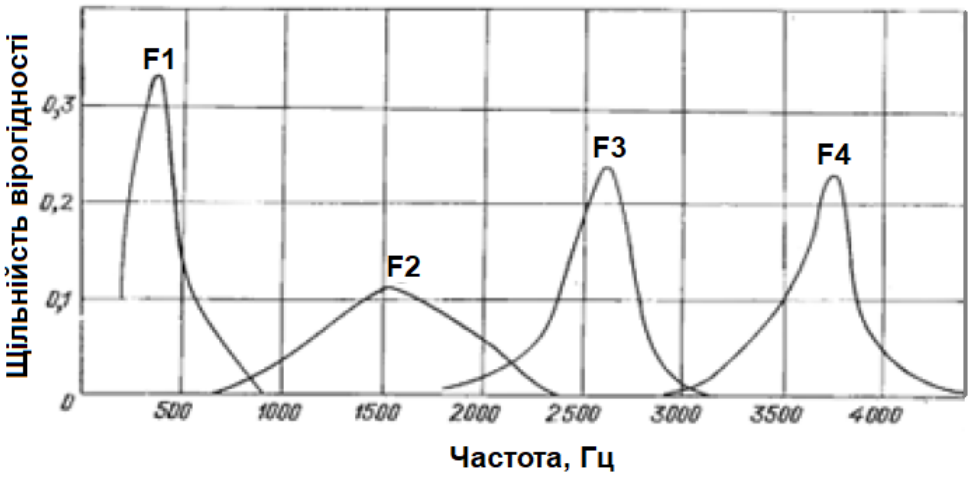


Рисунок 1.6 – Розподіл ймовірностей формантних частот реальних чоловічих голосів

Не викликає сумніву факт, що за допомогою формантів можна домогтися певної якості звучання. Але не тільки ці максимуми спектру визначають якість.

1.5 Модуляційна модель мовоутворення

Перші сумніви у спроможності акустичної теорії речеобразования виникли ще 1930-х гг. після створення смугового вокодера. На початку 1960-х років. на основі великого експериментального матеріалу була розроблена теорія розрахунку розбірливості мови, що базується на смуговому поданні мовного сигналу. У ньому не розглядалися форманти. Була висловлена гіпотеза про те, що фонетична якість звуків обумовлюється певним рівнем співвідношень потужності в спектральних смугах, а форманти є способом досягнення необхідних смугових співвідношень.

Відомо кілька фактів, які важко пояснити за допомогою класичної моделі мовлення Г. Л. Ф. Гельмгольця [5]:

– висока розбірливість мови в комунікаційних каналах з нестійкими амплітудно-частотними характеристиками (наприклад, телефонних каналах);

– висока розбірливість мови щодо впливу широкому діапазоні перешкод, шумів;

– розбірливість сигналу на виході з гортані навіть після видалення частин каналу артикуляції (в результаті хірургічної операції);

– синтез мовного сигналу пов'язаний зі специфікою слухового сприйняття.

Згідно з класичною схемою звукова хвиля формується в мовному тракті за рахунок порушення власних коливань звукового хвилеводу, яким вважається мовний тракт. З відкритого кінця хвилеводу хвилі поширюються в пружному середовищі і збуджують власні коливання хвилеводу, що представляє слуховий тракт. Ці коливання аналізуються мозком і сприймаються як мова. Така схема може існувати.

Однак такі уявлення погано узгоджуються з реальністю. По-перше, реальні спектри тих самих сигналів від різних індукторів можуть бути різними, що суперечить традиційній схемі. По-друге, при проходженні мовного сигналу через телефонний кабель його спектр значно спотворюється, але немає спотворення змістового змісту.

Яку роль грають значні просторові деформації мовного тракту процесі речеобразования? Чим зумовлена потреба в такій складній системі?

Відповіді ці питання дає модуляційна модель мовоутворення.

Згідно модуляційної моделі мовлення складається з двох етапів [6]:

– на першому етапі виникає звукова хвиля, яка не містить у собі жодної інформації та відіграє роль несійної;

– на другому етапі – несійна модулюється, і ця модуляція містить всю інформацію про промови.

Найбільш поширені модуляційні моделі мовного сигналу представляють мовний сигнал у вигляді:

– амплітудно-модульованого (АМ) коливання (рис.1.7, а) [1,5];

– частотно-модульованого коливання (рис.1.7, б) [6-8];

– коливання із амплітудно-частотною модуляцією [6].

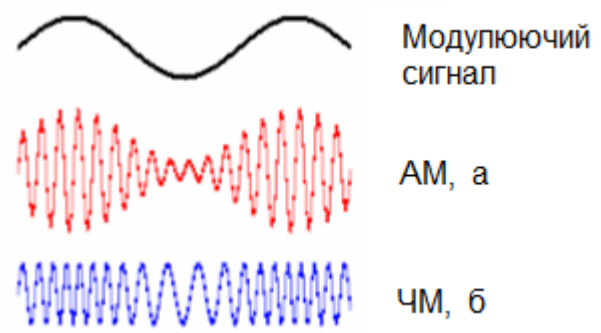


Рисунок 1.7 – Приклади модуляції сигналу несійної

У модуляційної моделі мовоутворення роль мовного тракту зводиться до ролі модулятора, і тому стають природними значні просторові деформації мовного тракту.

1.6 Постановка задачі аналізу голосу

1.6.1 Аналіз голосу в задачах розпізнавання мови

Задача аналізу голосу людини виникає в багатьох застосуваннях. Найпоширеніше сучасне застосування – це розпізнавання мови.

Розпізнавання мовлення - це процес перетворення мовного сигналу текстовий потік. Голосове керування – спосіб взаємодії з пристроєм за допомогою голосу. На відміну від розпізнавання мови, голосове управління призначене для введення команд - наприклад, "ввімкнути світло", "показати погоду на завтра", "вимкнути телевізор" тощо. Однак у будь-якому випадку необхідно виконати перетворення сигналу, який формується мікрофоном у слово або набір слів.

Для створення системи розпізнавання мови (рис.1.8) необхідно [7]:

– по-перше, проаналізувати спектральний склад сигналу, виділяючи з сигналу набір основних частот та амплітуд (виконати перетворення Фур'є);

– потім виділити з оцифрованого сигналу лінгвістичні конструкції (наприклад, фонеми), застосувавши різні математичні методи;

– після цього перетворити виділені лінгвістичні одиниці і конструкції в текстовий формат.

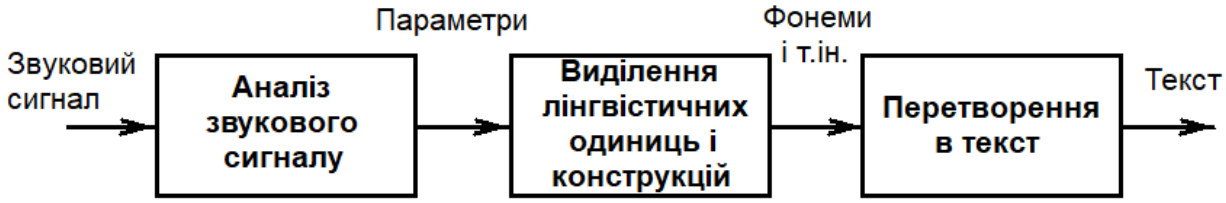


Рисунок 1.8 – Етапи задачі розпізнавання мовлення

При обробці мовленнєвих сигналів важливим є поняття «фонема», яке відрізняється від поняття «літера», оскільки враховує особливості звучання даного звуку в оточенні наступного та попереднього звуків. Крім того, для голосних звуків поняття «фонема» враховує, чи є даний звук наголошеним.

Фахівці в області лінгвістики нараховують від 50 до 150 фонем для певної мови. Конкретна кількість такого переліку залежить від авторської точки зору.

Приклад фонемного словника української мови із 58 фонем наведено на рис.1.9.

а, А, о, О, у, У, і, І, и, И, е, Е, б, бь,
в, вь, г, гь, ґ, ґь, д, дь, ж, жь, з, зь,
й, к, кь, л, ль, м, мь, н, нь, п, пь, р,
рь, с, сь, т, ть, ф, фь, х, хь, ц, ць, ч,
чь, ш, шь, дз, дзь, дж, джь, sil

Рисунок 1.9 – Приклад фонемного словника української мови

Для української мови найважливішими є перші дві форманти (резонансні частоти голосового тракту) голосного або приголосного звуку. Положення цих двох формант на осі частот дозволяє побудувати так зване «фонемне поле» голосних звуків (рис.1.10).

Відсутність подібного «формантного поля» для приголосних звуків пояснюється більш складним характером приголосних звуків, які можуть бути вокалізованими або невокалізованими, можуть мати круті передні фронти в часовій області та інше [8].

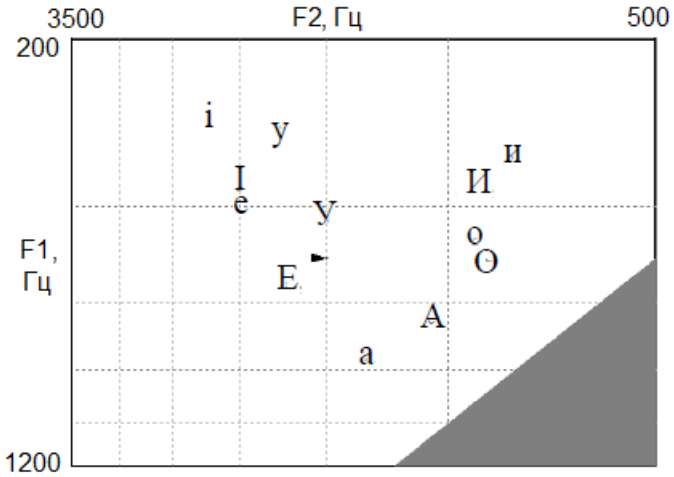


Рисунок 1.10 – Фонемне поле голосних звуків

Звуки мовлення характеризуються як частотними параметрами, такими як частоти формант, частота основного тону, шумоподібність спектру, кількість та виразність обертонів, так і часовими параметрами обвідної сигналу.

Зручним інструментом для відображення часо-частотних властивостей звуків мовлення є спектрограма (рис. 1.11) [9].

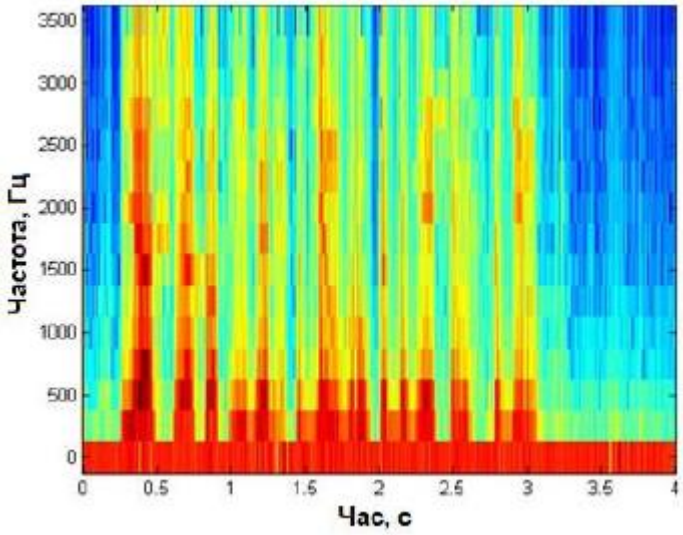


Рисунок 1.11 – Спектрограма мовлення

Спектрограму зручно використовувати, наприклад, для фонемної розмітки мовленнєвих сигналів при підготовці їх до етапу навчання систем автоматичного розпізнавання (рис.1.12) [10].

