

Міністерство освіти і науки України  
Національний авіаційний університет

Кваліфікаційна наукова праця  
на правах рукопису

**БЕЛОЗЬОРОВА ЯНА АНДРІЇВНА**

УДК 004.42: 004.93

**МЕТОД ЗАСТОСУВАННЯ ВЕЙВЛЕТ АНАЛІЗУ  
В ЗАДАЧАХ ІДЕНТИФІКАЦІЇ МОВНОЇ ІНФОРМАЦІЇ**

01.05.03 – математичне та програмне забезпечення  
обчислювальних машин та систем

Подається на здобуття наукового ступеня кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей,  
результатів і текстів інших авторів мають посилання на відповідне джерело  
\_\_\_\_\_ Белозьорова Я.А.

Науковий керівник – **Зибін Сергій Вікторович**,  
доктор технічних наук, доцент

Київ – 2021

## АНОТАЦІЯ

**Бєлєзьєрєвє Я.А. Мєтєд зєстєсєвєннє вєйвлєт єнєлїзє в зєдєчєх їдєнтїфїкєцї мєвнєї їнфєрмєцїї.** – Нє прєвєх рєкєпїсє.

Дїсєртєцїє нє здєбєттє нєкєвєгє стєпєнє кєндїдєтє тєхнїчнїх нєкє зє спєцїєлїнїстєю 01.05.03 – мєтємєтїчнє тє прєгрємнє зєбєзпєчєннє обчїслєвєлїнїх мєшїн ї сїстєм. – Нєцїєнєлїнїй євїєцїйнїй унївєрсїтєт, Кїїв, 2021.

Дїсєртєцїйнє рєбєтє прїсвєчєнє рєзрєбцї нє єснєвї єпєрєтє фрєктєлїнєгє тє вєйвлєт єнєлїзє мєдєлї тє мєтєдє, щє здїйснєють єбрєбкє, вїзнєчєннє їндивїдєлїнїх хєрєктерїстїк єсєбї тє їдєнтїфїкєцїє мєвнєгє сїгнєлє. У рєбєтї єбгрєнтєвєнє нєєбхїднїстє вїкєрїстєннє фрєктєлїнєгє тє вєйвлєт єнєлїзє длє їдєнтїфїкєцїї мєвнєї їнфєрмєцїї. Длє сєгмєнтєцїї мєвнєгє сїгнєлє рєзрєблєнє мєтєд, щє вїкєрїстєвєє фрєктєлїнї хєрєктерїстїкї сїгнєлє. Зєстєсєвєннє мєтєдє дєзвєлєє вїдїлєтї вєкєлїзєвєнї тє нєвєкєлїзєвєнї фрєгмєнтї сїгнєлє, нєзєлєжнє вїд мєвнїх дєнїх єсєбї. Зєпрєпєнєвєнє єлгєрїтм вїдїлєннє хєрєктерїстїк сємєпєдїбнїх стрєктєр в мєвнємє сїгнєлї нє єснєвї єнєлїзє мєкєсїмєлїв вєйвлєт-пєрєтвєрєннє нє рїзнїх рївнєх дєкємпєзїцїї мєвнєгє сїгнєлє, щє дєзвєлєє вїзнєчїтї пєрємєтрї чєстєтї єснєвнєгє тєнє тє чєстєт фєрмєнт у вїглєдї крївїх щїлїнєстї рєзпєдїлє вїрєгїднєстї. Рєзрєблєнє мєтєд їдєнтїфїкєцїї єсєбї, щє вїкєрїстєвєє мєтєд сєгмєнтєцїї мєвнєгє сїгнєлє тє єлгєрїтм вїдїлєннє хєрєктерїстїк сємєпєдїбнїх стрєктєр. Нє єснєвї вїкєнєнєгє дєслїджєннє ствєрєнє прєгрємнє сїстємє, єкє пє мєвнїм зєпїсєм здїйснєє євтємєтїчнїй рєзрєхєнєк мєвнїх хєрєктерїстїк, вїкєнєє рєнжєвєннє цїх хєрєктерїстїк в бєзї дєнїх, зє вїзнєчєнїмї в рєбєтї крїтєрїємї, тє їдєнтїфїкєцїє єсєбї в мєвнємє сїгнєлї.