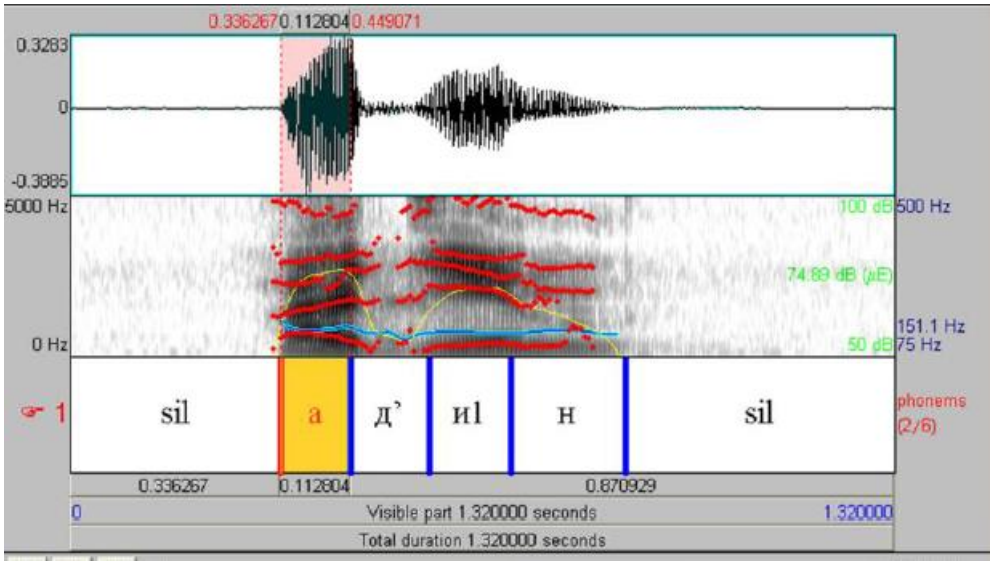


Рисунок 1.12 – Розмітка фонем

Розпізнавання та синтез мовлення – це не єдині задачі, що потребують аналізу голосу.

1.6.2 Аналіз голосу в задачах аналізу стану вокально-голосового апарату людини

При спектрально-часовому аналізі стану вокального апарату людини зручно розділяти часову та спектральну області. Спектрограму вокального сигналу показано на рис. 1.13.

На рис. 1.14 показано скріншот програмно-апаратного застосунку, що використовується в Київському інституті отоларингології ім. проф. О.С. Коломійченка. В спектральній області (верхня частина рис. 1.14) добре видно пік на частоті основного тону, піки на кратних частотах (обертони), а також наявність та розвиненість шумової складової вокального сигналу, яка може свідчити або про своєрідність тембру співака, або про наявність порушень голосового апарату.

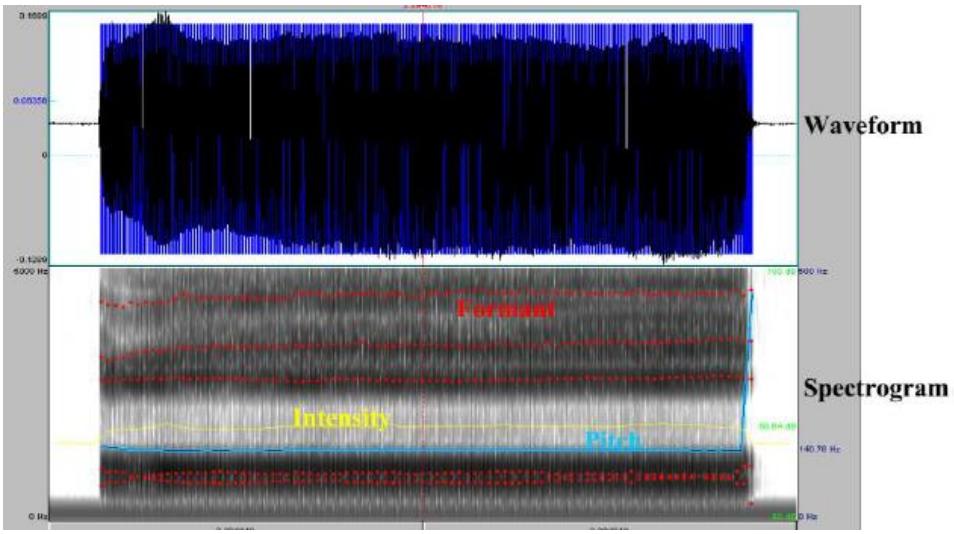


Рисунок 1.13 – Спектрограма вокального сигналу

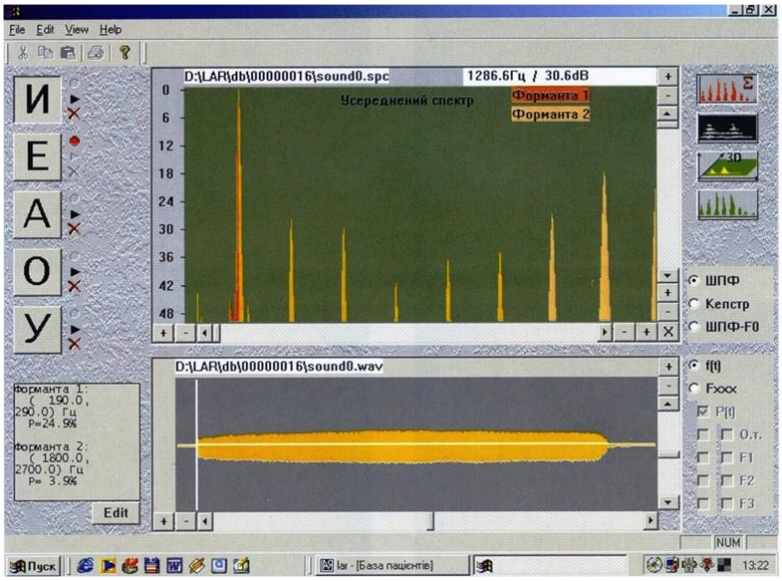


Рисунок 1.14 – Програмно-апаратний застосунок діагностики стану голосового апарату людини [2]

Важливою характеристикою стану голосового апарату співака є така часова характеристики як максимальна тривалість звуку, взятого на одному диханні (нижня частина рис. 1.6).

Іншими діагностичними параметрами є Shimmer та Jitter, що характеризують стабільність частоти основного тону та стабільність обвідної звуку (рис. 1.15 [11]), й таким чином свідчать про досконалість вокальної техніки.

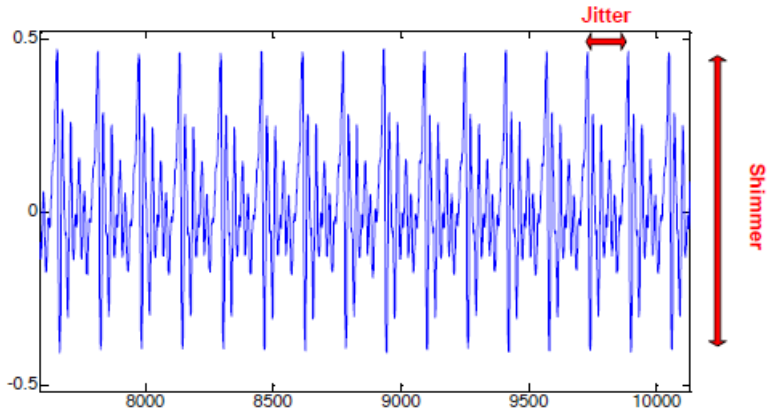


Рисунок 1.15 – Параметри Shimmer та Jitter як міри стабільності вокального звуку

Особливості спектрів голосних та приголосних звуків також дозволяють виділяти діагностичні ознаки. Наприклад, на рис. 1.14 показано спектр голосного звуку здорової людини. В [4] наведено спектр людини із захворюванням голосового апарату, де до дискретної структури додається потужна фонова структура, спричинена шумовим компонентом голосового сигналу.

1.7 Висновки по розділу 1

Задача аналізу голосу людини виникає в багатьох застосуваннях. Найпоширеніша сучасна задача – це розпізнавання мови, тобто процес перетворення мовного сигналу у текстовий потік. Також на основі параметрів мови відбувається зворотна операція – синтез мовних сигналів.

Іншою задачею є аналіз голосового апарату людини. Це здійснюється в лікарняних цілях, або для виявлення і класифікації голосових патологій, або для аналізу стану вокального апарату співаків.

Також аналіз голосу застосовується в судовій криміналістиці для розпізнавання статі, віку, емоційного стану людини, а також особи розмовника.

Розглянуті моделі формування мовлення – акустична і модуляційна. Акустична модель містить три основні складові: енергетичну, що генерує потік повітря, резонаторну частину, що формує частотну характеристику звуку, і артикуляційну частину, що формує обвідну звуку. Аналізуючи звук, можна отримати параметри моделі, що дозволить вирішувати багато прикладних задач.

Метою кваліфікаційної роботи є теоретичний аналіз і практичне дослідження деяких методів оцінки параметрів голосу з метою визначення їх переваг і недоліків для різних застосувань. Результати, отримані в даній роботі зможуть бути застосованими в лабораторному практикумі з дисципліни «Методи обробки звукової інформації».

2 ОЦІНЮВАННЯ ПАРАМЕТРІВ МОВНОГО ТРАКТУ ПО ЗВУКОВОМУ СИГНАЛУ

2.1 Непараметричне спектральне оцінювання

Сукупність синусоїдальних складових складного звуку, заданих за допомогою амплітуд та частот цих складових представляють акустичний спектр. Для спектрального аналізу сигналу використовується дискретне перетворення Фур'є (ДПФ) та швидке перетворення Фур'є (БПФ), яке представляє процедуру прискореного ДПФ.

Розглянемо кінцевий ряд дискретних сигналів $s(mT)$ при $m = 0,1,2...M-1$. Функція $S(K)$, яка визначається за формулою (2.1) називається дискретним перетворенням Фур'є для $s(mT)$ [12]:

$$S(K) = \sum_{m=0}^{M-1} s(mT) \exp\left(-2\pi j \frac{Km}{M}\right), \tag{2.1}$$

де $K = 0,1,2...M-1$.

Якщо знайдено ДПФ, можна відновити вихідний сигнал (зворотне перетворення Фур'є) по дискретним значенням сигналу.

Відповідно до теореми Котельникова, довільний сигнал, спектр якого не містить частот вище $F_в$ Гц, може бути повністю відновлений, якщо відомі відлікові значення цього сигналу, взяті через рівні проміжки часу дискретизації $T = 1/(2F_в)$.

Зворотне перетворення Фур'є визначається за формулою [12]

$$s(mT) = \frac{1}{M} \sum_{m=0}^{M-1} S(K) \exp\left(2\pi j \frac{Km}{M}\right), \tag{2.2}$$

де $m = 0,1,2...M-1$.

Реальний мовний сигнал має кінцеву тривалість, при поданні в частотній області спектр необмежений. Тому сигнал сегментують на ділянки близько 10 мс, на яких вважається стаціонарним.

Один із варіантів попередньої обробки мовного сигналу наведено на рис. 2.1 [12].

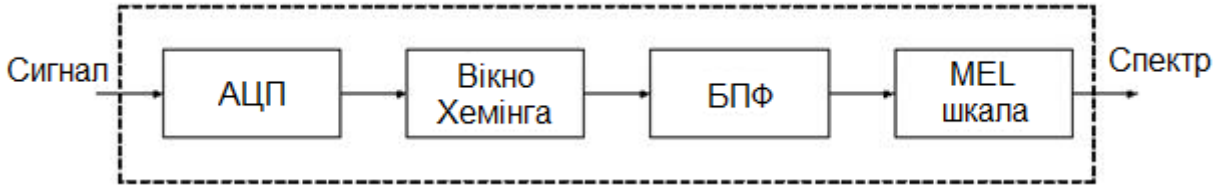


Рисунок 2.1 – Попередня обробка мовного сигналу

Зважування сигналу ваговою функцією вікна Хеммінга зменшує спектральні спотворення сигналу через граничні умови. Застосування часового вікна доцільно для інтервалів, що перевищують 15 мс або включають кілька періодів основного тону.

Значення вагової функції задається формулою [12]:

$$W_n = \begin{cases} 0{,}54 - 0{,}46 \cdot \cos\left(\dfrac{2\pi n}{N-1}\right), & 0 < n < N \\ 0 & \textit{інакше} \end{cases} \qquad (2.3)$$

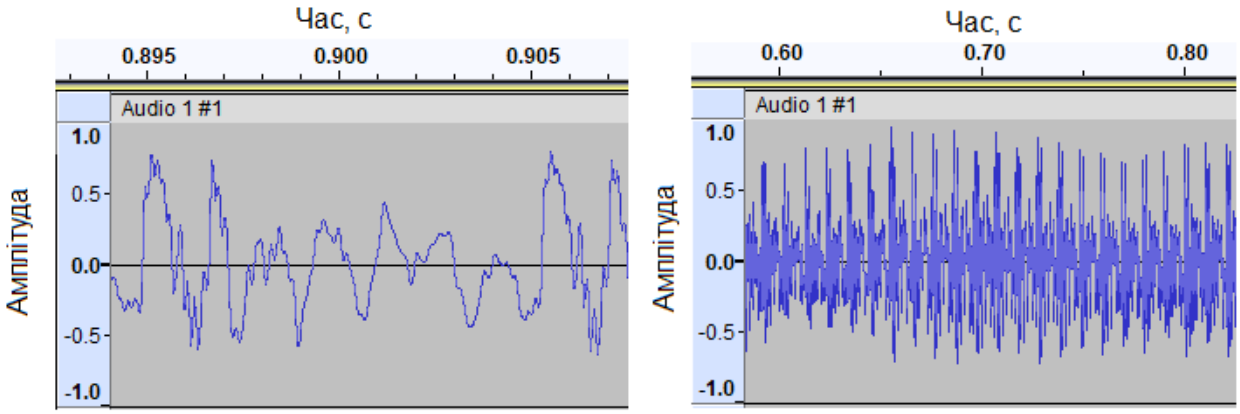
Інформативність різних частин спектра неоднакова: в низькочастотній області міститься більше інформації, ніж високочастотної. Тому стискають високочастотну область спектра у просторі частот. Найбільш поширений метод завдяки його простоті – логарифмічний стиск, або mel-стиск [13]

$$mel = 1125 \ln(1 + f / 700). \qquad (2.4)$$

де f – частота в спектрі, Гц;

mel – частота в новому стиснутому частотному просторі.

Зразки сегментів мовного сигналу голосної «а» наведено на рис. 2.2.



а) б)

Рисунок 2.2 – Зразки сегментів мовного сигналу голосної «а»:
а – інтервал 10 мс, б – інтервал 200 мс

На рис. 2.3 показаний результат частотного аналізу 16-бітного мовного сигналу із частотою дискретизації 44100 Гц, виконаний у вікні аналізу Analyze – Plot Spectrum звукового редактора Audacity. Подібний спектр коливань повітря формується голосовими зв'язками та джерелом звуку в ротовій порожнині шляхом вибіркового резонансу, що виникає при передачі звуку мовним трактом.

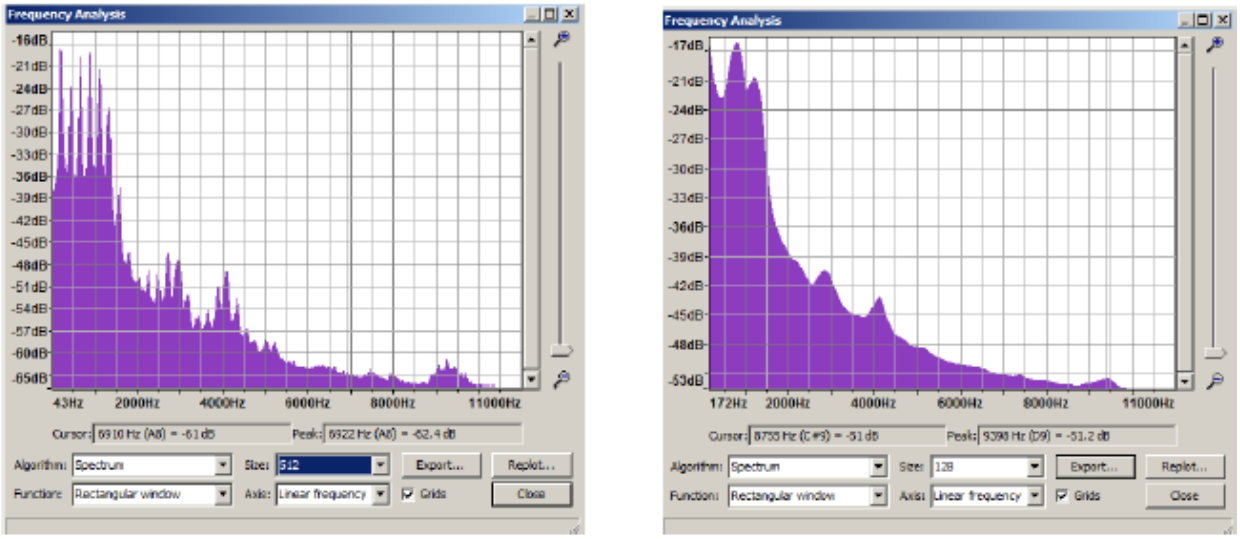
Мовний тракт утворюють горло, ротова порожнина, мова, носова порожнина тощо.

На рис.2.3 показані непараметричні оцінки спектра із високою (а) та низькою (б) роздільною здатністю за частотою. Звук «а». Жіночий голос. *N*seg = 512 (рис. 2.3, а) та *N*seg = 128 (рис. 2.3, б), *F*s= 44100 Гц.

На рис.2.4 показані непараметричні оцінки спектра із високою (а) та низькою (б) роздільною здатністю за частотою. Звук «і». Жіночий голос. *N*seg = 512 (рис. 2.4, а) та *N*seg = 128 (рис. 2.4, б), *F*s= 44100 Гц.

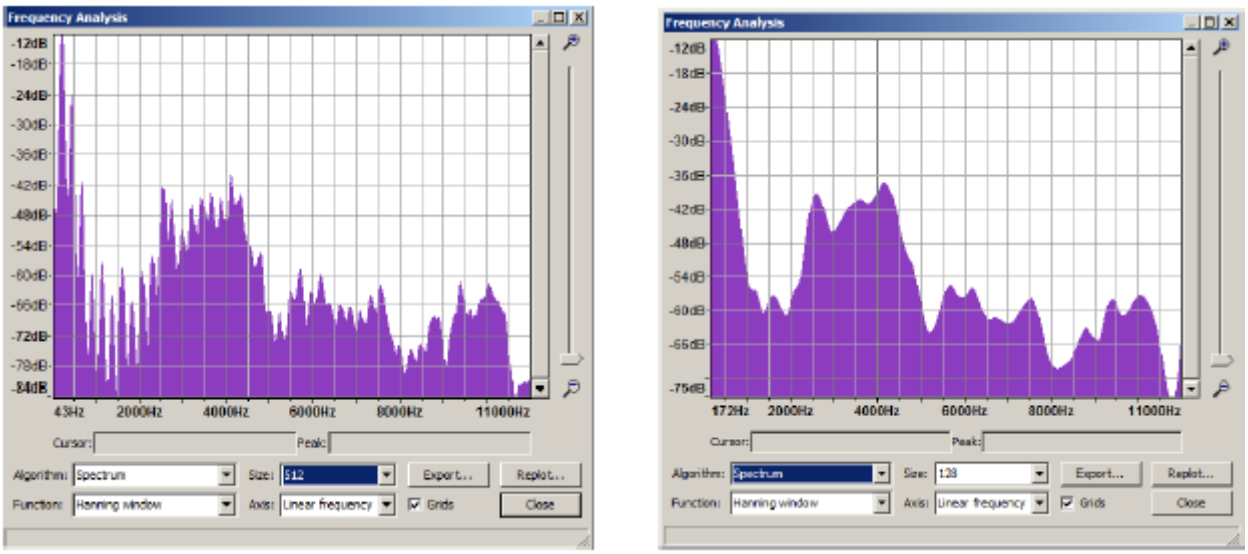
Форманти – максимуми розподілу енергії звукового сигналу в координатах амплітуда, частота, час. Для отримання хорошої якості сигналу достатньо задати параметри кількох старших формант основного тону. Коли

потрібно досягти високої якості, використовуються деякі з перерахованих параметрів або їх комбінації.



а) б)

Рисунок 2.3 – Непараметричні оцінки спектра звуку «а» із високою (а) та низькою роздільною здатністю



а) б)

Рисунок 2.4 – Непараметричні оцінки спектра звуку «і» із високою (а) та низькою роздільною здатністю

В табл.2.1 наведено порівняльні результати оцінювання частот перших 2-3-х формант.

Таблиця 2.1 – Порівняльні результати оцінювання частот перших 2-х формант

Голосні	Частота основного тону, Гц	Форманта 1, Гц	Форманта 2, Гц
Чоловічий голос			
а	110	1050	2770
і	185	2670	3440
Жіночий голос			
а	110	950	2870
і	185	2617	4140

Порівнюючи дані, отримані в табл.2.1 з відомими параметрами формант [13], можна відмітити наступне. Так, наприклад, голосний звук «а» незалежно від свого основного тону, тобто незалежно від того, на якій висоті голосу він вимовлений, має характерну для цього звуку форманту, що охоплює область від 1000 до 1400 Гц, звук «і» – від 2800 до 4200 Гц. Тобто, проведені оцінки непараметричним методом близькі до відомих експериментальних результатів.

Мовний сигнал має низку особливостей, які необхідно враховувати:

– властивості сигналу не постійні на вибраному для аналізу відрізку завдовжки у слово, це нестаціонарний випадковий процес;

– складність форми сигналу (мова нагадує швидше шум, ніж регулярний сигнал).

Для подолання цих труднощів, як зазначалося вище, дискретний випадковий процес оцифрованого мовного сигналу вважається стаціонарним на інтервалі близько 10 мс, оскільки параметри голосового тракту цьому інтервалі значно змінюються. Це обґрунтований експериментально часовий інтервал.

Основне завдання обробки сигналу полягає у обчисленні за вхідним сигналом сукупності параметрів (ознак), які містять інформацію про сигнал, що використовується при синтезі та розпізнаванні.

2.2 Структурна схема обробки звукового сигналу для оцінки параметрів голосу

Зазвичай визначають такі параметри сигналу:

– частоту основного тону та формування траєкторії періоду основного тону;

– короткочасну енергію для синтезу траєкторії короткочасної енергії;

– коефіцієнти лінійного передбачення (КЛП) для побудови траєкторії передавальної функції мовного тракту;

– формантні частоти для відтворення траєкторії формантних частот.