**Клєчєвї слєвє:** мєвнїй сїгнєл, вєйвлєт пєрєтвєрєннє, фрєктєлїнє рєзмїрнїстє, чєстєтє єснєвнєгє тєнє, сєгмєнтєцїє сїгнєлє, вєкєлїзєвєнїй фрєгмєнт.

## АННОТАЦИЯ

**Белозерова Я.А. Метод использования вейвлет анализа в задачах идентификации голосовой информации.** – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 01.05.03 – математическое и программное обеспечение вычислительных машин и систем. – Национальный авиационный университет, Киев, 2021.

Диссертационная работа посвящена разработке на основе аппарата фрактального и вейвлет анализа модели и метода, осуществляющих обработку, определение индивидуальных характеристик личности и идентификацию голосового сигнала. В работе обоснована необходимость использования фрактального и вейвлет анализа для идентификации голосовой информации. Для сегментации голосового сигнала разработан метод, который использует фрактальные характеристики сигнала. Применение метода позволяет выделять вокализованные и невокализованные фрагменты сигнала, независимо от голосовых данных личности. Предложен алгоритм выделения характеристик самоподобных структур в голосовом сигнале на основе максимумов вейвлет-преобразования на разных уровнях декомпозиции голосового сигнала, что позволяет определить частоту основного тона и частот формант в виде кривых плотности распределения вероятности. Разработанный метод идентификации личности, использует метод сегментации голосового сигнала и алгоритм выделения характеристик самоподобных структур. На основе выполненного исследования создана программная система, которая по голосовым записям осуществляет автоматический расчет голосовых характеристик, выполняет ранжирование этих характеристик в базе данных, по определенным в работе критериям, и идентификацию личности в голосовом сигнале.

**Ключевые слова:** программное обеспечение, голосовой сигнал, вейвлет преобразование, фрактальная размерность, частота основного тона, сегментация сигнала, вокализованный фрагмент.

## ANNOTATION

**Bielozorova Y.A. Method of application of wavelet analysis in problems of identification of speech information.** – As a manuscript.

The thesis for obtaining a scientific degree of the candidate of technical sciences on the speciality 01.05.03 – mathematical and software of computing machines and systems. – National Aviation University, Kyiv, 2021.

The thesis is devoted to the development of the model and method that process and determine the individual characteristics of the person and the identification of the speech signal. Models and methods are based on the apparatus of fractal and wavelet analysis. The paper substantiates the need to use fractal and wavelet analysis to identify speech information. The method using fractal signal characteristics has been developed for speech signal segmentation. The application of the method allows to distinguish vocalized and unvocalized fragments of the signal. This happens regardless of the person's language data. The algorithm for selecting parameters of self-similar structures in a speech signal has been proposed, it based on wavelet transform maxima at different levels of voice signal decomposition. This allows us to determine the frequency parameters of the fundamental tone and the formant envelop in the form of probability density curves. The method of personal identification has been developed. This method uses the method of speech signal segmentation and the algorithm for selecting parameters of self-similar structures. The software system was created on the basis of the performed research. This software system automatically calculates the voice characteristics of the voice records. In addition, it performs the ranking of these characteristics in the database, according to the criteria, which was defined in the work and the identification of the person in the speech signal.

**Key words:** speech signal, wavelet transform, fractal dimension, fundamental frequency, signal segmentation, vocalized fragment.

## СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

*Статті у наукових фахових виданнях України:*

1. Соловьев В.И., Брюханова Я.А. Идентификация заданных фрагментов в звуковых файлах / В.И.Соловьев, Брюханова Я.А. // Науковий журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2008. – №9(127) ч.2. – С. 30 – 33. **Внесок автора:** запропоновано структуру аналізу мовного файлу та ідентифікацію заданих фрагментів на основі інформативного компонента та його обмежувача.

2. Рыбальский О.В., Белозеров Е.В., Соловьев В.И., Белозерова Я.А. Методология проверки подлинности сигналограмм выделением самоподобных структур / О.В. Рыбальский, Е.В. Белозеров, В.И. Соловьев, Я.А. Белозерова // Научно-технический журнал. Захист інформації. – 2010. – № 2(47). – С. 35 – 43. **Внесок автора:** запропоновано метод представлення мовного сигналу та ідентифікації стійких самоподібних (мультифрактальних) структур на основі використання вейвлет аналізу.