На рис.2.5 показано структурну схему обробки звукового сигналу для оцінки зазначених вище параметрів.

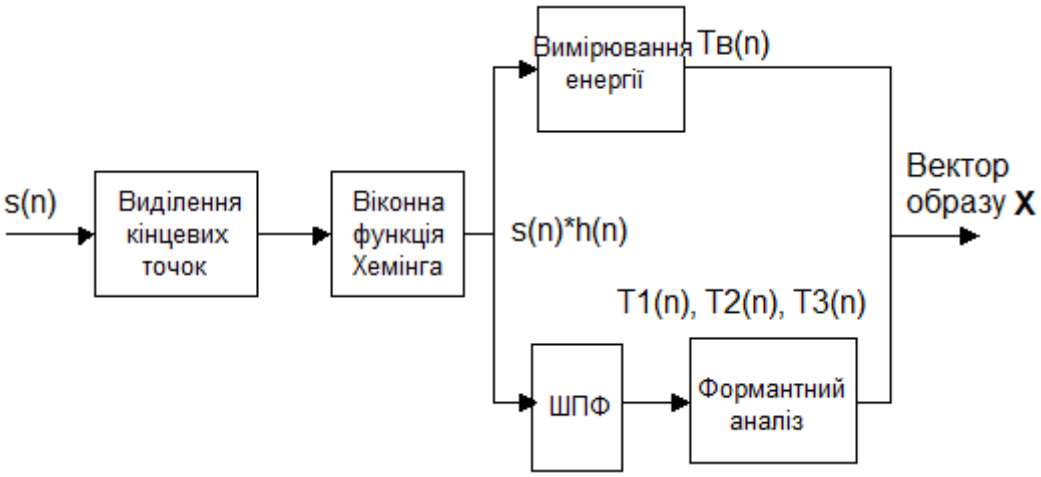


Рисунок 2.5 – Структурна схема обробки звукового сигналу для оцінки зазначених параметрів голосу

Один з алгоритмів виділення фрази (запропонований Л. Рабінером) заснований на вимірюванні двох простих характеристик – енергії і числа переходів через нуль. При підрахунку середнього значення енергії використовується вікно 10 мс (приблизно 110 відліків), у якому підсумовуються квадрати відліків (рис. 2.6).

Передбачається, що перші 50 мс сигнал не містять мовного сигналу.

Рисунок 2.6 – Виділення фрази

У межах обраного часового сегмента обчислюється середнє значення енергії шуму E і поріг P, який береться рівним подвоєної енергії шуму. При подальшій обробці, якщо середнє значення енергії перевищило поріг, фіксується момент запису мовного сигналу (початок фрази), який запам'ятовується. Якщо середнє значення енергії поменшає порога, то запам'ятовується кінець фрази.

На рис. 2.7 наведено графіки зміни енергії сигналу для фонеми "Р" в слові «РОК».

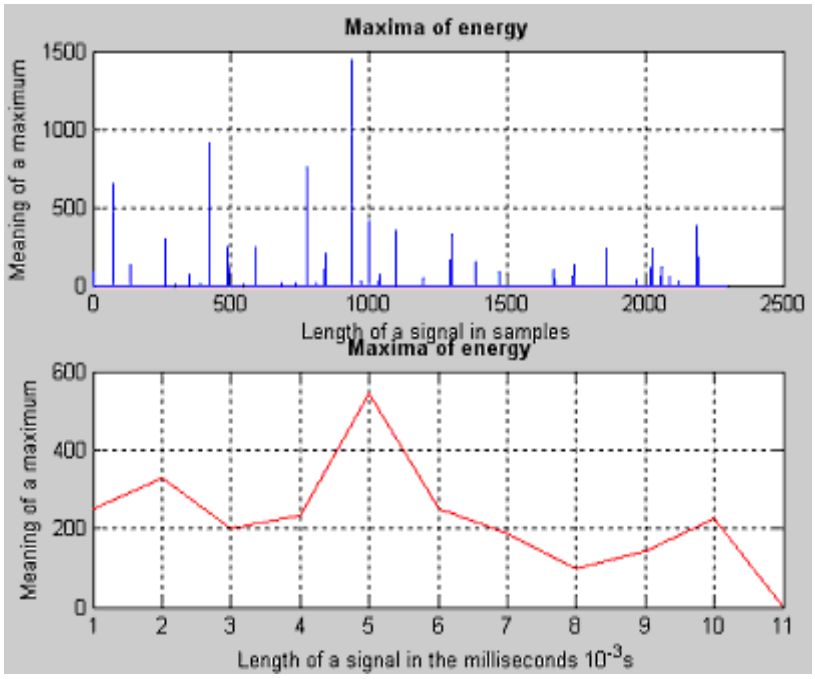


Рисунок 2.7 – Максимуми енергії у спектрі фонеми "Р" у слові «РОК»

Частота основного тону, енергія та тривалість забезпечують формування просодичних характеристик мови.

2.3 Параметричне спектральне оцінювання

У параметричних моделях аналізованому випадковому процесу ставиться у відповідність модель часового ряду. Приймається, що модель збуджується білим шумом і має раціональні системні функції. Вихідні процеси в цій моделі описуються за допомогою параметрів моделі та дисперсії білого шумового процесу.

При використанні класичних методів спектрального оцінювання відсутні дані або дані за межами вікна неявно вважаються рівними нулю, що призводить до спотворень спектральних оцінок. Параметричні моделі позбавлені цього недоліку.

Параметричний метод спектрального оцінювання складається із трьох етапів:

– на першому етапі проводиться вибір параметричної моделі часового ряду. Вибрана нами модель авто регресії (АР) дає спектри з гострими піками;

– на другому етапі обчислюються оцінки параметрів моделі:

– на третьому етап оцінені значення параметрів вводяться у вираз для спектральної щільності потужності, що відповідає обраній моделі.

Ступінь покращення розрізнення та підвищення достовірності спектральних оцінок визначається відповідністю обраної моделі аналізованому процесу та можливістю апроксимації виміряних даних за допомогою декількох параметрів моделі.

Модель часового ряду, яка апроксимує аналізований процес, описується виходом фільтра, що виражається наступним лінійним різницевим рівнянням [14]:

$$s[n] = -\sum_{k=1}^{p} a[k]x[n-k] + \sum_{k=0}^{q} b[k]s_{вх}[n-k] = \sum_{k=0}^{\infty} h[k]s_{вх}[n-k], \qquad (2.5)$$

де $s[n]$ – послідовність на виході казуального фільтра;

$s_{вх}[n]$ – вхідна збуджуюча послідовність.

Оскільки вхідний процес $s_{вх}[n]$ зазвичай недоступний для спостереження, приймемо припущення, що це білий шум з нульовим середнім значенням та дисперсією ρ. Якщо в (2.5) всі $b[k]$, крім $b[0]=1$, покласти рівними нулю, то

$$s[n] = -\sum_{k=1}^{p} a[k]x[n-k] + s_{вх}[n-k], \tag{2.6}$$

і отримуємо АР-процес порядку p.

Оцінки параметрів АР-моделі можна отримати як розв'язання лінійних рівнянь, в той же час, як інші параметричні оцінки параметрів, вимагають вирішення нелінійних рівнянь.

Модифікований коваріаційний метод [15] забезпечує найкращі результати при наявності в даних синусоїдальних компонентів. Цей метод фактично дає оцінки лінійного передбачення, які потім використовуються як оцінки АР-параметрів.

За обчисленими оцінками АР-параметрів визначається авторегресійна оцінка спектральної щільності потужності на частоті f

$$P_{AP}(f) = \frac{T\rho}{\left|1 + \sum_{n=1}^{p} a[n]\exp(-j2\pi fnT)\right|^2}, \tag{2.7}$$

де T – інтервал відліків.

На рис. 2.8 наведено графік параметричної оцінки спектру звуку "а" української мови (F1 = 700 Гц; F2 = 1200 Гц; F3 = 2500 Гц). Цікаво порівняти цей графік із табл. 2.1. Як бачимо, має місце суттєва схожість.

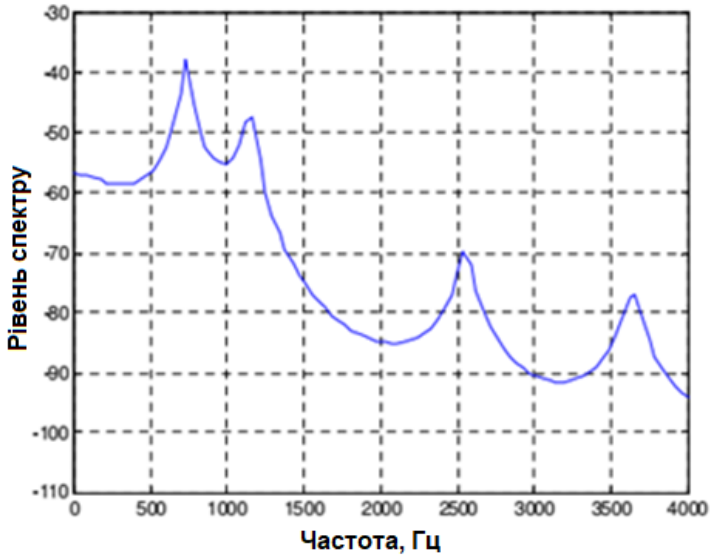


Рисунок 2.8 – Параметрична оцінка спектру звуку "а"

Аналогічну схожість результатів маємо для звуку «і» в табл. 2.1 для української мови (F1 = 200 Гц; F2 = 2250 Гц; F3 = 3300 Гц).

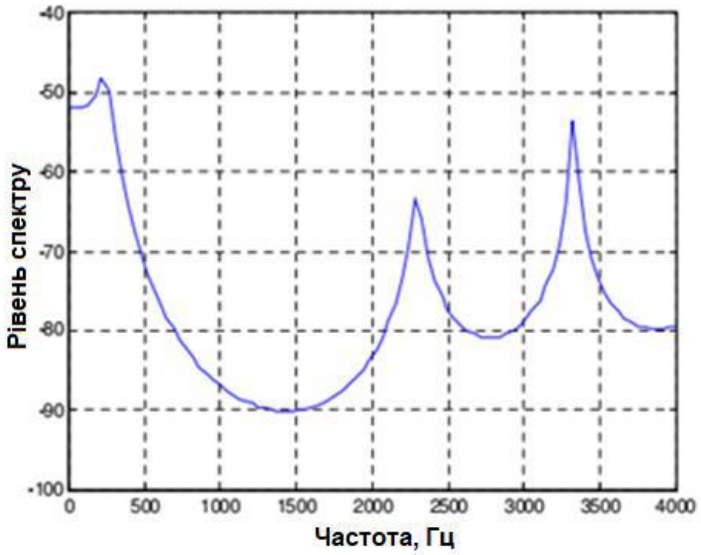


Рисунок 2.9 – Параметрична оцінка спектру звуку "і"

Отже, при використанні класичних методів спектрального оцінювання відсутні дані або дані за межами вікна неявно вважаються рівними нулю, що призводить до спотворень спектральних оцінок. Параметричні моделі позбавлені цього недоліку.

2.4 Кепстральний аналіз

Кепстр (кепструм) – це зворотне перетворення Фур'є від натурального логарифму квадрата спектральної щільності випадкового процесу, що відображається в вигляді функції $C(q)$ від так званого кепстрального часу q

$$C(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \ln\left|S(j\omega)\right|^2 \exp(j\omega q) d\omega, \tag{2.8}$$

$$S(j\omega) = \int_{-\infty}^{\infty} s(t) \exp(-j\omega t) dt. \tag{2.9}$$

В даному випадку $s(t)$ – вхідний коливальний процес, що в принципі допускає полігармонічну апроксимацію за допомогою адитивних комбінацій наборів синусоїдальних функцій:

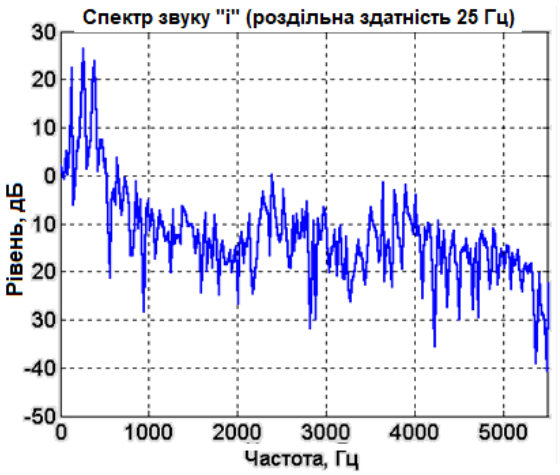
$$s(t) = s_1(t) + s_1(t) + \ldots s_n(t) + \ldots, \tag{2.10}$$

де $s_n(t) = A_n \sin(2\pi f_n + \varphi_n)$ – гармонійні звукові сигнали, представлений у вигляді гармонійних складових з амплітудами A_n, частотами f_n − та фазами φ_n. Реальні звукові сигнали можуть не мати вираженої періодичності і стаціонарності. Тут $S(j\omega)$ – комплексна функція прямого перетворення Фур'є.

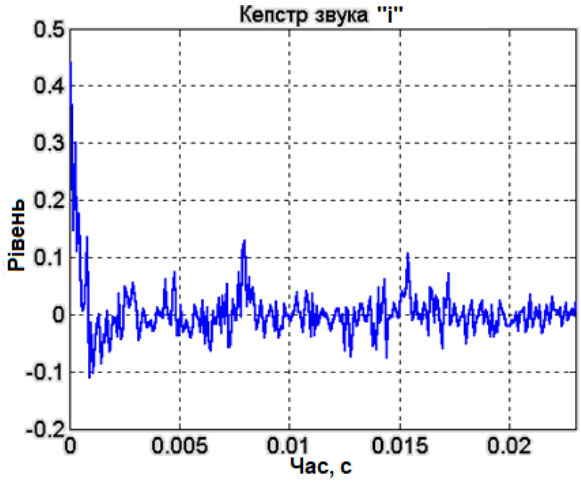
Параметр q має розмірність часу, у кепстральному аналізі він умовно називається сачтотою, а його застосування забезпечує рознесення результуючих енергетичних сплесків по осі кепстрального часу. Як правило, енергетичні сплески в реальних звукових сигналах розміщуються вздовж осі сачтот q з віддаленням від нульової позначки.

На рис. 2.10 показано результати обчислення обвідної спектру звуку «і» через спектр цього звуку, а також кепстр звуку «і». При цьому параметри

обчислень мали наступні значення: $T_{фр}$ = 46,4 мс; $T_{пер}$ = 23,2 мс; dF_1 = 25 Гц; τ = 11,2 мс; dF_2 = 100 Гц.



а)



б)

в)

Рисунок 2.10 – Спектр (а), кепстр (б) та обвідна спектру (в) звуку «і»

2.5 Розробка програмного забезпечення в системі MATLAB для розрахунків і досліджень

2.5.1 Підготовка до обчислень спектрограми

Ознайомившись із синтаксисом Matlab, зчитуємо звуковий файл із диска, будуємо графік сигналу (рис.1.1) та прослуховуємо сигнал:

```
[x, fs] = audioread('speech.wav');
t = 1/fs:1/fs:length(x)/fs;
figure; plot(t,x);
xlabel('Time (sec)'); ylabel('Level');
soundsc(x,fs)
```

На рис. 2.11 показана форма сигналу для слів «Двадцать один».

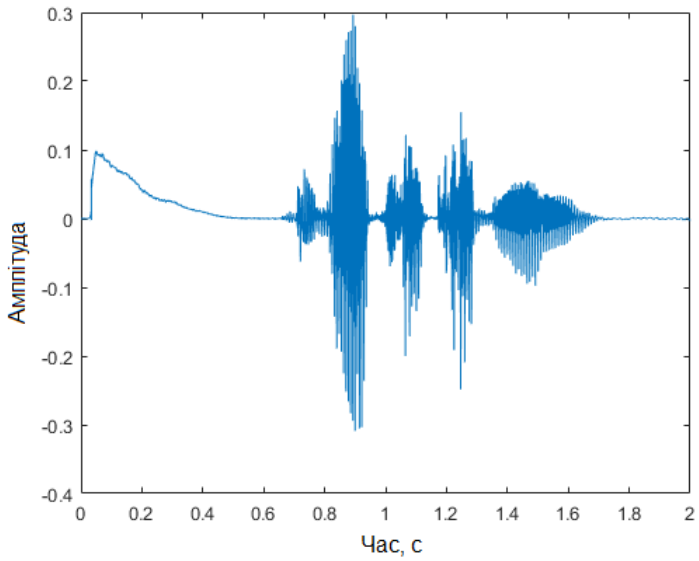


Рисунок 2.11 – Форма сигналу для слів «Двадцать один»

Підготовка до обчислень спектрограми:

```
Tseg = 0.1;
Nseg = round(Tseg*fs);
window = hamming(Nseg);
nfft = Nseg;
noverlap = round(Nseg/2);
```

2.5.2. Обчислення спектрограми за допомогою бібліотечної функції Matlab

Синтаксис:

```
spectrogram(x,window,noverlap,nfft,fs)
```

або

```
S = spectrogram(x,window,noverlap,nfft,fs)
```

В першому випадку побудова графіка спектрограми (рис. 2.12) виконується автоматично.

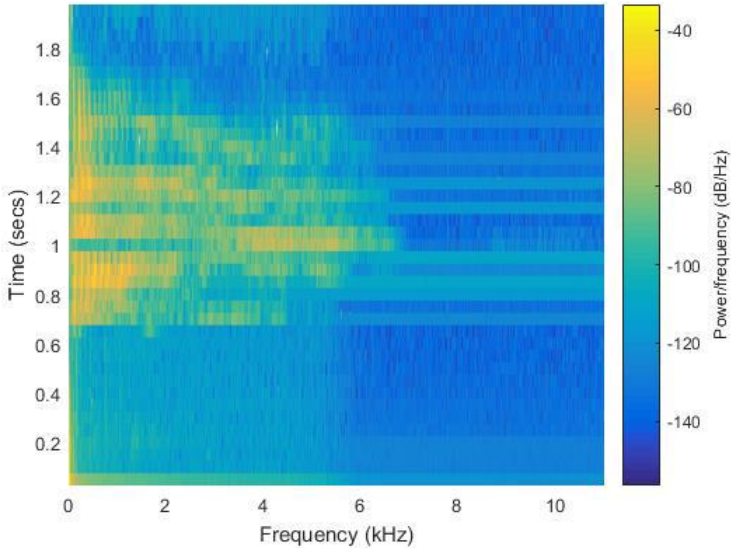


Рисунок 2.12 – Спектрограма мовного сигналу «Двадцать один»

Звернемо увагу на графік рис. 2.12: частоти розташовані уздовж осі x, час – уздовж осі y. Це не завжди є зручним. Наприклад, в ряді програм, таких як Sound Forge, Audacity та інших, вісь часу розташована горизонтально, а вісь частот – вертикально. Щоби отримати такий графік (рис. 2.13), треба додати в аргумент функції опцію

```
‘yaxis’:
spectrogram(x, window, noverlap, nfft, fs, 'yaxis');
```

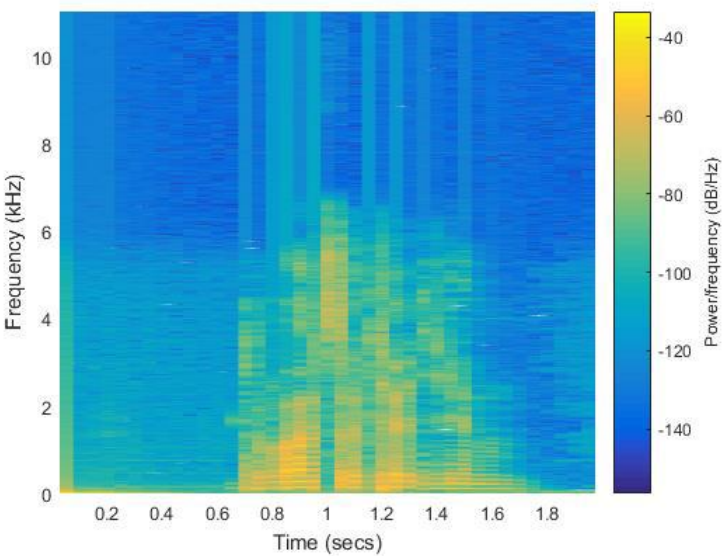


Рисунок 2.13 – Спектрограма сигналу «Двадцать один» – повернуто вісі

2.5.3 Обчислення спектрограми за допомогою самостійно розробленої програми-функції

Аргументом на користь такої ідеї є прагнення спростити синтаксис. Результат обчислень представлено на рис. 2.14.

```
функція S = myspecgram(filename,dt,df)
% === обчислення спектрограми ===
%
% ВХІДНІ ДАНІ:
%filename - ім'я звукового файлу
% dt - дозвіл за часом
% df - роздільна здатність за частотою
% ВИХІДНІ ДАНІ:
%S – спектрограма
%
% === читання звукового файлу ===
[x, fs] = audioread(filename);
%
% === підготовка до поділу на кадри ===
nfft = round(fs/df); % Довжина кадру = параметр БПФ = Довжина вікна
shft = round (dt * fs); % зсув кадрів, виражений у кількості вибірок
nfrm = floor((length(x)-nfft)/shft); % кількість кадрів
%
% === розподіл x(n) на кадри ===
xfrm = zeros (nfft, nfrm); % виділення області
for k = 1:nfrm
nach = 1 + (k-1) * shft; kon = nach + nfft-1; % номера меж кадрів
xfrm(:,k) = x(nach:kon); % матриця кадрів
end
%
% === зважування кадрів вікном Хеммінга ===
w = hamming (nfft);
winmatr = repmat (w, 1, nfrm);
xwin = winmatr. * xfrm;
%
% === БПФ від кожного кадру ===
sp = fft (xwin);
S1 = sp. * conj (sp);
%
```

```
% === відкидаємо заперечень частоти, беремо модуль і нормир. по макс. ===
S1 = S1(1:round(nfft/2),:);
Sabs = abs (S1);
Samax = max(max(Sabs));
S2 = Sabs/Samax;
S = im2uint16(S2);
%
% === виведення графіків ===
tx = 0:1/fs:(length(x)-1)/fs;
figure; subplot(2,1,1); plot(tx,x);
xlabel('Час, с'); ylabel('Рівень');
t = 0:dt:(size(S,2)-1)*dt;
f = 0:df:(size(S,1)-1)*df;
subplot(2,1,2); image(t,f,S);
set(gca, 'YDir', 'normal');
colormap(jet);
xlabel('Час, с'); ylabel('Частота, Гц');
```

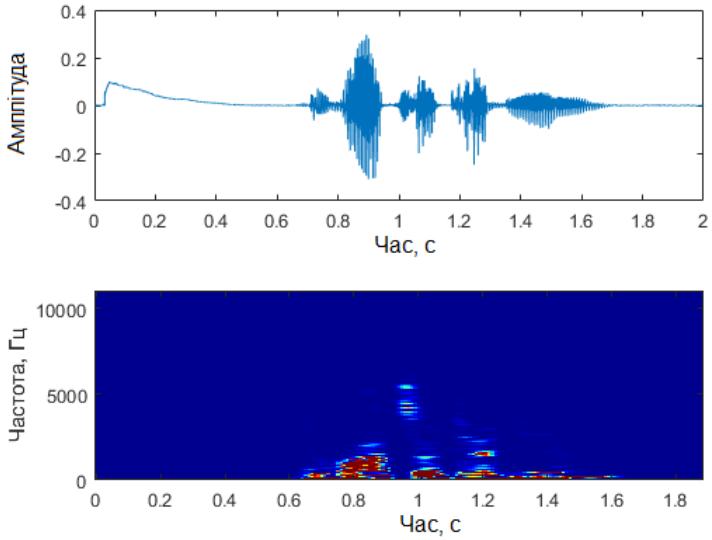


Рисунок 2.14 – Побудова «саморобної» спектрограми

Отже, розроблені програми в системі MATLAB дозволять провести експериментальні дослідження параметрів голосу.