3. Соловьев В.И., Белозерова Я.А. Использование фрактальной размерности аудиофайлов в задаче сегментации звукового файла / В.И. Соловьев., Я.А. Белозерова // Науковий журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2013. – № 5(194) ч.2. – С. 165 – 168. **Внесок автора:** запропоновано метод сегментації мовного сигналу на основі використання фрактальної розмірності.

4. Соловьев В.И., Белозерова Я.А. Анализ алгоритмов построение системы идентификации диктора / В.И. Соловьев., Я.А. Белозерова // Науковий журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2013. – № 6(195) ч.1. – С. 62 – 66. **Внесок автора:** запропоновано структуру системи тектонезалежної ідентифікації мовного

сигналу, критерії та алгоритми визначення основних параметрів мовного сигналу.

5. Белозерова Я.А. Выделение идентификационных характеристик диктора на основе вейвлет-разложения голосового сигнала / Я.А. Белозерова // Науковий журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2015. – № 1(218). – С. 62 – 66.

*Статті у наукових фахових виданнях України, які входять до міжнародних наукометричних баз даних:*

6. Белозьорова Я.А. Ідентифікація диктора на основі кратномасштабного аналізу / Я.А. Белозьорова // Науковий журнал. Інженерія програмного забезпечення. – 2017. – № 1(29). – С. 15 – 25.

7. Зибін С.В., Белозьорова Я.А. Побудова архітектури програмної системи ідентифікації диктора / С.В., Зибін, Я.А. Белозьорова // Науковий журнал. Інженерія програмного забезпечення. – 2017. – № 2(30). – С. 24 – 31.  
**Внесок автора:** запропоновано загальна структура, особливості побудови програмної системи ідентифікації мовного сигналу та її практична реалізація.

8. Зибін С.В., Белозьорова Я.А. Поліпшення точності ідентифікації мовного сигналу за рахунок уточнювального підходу до підбору характеристик ідентифікації / С.В., Зибін, Я.А. Белозьорова // Науковий журнал. Інженерія програмного забезпечення. – 2017. – № 3(31). – С. 34-38.  
**Внесок автора:** визначено критерії втрати точності при ідентифікації мовного сигналу, запропоновано метод ідентифікації мовного сигналу.

*Статті у закордонних виданнях, які входять до міжнародних наукометричних баз даних:*

9. Solovyov Victor, Byelozorova Yana. Multifractal approach in pattern recognition of an announcer's voice / Victor Solovyov, Yana Byelozorova // ТЕКА.

Commission of motorization and energetics in agriculture. – 2014. – Vol. 14, № 2. – pp. 164 – 170. **Внесок автора:** запропоновано мовна модель, що використовує самоподібні (мультифрактальні) структури, доведено можливість компонування фонем як атомарних структур на основі афінних перетворень.

10. Белозьорова Я.А. Виділення самоподібних структур мовних сигналів в задачах ідентифікації диктора / Я.А. Белозьорова // Scientific Journal «ScienceRice». – 2017. – Vol.5. – №2 (34). – pp. 22 – 27.

11. Bielorozova Yana. Analyse and develop the software of automatic search for an anonymous person in the voice database / Yana Bielorozova // International Journal “Information Technologies & Knowledge”. – 2019. – Vol.13. – № 2. – pp. 152 – 164.

12. Zybin Serhii, Bielorozova Yana. Practical approach to speech identification / Serhii Zybin, Yana Bielorozova // International Journal “Information models & analyses”. – 2020. – Vol.9. – № 3. – pp. 224 – 231. **Внесок автора:** розроблено критерії ідентифікації особи в мовному сигналі та визначено практичні рекомендації виділення ознак особи в мовному сигналі.

13. Zybin Serhii, Bielorozova Yana. Risk-based decision-making system for information processing systems / Serhii Zybin, Yana Bielorozova // International Journal of Modern Education and Computer Science (IJMECS). – 2021. – Vol.13. – № 4. – pp. 34 – 48. **Внесок автора:** розроблено підхід до прийняття рішень в складних системах, що може використовуватися при побудові систем ідентифікації мовної інформації.

*Тези наукових конференцій:*

14. Белозеров Е.В., Белозерова Я.А. Оценка аутентичности цифровых медиа-файлов с использованием особенностей устройств их формирования / Е.В. Белозеров, Я.А. Белозерова // III Міжнародна науково-технічна

конференція «Комп'ютерні системи та мережні технології», Київ, 15-17 червня 2010 р. – Київ: Національний авіаційний університет, 2010. – С. 17.

15. Соловьев В.И., Белозерова Я.А. Модель слуховой системы человека для задач идентификации диктора / В.И. Соловьев, Я.А. Белозерова // XVI Международная научно-практическая конференция «Современные информационные и электронные технологии», Одесса, 25-29 мая 2015 г. – Одеса: Одеський національний політехнічний університет, 2017. – С. 46.

16. Белозьорова Я.А. Ідентифікація диктора на основі поліпшеного алгоритму виділення частоти основного тону / Я.А. Белозьорова // Міжнародна науково-технічна конференція «Інтелектуальні технології лінгвістичного аналізу», Київ, 25-25 жовтня 2017 р. – Київ: Національний авіаційний університет, 2017. – С. 9.

17. Белозьорова Я.А. Особливості побудови системи ідентифікації диктора на основі мультифрактального підходу / Я.А. Белозьорова // Всеукраїнська науково-практична конференція «Безпека соціально-економічних процесів в кіберпросторі», Київ, 27 березня 2019 р. – Київ: Київський національний торговельно-економічний університет, 2019. – С. 126-127.

18. Bielozorova Yana. Analyse and develop the software of automatic search for an anonymous person in the voice database / Yana Bielozolorova // International Conference on Software Engineering, Kyiv, June 03 – 06, 2019. – Kyiv: National Aviation University, 2019. – pp. 60 – 64.

19. Bielozorova Yana. Practical approach to building a speaker identification system / Yana Bielozolorova // International Conference on Software Engineering «SoftEngine 2020», Kyiv, April 13 – 15, 2020. – Kyiv: National Aviation University, 2020. – pp. 59 – 61.

20. Serhii Zybin, Yana Bielozorova. Automatic search for involvants of information messages in the voice databases / Serhii Zybin, Yana Bielozorova // International Conference on Software Engineering «SoftEngine 2021», Kyiv, April 12 – 14, 2021. – Kyiv: National Aviation University, 2021. – pp. 47 – 50.



21. Белозьорова Я.А., Зибін С.В. Критерії побудови системи мовної ідентифікації особи / Я.А. Белозьорова, С.В. Зибін // 13 Всеукраїнська науково-практична конференція «Стан та удосконалення безпеки інформаційно-телекомунікаційних систем «SITS 2021»», Миколаїв – Коблево, 23 – 25 червня 2021 р. С. 74 – 76.

*Патент на винахід (корисну модель)*

22. Пат. 76105 Україна, МПК G 11 b 27/00, 27/36. Спосіб виявлення слідів цифрової обробки цифрових графічних сигналів [Текст] / Я.А. Белозьорова, М.О. Сидоров, К.С. Лебеденко., Є.В. Белозьоров; заявник и патентовласник: Національний авіаційний університет. – заявл. 31.05.2012; опубл. 25.12.2012, Бюл. №24.