2.6 Висновки по розділу 2

В другому розділі розглянуті методи оцінювання параметрів мовного тракту по звуковому сигналу. Зазвичай визначають такі параметри: частоту

основного тону, короткочасну енергію, формантні частоти. Також відслідковують траєкторію цих параметрів в часі.

Першими розглянуті непараметричні оцінки, основані на швидкому перетворенні Фур'є з вікном Хемінга. Реальний мовний сигнал має кінцеву тривалість, при поданні в частотній області спектр необмежений. Тому сигнал сегментують на ділянки близько 10 мс, на яких вважається стаціонарним.

Отримано спектральні оцінки чоловічого і жіночого голосів при промові фонем «а» та «і» у високому і низькому спектральному розрізненні. Високе спектральне розрізнення звучно використовувати при оцінці частоти основного тону. Низьке – при оцінюванні формантних резонансних частот звукового тракту. Отримані оцінки приблизно співпадають з відомими даними для української мови. Відмінність у результатах спостерігається в оцінці формант і викликано багатомодовістю спектру Фур'є навіть при низькому спектральному розрізненні, що ускладнює оцінку формант.

Розроблено структурну схему обробки звукового сигналу для оцінки зазначених вище параметрів. Виділення фраз відбувається таким шляхом. У межах обраного часового сегмента обчислюється середнє значення енергії шуму E і поріг P, який береться рівним подвоєної енергії шуму. При подальшій обробці, якщо середнє значення енергії перевищило поріг, фіксується момент запису мовного сигналу (початок фрази), який запам'ятовується. Якщо середнє значення енергії поменшає порога, то запам'ятовується кінець фрази.

Досліджено параметричний метод спектрального оцінювання. Він складається із трьох етапів: спочатку проводиться вибір параметричної моделі часового ряду. Вибрана модель авторегресії дає спектри з гострими піками. На другому етапі обчислюються оцінки параметрів моделі. На третьому етап оцінені значення параметрів вводяться у вираз для спектральної щільності потужності, що відповідає обраній моделі.

Отримано графік параметричної оцінки спектру звуків "а" і «і» української мови. Оцінки формантних частот набагато краще відповідають відомим в літературі результатам, ніж непараметричні оцінки.

Нарешті, розглянуто кепстральний аналіз. Кепстр– це зворотне перетворення Фур'є від натурального логарифму квадрата спектральної щільності випадкового процесу, що відображається в вигляді функції $C(q)$ від так званого кепстрального часу q. Параметр q має розмірність часу, у кепстральному аналізі він умовно називається сачтотою, а його застосування забезпечує рознесення результуючих енергетичних сплесків по осі кепстрального часу. Як правило, енергетичні сплески в реальних звукових сигналах розміщуються вздовж осі сачтот q з віддаленням від нульової позначки.

Отримано результати обчислення обвідної спектру звуку «і» через спектр цього звуку, а також кепстр звуку «і». Кепстр можна використовувати як вхідні дані для нейронних мереж, які розпізнають звук.

В середовищі MATLAB розроблено ряд программ, що дозволяють обчислювати кепстри, спектрограми, і параметричні спектри. Ці програми були використані в експериментальних дослідженнях.

3 ЕКСПЕРИМЕНТАЛЬНЕ ВИЗНАЧЕННЯ ФОРМАНТНИХ ЧАСТОТ ІЗ ВИКОРИСТАННЯМ СПЕКТРАЛЬНОГО РОЗКЛАДАННЯ МОВНОГО СИГНАЛУ

3.1 Методика дослідження

Форманти є одними з основних елементів ідентифікації особи в мовному сигналі тому, що природа їхнього походження пов'язана з порожнинами людського мовного тракту. Зважаючи на індивідуальність подібних компонентів для кожної людини, можна дійти висновку, що визначення формантних частот є важливим компонентом побудови системи ідентифікації мовної інформації.

Дослідження формантних атрибутів найчастіше виконується шляхом:

– порівняння спектра формант однакових фонемічних елементів (ударних голосних, голосних у кінці або на початку слів тощо);

– порівняння спектрів формант для зрізів спектрограм (кожного елемента спектрограми або всередині слів, складів та ін.);

– порівняння динамічних змін частот формант уздовж усього мовного сигналу чи у важливих його компонентах.

Виокремлення формант супроводжує ряд проблем, пов'язаних з їхньою динамічною зміною у процесі мовлення. Навіть однакові голосні змінюють формантний набір залежно від свого розташування у складі слів, складів тощо. Труднощі також визивають проблеми, пов'язані з близьким розташуванням піків під час аналізу спектрограм і проблемами правильного визначення піків максимумів формант на спектрограмі. Визначення розташування формант на спектрограмах мовного сигналу достатньо легко виконується людиною, але автоматизація цього процесу визиває деякі труднощі. Типове представлення спектрограми з розміченими людиною розташуваннями формант представлено на рис. 3.1, що відповідає зонам

розташування формант, причому характеризується не тільки середньою частотою форманти, а також її шириною.

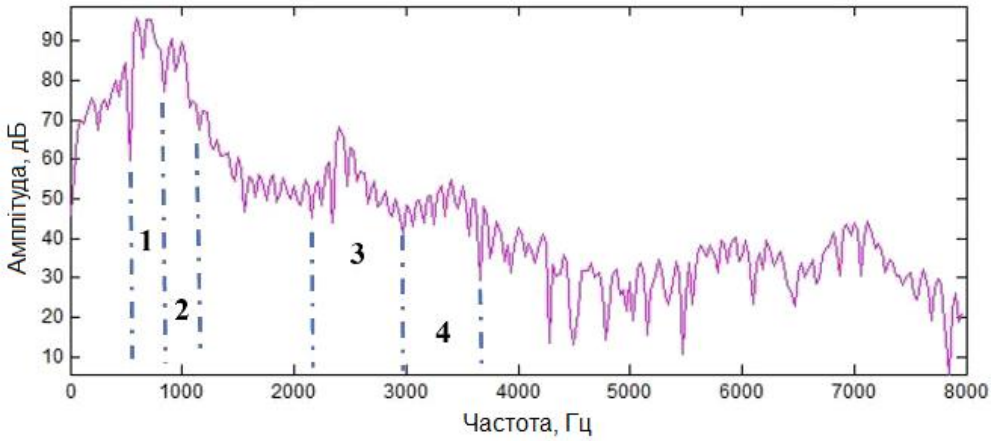


Рисунок 3.1 – Визначення формант людиною

Проведення подібного розмічення є досить складною задачею, зважаючи не велику кількість конкуруючих частотних піків. Трудомісткість подібного роду маркування формант досить висока, тому в експериментальних дослідженнях використовують заздалегідь розмічені дані, що можна знайти в різних типах мовних датасетів. В наступному дослідженні будемо використовувати датасет VTR-TIMIT, що має підготовлений набір розмічених даних подібного типу.

Крім того, під час проголошення окремих видів звуків на положення формант можуть впливати безліч факторів, що може приводити до коливань формантних частот, а на окремих фрагментах навіть відсутності деяких із них.

Існує кілька підходів для визначення положень формант на частотній шкалі, але всі вони базуються на аналізі та перетвореннях спектрограмами мовного сигналу. При виділенні формантних частот першим етапом дослідження завжди є побудова спектрограми за визначеними дослідником критеріями. Серед них є ширина кадру, тип спектрального аналізу, частотний діапазон, вид вікна та ін.

Наступним етапом є таке представлення спектрограми, що дозволить провести сегментацію формантних частот на основі різних математичних

методів, найчастіше всі вони базуються на алгоритмах кластеризації або на побудові обвідної спектра. Найвідомішими з алгоритмів, що використовують обвідну спектра, є:

– метод лінійного передбачення – коли обвідна спектра будується на алгоритмі LPC [18];

– апроксимації спектра кубічними сплайнами або іншими видами функцій [18].

Однак згідно з відомими дослідженнями обидва алгоритми мають практично однакову точність при різній обчислювальній складності [18].

Приклад обвідної на основі алгоритму лінійного передбачення виділеного спектра для стандартного фрагмента дослідження (20 мс) зображено на рис. 3.2.

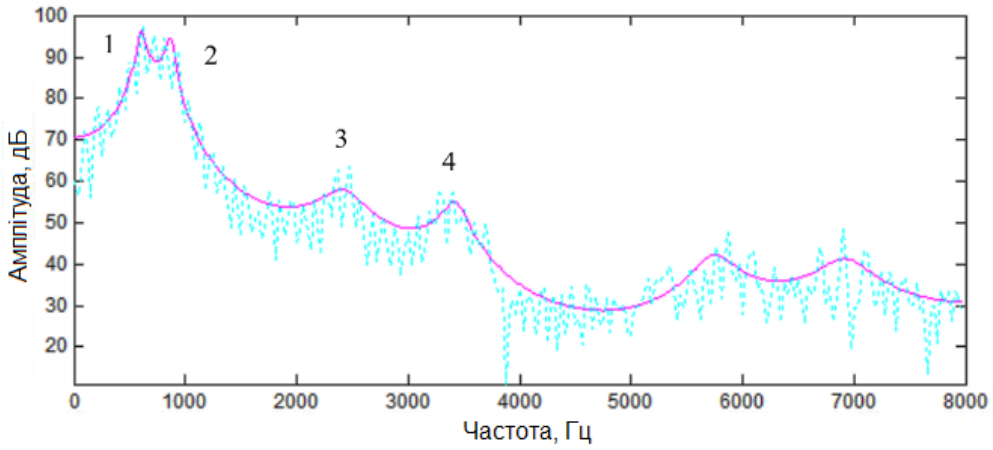


Рисунок 3.2 – Побудова обвідної для спектра мовного сигналу (числами зображено формантні діапазони 1–4)

На основі цього рисунка виділяють локальні максимуми обвідної для спектра мовного сигналу, які розглядають як центри формантних частот.

Проведені попередні дослідження вказують на відповідність розподілу частот відносно голосних звуків, які вносять найбільшу вагу в формування формант в мовному сигналі. Установлено, що частина голосних звуків у більшості мов розташована в частотному діапазоні 200–500 Гц, а інша частина голосних звуків у діапазоні від 500 до 1500 Гц. Зважаючи на це,

раціональним є окремий розгляд цих частотних діапазонів, під час формування характерних ознак мовного сигналу, що дозволяє підвищити кількість параметрів, та набирати більшу статистику при визначенні максимумів формантних частот.

3.2 Алгоритм визначення формантних частот

Результатом проведеного огляду підходів до визначення формантних частот став алгоритм (рис. 3.3) [18], що складається з таких етапів.

Рисунок 3.3 – Алгоритм визначення формантних частот

Алгоритм складається з таких етапів:

– сегментація мовного сигналу на вокалізовані фрагменти та паузи;

– розбиття вокалізованих фрагментів сигналу на часові фрейми;

– для кожного фрагмента отримання спектра на основі авторегресійної моделі;

– побудова обвідної спектра;

– знаходження всіх максимумів;

– визначення положень формантних діапазонів;

– отримання максимумів формантних діапазонів;

– побудова графіків траєкторії положення формант (рис. 3.4);

– розрахунок залежності щільності ймовірності розподілу кожної із чотирьох формантних частот (максимумів формантних частот).

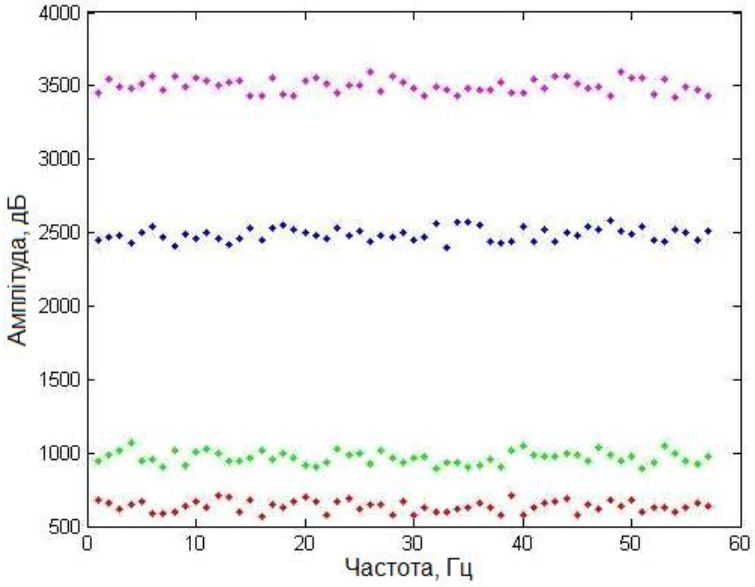


Рисунок 3.4 – Траєкторія положення формант за номерами фреймів

Сегментація мовного сигналу на вокалізовані фрагменти й паузи виконується методом оцінювання змін фрактальної розмірності [18]. У роботі визначено, що фрактальна розмірність D для невокалізованих фрагментів у 99 % випадків перебуває в межах $1{,}04 \leq D \leq 1{,}45$, а фрактальна розмірність вокалізованих фрагментів не спостерігалася менше $D = 1{,}55$ для часового вікна розміром 20 мс.

Операцію розбиття виділених вокалізованих фрагментів мовного сигналу на часові фрейми виконувалась фреймами 10–20 мс для тестування робочої спроможності методу.

3.3 Отримання обвідної спектра мовного сигналу

Отримання обвідної спектра мовного сигналу виконувалося з використанням моделі авторегресії.

Робота полюсного фільтра $H(z)$, який моделює форму голосового тракту в момент виголошення звуку, може бути описана в різницевій формі:

$$s(n) = -\sum_{k=1}^{p} a_k s(n-k) + gw(n), \qquad (3.1)$$

де $s(n)$ – мовний сигнал;

$w(n)$ – збуджуючий процес;

g – коефіцієнт посилення;

a_k, k = 1, 2, . . ., p – авторегресійні коефіцієнти.

Порядок авторегресійної моделі p вибирається, як правило, рівним 8…20. Очевидно, що підвищення порядку моделі дозволяє більш точно оцінювати спектральні характеристики мови.

Основною властивістю параметрів авторегресійної моделі є те, що вони відносно повільно змінюються з часом. Можна вважати, що ці параметри залишаються незмінними на відрізках довжиною 10…30 мс (властивість квазістаціонарності мовних сигналів).

Традиційно авторегресійні коефіцієнти визначаються за допомогою методів лінійного передбачення, зокрема автокореляційного методу. Тому авторегресійні коефіцієнти також часто називають коефіцієнтами лінійного передбачення. На рис. 3.5 показано приклад апроксимації спектра потужності мовного сигналу (тривалістю 10 мс) за допомогою функції

$$\left|H(e^{j\omega})\right|^2 = \frac{g^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-jk\omega}\right|^2}, \qquad (3.2)$$

обчисленою за допомогою автокореляційного методу.

Визначаючи максимуми обвідної спектра (рис. 3.5), отримуємо максимуми чотирьох перших формант. Набір формант для кожного фрейму зображено на графіку відповідно до частоти форманти та номера фрейму, з якого вона була отримана (рис. 3.4).

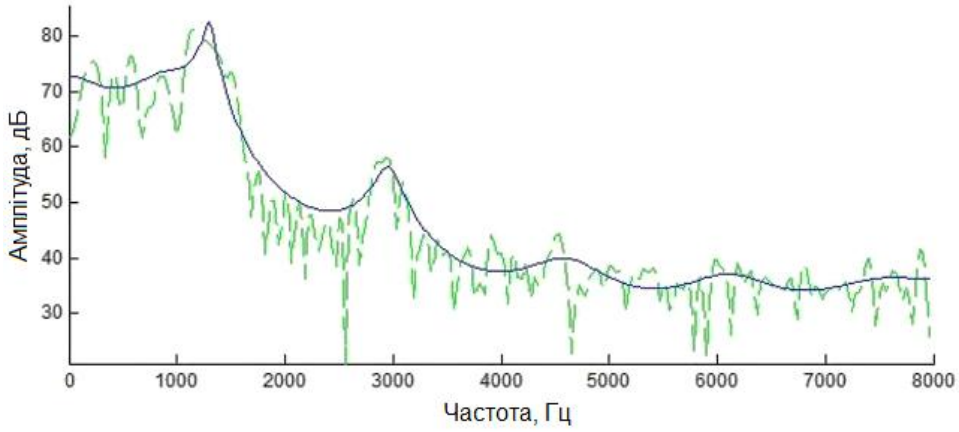


Рисунок 3.5 – Обвідна спектра мовного сигналу

Аналіз цього графіка показує достатньо високу стабільність визначення формантних частот для розглянутого мовного сигналу.

3.4 Порівняльне дослідження трекінгу формант

Для дослідження розглядались формант-трекери PRAAT і SNACK [19, 20]. Налаштування кожного з них здійснювалося на основі набору параметрів за замовчуванням, що було закладено розробниками цих трекерів.

Набір налаштувань для кожного з трекерів представлено у табл. 3.1. Основні параметри налаштувань відомих трекерів визначено на основі [19, 20].

Таблиця 3.1 – Налаштування формант-трекерів, використаних у дослідженні

Параметр	PRAAT	SNACK	Досліджений метод
Число формант	5	4	4
Порядок LPC	10	12	10
Вікно	Гаус	Cos4	Гаус
Розмір вікна, мс	25	25	25
Крок, мс	10	10	10

Необхідно також зазначити, що кожен із цих відомих трекерів має особливості використання відповідно до статі особи, так [19, 20]:

– PRAAT оптимізовано для мовних сигналів жінок,

– SNACK оптимізовано під мовні сигнали чоловіків.

Налаштування часових інтервалів досліджуваного методу має відповідати розміру подібних інтервалів для інших формант-трекерів для коректного порівняння, тому використано розмір вікна 25 мс із кроком 10 мс.

У дослідженні трекери самостійно виконували сегментацію на вокалізовані фрагменти і паузи, застосовуючи датасет VTR-TIMIT, у випадку некоректної сегментації (невокалізований фрагмент вважався вокалізованим) помилкові результати сегментації вилучалися з розгляду.

Як параметри порівняння використовували середньоквадратичне відхилення формант між еталонною розміткою мовних сигналів та результатами формант-трекерів (табл. 3.2).

Таблиця 3.2 – Середньоквадратичне відхилення визначення формант

Форманта	Стать	PRAAT	SNACK	Досліджений метод
F1, Гц	ж	97	104	78
	ч	161	92	70
F2, Гц	ж	185	197	89
	ч	215	209	91
F3, Гц	ж	194	183	167
	ч	247	261	144

Розгляд саме трьох формант, замість чотирьох, пов'язаний із тим, що в датасеті VTR-TIMIT розмічені експертами лише три форманти. Крім того, проведений аналіз цього датасету [18] показав середнє відхилення частоти максимуму для 1–3 формант відповідно 78, 100, 111 Гц, тому отримані значення достатньою мірою відповідають раніше проведеним дослідженням.

Порівняльний аналіз показує достатньо високу точність визначення формантних частот порівняно з існуючими формант-трекерами. Поряд із

цим, необхідно зазначити простоту реалізації, низьку обчислювальну складність, швидкість і відповідність методу наявним фізичним процесам.

3.5 Висновки по розділу 3

В розділі 3 виконано експериментальне визначення формантних частот із використанням спектрального розкладання мовного сигналу.

Виділення формант супроводжує ряд проблем, пов'язаних з їхньою динамічною зміною у процесі мовлення. Навіть однакові голосні змінюють формантний набір залежно від свого розташування у складі слів, складів тощо. Визначення розташування формант на спектрограмах мовного сигналу достатньо легко виконується людиною, але автоматизація цього процесу визиває деякі труднощі.

Досліджено алгоритм динамічного виділення формантних частот. Він складається з таких етапів: сегментація мовного сигналу на вокалізовані фрагменти та паузи; розбиття вокалізованих фрагментів сигналу на часові фрейми; для кожного фрагмента отримання спектра на основі вейвлет-перетворення; побудова обвідної спектра; знаходження всіх максимумів; визначення положень формантних діапазонів; отримання максимумів формантних діапазонів; побудова графіків траєкторії положення формант; розрахунок залежності щільності ймовірності розподілу кожної із чотирьох формантних частот.

Авторегресійні коефіцієнти визначалися за допомогою автокореляційного методу. Визначаючи максимуми обвідної спектра, отримуємо максимуми чотирьох перших формант. Аналіз отриманих спектрів показує достатньо високу стабільність визначення формантних частот для досліджених мовних сигналів.

Для порівняльного дослідження розглядались знамениті формант-трекери PRAAT і SNACK. Налаштування кожного з них здійснювалося на основі набору параметрів за замовчуванням, що було закладено

розробниками цих трекерів. Налаштування часових інтервалів досліджуваного методу має відповідати розміру подібних інтервалів для інших формант-трекерів для коректного порівняння, тому використано розмір вікна 25 мс із кроком 10 мс.

У дослідженні трекери самостійно виконували сегментацію на вокалізовані фрагменти і паузи, застосовуючи датасет VTR-TIMIT, у випадку некоректної сегментації (невокалізований фрагмент вважався вокалізованим) помилкові результати сегментації вилучалися з розгляду.

Як параметри порівняння використовували середньоквадратичне відхилення формант між еталонною розміткою мовних сигналів та результатами формант-трекерів.

Порівняльний аналіз показує достатньо високу точність визначення формантних частот порівняно з існуючими формант-трекерами. Поряд із цим, необхідно зазначити простоту реалізації, низьку обчислювальну складність, швидкість і відповідність дослідженого методу наявним фізичним процесам мовлення.