## ЗМІСТ

ВСТУП.....	13
РОЗДІЛ 1. АНАЛІЗ СУЧАСНОГО СТАНУ І НАПРЯМКІВ РОЗВИТКУ ДОСЛІДЖЕНЬ ЗАДАЧ ІДЕНТИФІКАЦІЇ МОВНОЇ ІНФОРМАЦІЇ.....	19
1.1. Аналіз механізмів утворення мовного сигналу .....	19
1.2. Аналіз механізмів сприйняття мовного сигналу .....	22
1.3. Огляд характеристик опису мовного сигналу .....	26
1.4. Огляд методів визначення частоти ОТ .....	30
1.5. Огляд методів сегментації мовного сигналу .....	33
1.6. Прийняття рішення в задачах ідентифікації мовної інформації.....	38
1.7. Постановка мети і задач дослідження.....	41
Висновки за 1 розділом.....	43
РОЗДІЛ 2. РОЗРОБКА МОДЕЛІ СЛУХОВОЇ СИСТЕМИ ТА МОДЕЛІ ПРЕДСТАВЛЕННЯ МОВНОГО СИГНАЛУ .....	45
2.1. Модель слухової системи людини .....	45
2.2. Фрактальні структури, як елементарні складові мовного сигналу.....	48
2.3. Обґрунтування вибору моделі представлення мовного сигналу.....	56
2.4. Модель представлення мовного сигналу на основі вейвлет аналізу та визначення характеристик його самоподібних структур.....	58
Висновки за 2 розділом.....	65
РОЗДІЛ 3. РОЗРОБКА МЕТОДУ ТА АЛГОРИТМІВ, НЕОБХІДНИХ ДЛЯ ПОБУДОВИ СИСТЕМИ ІДЕНТИФІКАЦІЇ МОВНОЇ ІНФОРМАЦІЇ .....	66
3.1. Вибір методу розрахунку фрактальної розмірності .....	66
3.2. Розробка методу сегментації на основі фрактальних характеристик сигналу.....	72

3.3. Розробка алгоритму виділення характеристик самоподібних структур в якості ідентифікаційних ознак мовного сигналу.....	75
3.4. Визначення частотних характеристик формант .....	85
3.5. Порівняння характерних ознак мовного сигналу .....	95
Висновки за 3 розділом.....	98
<b>РОЗДІЛ 4. ПІДБІР ПАРАМЕТРІВ ДЛЯ ПОБУДОВИ МЕТОДУ ІДЕНТИФІКАЦІЇ МОВНОЇ ІНФОРМАЦІЇ.....</b>	<b>100</b>
4.1. Ідентифікація мовного сигналу як клас задач загальної ідентифікації.....	100
4.2. Вибір параметрів та дослідження ефективності методу сегментації ..	102
4.3. Вибір параметрів та дослідження ефективності методу підвищення інформативності частоти основного тону .....	108
4.4. Оцінка ефективності алгоритму визначення формантних частот .....	117
4.5. Створення програмної системи мовної ідентифікації особи .....	119
Висновки за 4 розділом.....	127
<b>ВИСНОВКИ .....</b>	<b>129</b>
<b>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....</b>	<b>131</b>
<b>ДОДАТКИ</b>	

**ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ**

МС – мовний сигнал

ОТ – основний тон

ПС – програмна система

ШФП – швидке Фур'є перетворення

ШНМ – штучні нейронні мережі

БВП – безперервне вейвлет-перетворення

MFCC – Mel-frequency Cepstral Coefficient

CMM – Coerced Markov Model

LPC – Linear Prediction Coefficient

DTW – Dynamic Time Warping

NIST – National Institute of Standards and Technology

CER – Crossover Error Rate

FRR – False Reject Rate

FAR – False Accept Rate

SDI – Standard Deviation Index

UML – Unified Modeling Language

SQL – Structured Query language

## ВСТУП

**Актуальність теми.** Створення алгоритмів та програмних засобів для систем комп'ютерного розпізнавання мовної інформації є частиною галузі математичного та програмного забезпечення обчислювальних машин і систем. Одним з важливих напрямків досліджень є розробка інтелектуальних прикладних програмних систем образного сприйняття мовної інформації, серед яких значну роль відіграють системи розпізнавання мовних сигналів (МС).

Дослідженню мовного апарата і математичному обґрунтуванню частотних характеристик звуків мовлення були присвячені роботи А. Бела, Г. Фанта, Д. Фланагана, Р. Якобсона та ін. Поява ЕОМ сприяла необхідності розвитку методів цифрової обробки МС. Важливу роль у цій області зіграли роботи Д. Габора, Б. Гоулда, Д. Маркела, Л. Рабінера, Д. Рейді, Р. Шафера, Б. Янга, Г.Ф. Конаховича та ін. Значний вклад у розвиток технологій розпізнавання МС внесли відомі вчені Х. Сакое, С. Чіба, Ф. Ітакура, В.М. Величко, Н.Г. Загоруйко, В.М. Сорокін, М.А. Сапожков, О.В Рибальський, В.І. Соловійов. Розроблені методи базувалися, в основному, на статистичному підході з використанням прихованих Марківських ланцюгів, критерію максимальної правдоподібності та байєсовських правил.