ВИСНОВКИ

Задача аналізу голосу людини виникає в багатьох застосуваннях. Найпоширеніша сучасна задача – це розпізнавання мови, тобто процес перетворення мовного сигналу у текстовий потік. Також на основі параметрів мови відбувається зворотна операція – синтез мовних сигналів.

Іншою задачею є аналіз голосового апарату людини. Це здійснюється в лікарняних цілях, або для виявлення і класифікації голосових патологій, або для аналізу стану вокального апарату співаків.

Також аналіз голосу застосовується в судовій криміналістиці для розпізнавання статі, віку, емоційного стану людини, а також особи розмовника.

Метою кваліфікаційної роботи є теоретичний аналіз і практичне дослідження деяких методів оцінки параметрів голосу з метою визначення їх переваг і недоліків для різних застосувань. Результати, отримані в даній роботі зможуть бути застосованими в лабораторному практикумі з дисципліни «Методи обробки звукової інформації».

В розділі 1 розглянуті моделі формування мовлення – акустична і модуляційна. Акустична модель містить три основні складові: енергетичну, що генерує потік повітря, резонаторну частину, що формує частотну характеристику звуку, і артикуляційну частину, що формує обвідну звуку. Аналізуючи звук, можна отримати параметри моделі, що дозволить вирішувати багато прикладних задач.

В другому розділі розглянуті методи оцінювання параметрів мовного тракту по звуковому сигналу. Зазвичай визначають такі параметри: частоту основного тону, короткочасну енергію, формантні частоти. Також відслідковують траєкторію цих параметрів в часі.

Першими розглянуті непараметричні оцінки, основані на швидкому перетворенні Фур’є з вікном Хемінга. Реальний мовний сигнал має кінцеву тривалість, при поданні в частотній області спектр необмежений. Тому

сигнал сегментують на ділянки близько 10 мс, на яких вважається стаціонарним.

Отримано спектральні оцінки чоловічого і жіночого голосів при промові фонем «а» та «і» у високому і низькому спектральному розрізненні. Високе спектральне розрізнення звучно використовувати при оцінці частоти основного тону. Низьке – при оцінюванні формантних резонансних частот звукового тракту. Отримані оцінки приблизно співпадають з відомими даними для української мови. Відмінність у результатах спостерігається в оцінці формант і викликано багатомодовістю спектру Фур'є навіть при низькому спектральному розрізненні, що ускладнює оцінку формант.

Розроблено структурну схему обробки звукового сигналу для оцінки зазначених вище параметрів. Виділення фраз відбувається таким шляхом. У межах обраного часового сегмента обчислюється середнє значення енергії шуму E і поріг P, який береться рівним подвоєної енергії шуму. При подальшій обробці, якщо середнє значення енергії перевищило поріг, фіксується момент запису мовного сигналу (початок фрази), який запам'ятовується. Якщо середнє значення енергії поменшає порога, то запам'ятовується кінець фрази.

Досліджено параметричний метод спектрального оцінювання. Він складається із трьох етапів: спочатку проводиться вибір параметричної моделі часового ряду. Вибрана модель авторегресії дає спектри з гострими піками. На другому етапі обчислюються оцінки параметрів моделі. На третьому етап оцінені значення параметрів вводяться у вираз для спектральної щільності потужності, що відповідає обраній моделі.

Отримано графік параметричної оцінки спектру звуків "а" і «і» української мови. Оцінки формантних частот набагато краще відповідають відомим в літературі результатам, ніж непараметричні оцінки.

Нарешті, розглянуто кепстральний аналіз. Кепстр – це зворотне перетворення Фур'є від натурального логарифму квадрата спектральної щільності випадкового процесу, що відображається в вигляді функції $C(q)$

від так званого кепстрального часу q. Параметр q має розмірність часу, у кепстральному аналізі він умовно називається сачтотою, а його застосування забезпечує рознесення результуючих енергетичних сплесків по осі кепстрального часу. Як правило, енергетичні сплески в реальних звукових сигналах розміщуються вздовж осі сачтот q з віддаленням від нульової позначки.

Отримано результати обчислення обвідної спектру звуку «і» через спектр цього звуку, а також кепстр звуку «і». Кепстр можна використовувати як вхідні дані для нейронних мереж, які розпізнають звук.

В середовищі MATLAB розроблено ряд программ, що дозволяють обчислювати кепстри, спектрограми, і параметричні спектри. Ці програми були використані в експериментальних дослідженнях.

В розділі 3 виконано експериментальне визначення формантних частот із використанням спектрального розкладання мовного сигналу.

Виділення формант супроводжує ряд проблем, пов'язаних з їхньою динамічною зміною у процесі мовлення. Навіть однакові голосні змінюють формантний набір залежно від свого розташування у складі слів, складів тощо. Визначення розташування формант на спектрограмах мовного сигналу достатньо легко виконується людиною, але автоматизація цього процесу визиває деякі труднощі.

Досліджено алгоритм динамічного виділення формантних частот. Він складається з таких етапів: сегментація мовного сигналу на вокалізовані фрагменти та паузи; розбиття вокалізованих фрагментів сигналу на часові фрейми; для кожного фрагмента отримання спектра на основі вейвлет-перетворення; побудова обвідної спектра; знаходження всіх максимумів; визначення положень формантних діапазонів; отримання максимумів формантних діапазонів; побудова графіків траєкторії положення формант; розрахунок залежності щільності ймовірності розподілу кожної із чотирьох формантних частот.

Авторегресійні коефіцієнти визначалися за допомогою автокореляційного методу. Визначаючи максимуми обвідної спектра, отримуємо максимуми чотирьох перших формант. Аналіз отриманих спектрів показує достатньо високу стабільність визначення формантних частот для досліджених мовних сигналів.

Для порівняльного дослідження розглядались знамениті формант-трекери PRAAT і SNACK. Налаштування кожного з них здійснювалося на основі набору параметрів за замовчуванням, що було закладено розробниками цих трекерів. Налаштування часових інтервалів досліджуваного методу має відповідати розміру подібних інтервалів для інших формант-трекерів для коректного порівняння, тому використано розмір вікна 25 мс із кроком 10 мс.

У дослідженні трекери самостійно виконували сегментацію на вокалізовані фрагменти і паузи, застосовуючи датасет VTR-TIMIT, у випадку некоректної сегментації (невокалізований фрагмент вважався вокалізованим) помилкові результати сегментації вилучалися з розгляду.

Як параметри порівняння використовували середньоквадратичне відхилення формант між еталонною розміткою мовних сигналів та результатами формант-трекерів.

Порівняльний аналіз показує достатньо високу точність визначення формантних частот порівняно з існуючими формант-трекерами. Поряд із цим, необхідно зазначити простоту реалізації, низьку обчислювальну складність, швидкість і відповідність дослідженого методу наявним фізичним процесам мовлення.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Yegnanarayana, B., Veldhuis, R. N. J. Extraction of vocaltract system characteristics from speech signals, IEEE Trans. Speech Audio Process, 6 (4), 1998, 313–327.

2. Kim, C., Seo, K., & Sung, W. A Robust Formant Extraction Algorithm Combining Spectral Peak Picking and Root Polishing. EURASIP Journal on Applied Signal Processing, 2006, 1–16.

3. Wet, F. D., Weber, K., Boves, L., Cranen, B., Bengio, S., & Bourlard, H. (2004). Evaluation of Formant-Like Features for Automatic Speech Recognition. Journal of the Acoustical Society of America, 116, 1781–1791.

4. Mallat, S. A Wavelet Tour of Signal Processing. Academic Press. 1999.

5. Yan, Q., Vaseghi, S., Zavarehei, E., Milner, B., Darch, J., White, P., & Andrianakis, I. Formant Tracking Linear Prediction Model using HMMs and Kalman Filters for Noisy Speech Processing. Computer Speech and Language, vol. 21, Jul. 2007, pp. 543–561.

6. Messaoud, Z. B., Gargouri, D., Zribi, S., & Hamida, A. B.. Formant Tracking Linear Prediction Model using HMMs for Noisy Speech Processing. International Journal of Signal Processing, vol. 5, 2009, pp. 291–296.

7. Cooke, M., Barker, J., Cunningham, S., & X. Shao. An audio-visual corpus for speech perception and automatic speech recognition. Journal of the Acoustical Society of America, vol. 120. 2006.

8. Acero, A. Formant Analysis and Synthesis using Hidden Markov Models. In Proc. of the Eurospeech Conference. Budapest. 1999.

9. Veldhuis, R.. A computationally e$cient alternative for the LF model and its perceptual evaluation. J. Acoust. Soc., 103 (1), 1997, 566–571.

10. Bazzi, I., Acero, A., & Deng, L. An expectation maximization approach for formant tracking using a parameter-free non-linear predictor. In Proc. ICASSP, vol. 1, 2003, 464–467.

11. Ali, J. A. M. A., Spiegel, J. V. D., & Mueller P. Robust Auditory-based Processing using the Average Localized Synchrony Detection. In IEEE Transaction Speech and Audio Processing. 2002.

12. Vakman, D. On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency. IEEE Trans. Signal Process, SP-44, 1996, 791–797.

13. Semenets V, Kartashov V, Sergiyenko O, Tikhonov V, Mercorelli P, Sheiko S, Kudriavtseva NC, Rodriguez-Quinonez JC, Flores-Fuentes W. The Use of Factorization and Multimode Parametric Spectra in Estimating Frequency and Spectral Parameters of Signal. In: IEEE International Symposium on Industrial Electronics [Internet]; 2020, p. 215-9.

14. Kartashov VM, Tikhonov VA, Voronin VV. Features of construction and application of complex systems for the atmosphere remote sounding. Telecommun Radio Eng [Internet]. 2017; 76(8): 743-9.

15. Тихонов В. А., Карташов В. М., Олейников В. Н., Леонидов В. И. Обнаружение-распознавание беспилотных летательных аппаратов с использованием составной модели авторегрессии их акустического излучения. Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia, 2020, Iss. 75.

16. Карташов В.М., Тихонов В.А., Воронин В.В., Кошевой В.В. Подавление акустических помех в системах дистанционного мониторинга атмосферы с использованием решетчатых фильтров// Інформаційно-керуючі системи на залізничному траспорті. – 2019. – №2 (135). – С. 40-48.

17. Карташов В.М., Тихонов В.А., Воронин В.В., Селезнев И.С. Автогрегрессионные фильтры подавления помех в системах акустического зондирования атмосферы// Радіотехніка (Харків). – 2019.– Вип. 196. – С. 106-111.

18. Зибін С., Бєлозьорова Я. Метод визначення формантних частот із використанням спектрального розкладання мовного сигналу / Безпека інформаційних систем і технологій, № 1(6), 2023. – с.51 – 60.

19. Boersma, P., & D. Weenink. Praat: doing phonetics by computer [Електронний ресурс]. URL: http://www.praat.org/ [Дата доступу 17.11.2023].

20. Kåre Sjölander. The Snack Sound Toolkit [Електронний ресурс]. URL: https://www.speech.kth.se/snack/ [Дата доступу 17.11.2023].

21. Методичні вказівки з виконання атестаційної магістерської роботи за спеціальністю 8.05090102 «Апаратура радіозв'язку, радіомовлення і телебачення». Освітньо-кваліфікаційний рівень – магістр / Упоряд. В.М. Карташов, В.А. Тихонов, І.В. Савченко – Харків: ХНУРЕ, 2012. – 68 с.