Проблеми, що виникають при розпізнаванні МС, пов'язані з варіативністю сигналу, шумом навколишнього середовища та звукозаписуючого обладнання. Для опису локальних особливостей неоднорідних сигналів і зниження рівня шуму ефективно використовується вейвлет-перетворення, теоретичні основи якого були викладені у працях А.Гроссмана, Ж. Морле, І. Добеші, С. Малла, І. Мейера, Ч. Чуї та ін. Використання вейвлетів може значно розширити алгоритмічну і методичну базу для створення інформаційних технологій обробки і аналізу МС.

Дана дисертаційна робота присвячена розробці методу, спрямованого на підвищення точності розпізнавання МС в умовах шуму, а також пошуку параметрів, описуючих акустичні характеристики мови особистості на основі

вейвлет аналізу та інваріантних до інтенсивності сигналу. Саме в цьому полягає актуальність дисертаційної роботи.

**Зв'язок роботи з науковими програмами, планами, темами.**

Дисертацію виконано в Національному авіаційному університеті на кафедрі інженерії програмного забезпечення в рамках таких науково-дослідних робіт, у яких Белозьорова Я.А. брала участь: № 29/09.01.02 «Онтології в інженерії програмного забезпечення», № 58/09.01.02 «Методологія підвищення ефективності процесів життєвого циклу розробки програмного забезпечення у гнучких підходах його розробки» (2019-2022р.).

**Мета та задачі дослідження.** Метою дисертаційної роботи являється розробка, на основі апарату фрактального та вейвлет аналізу, моделі і методів, що здійснюють обробку, визначення індивідуальних характеристик особи та ідентифікацію МС.

Для досягнення поставленої мети необхідно вирішити наступні **задачі**:

1. Провести аналіз математичних моделей формування та сприйняття МС для задач їх цифрової обробки.

2. Використовуючи фрактальний аналіз МС, розробити метод його сегментації.

3. На базі вейвлет-аналізу розробити алгоритм виділення характеристик самоподібних структур в МС, що можуть використовуватись в якості унікальних ознак при ідентифікації особи.

4. На основі методу сегментації МС та алгоритму виділення характеристик самоподібних структур в МС розробити метод ідентифікації особи.

5. Провести експериментальні дослідження методів та алгоритму, для мовної ідентифікації особи, проаналізувати та порівняти результати з відомими.

6. Створити програмну систему мовної ідентифікації особи та провести порівняльний аналіз з існуючими аналогами.

**Об'єкт дослідження** – процес ідентифікації МС.

**Предмет дослідження** – моделі та методи цифрової обробки й ідентифікації МС.

**Методи дослідження.** У дисертаційній роботі використовуються: методи фрактального аналізу, для сегментації МС; алгоритми вейвлет-аналізу для одержання самоподібних структур, в якості ознак для класифікації особи; методи цифрової обробки МС, для зниження рівня шумів й одержання спектральних характеристик звуків мовлення; методи математичної статистики, для розв'язання задач класифікації самоподібних структур й мовної ідентифікації особи.

**Наукова новизна роботи** полягає в розробці нових, уточненні і доповненні наявних науково-методичних положень та практичних рекомендацій з ідентифікації мовного сигналу. У ході виконання дисертаційного дослідження були отримані наступні основні результати, що відображають наукову новизну роботи:

*уперше:*

– для мовної ідентифікації особи, в якості набору ознак було запропоновано самоподібні структури, що дозволило сформулювати новий спосіб представлення індивідуальних ознак для розпізнавання цієї особи та зменшити ймовірності виникнення похибок першого та другого роду при його ідентифікації за цими ознаками. Це дозволило досягти точності ідентифікації особи в 96,5 %, при побудові системи ідентифікації МС;

– запропоновано метод підвищення інформативності частоти основного тону (ОТ) для мовної ідентифікації особи, в якому, на відміну від існуючих, за ознаку для розпізнавання використано значення коефіцієнтів вейвлет-перетворення на відрізках МС, де спостерігаються екстремуми кореляційних функцій частоти ОТ. Це дозволило досягти ймовірності виникнення похибок другого роду 2,12 % та ймовірності виникнення похибок першого роду 4,59 % при розпізнаванні особи за цією ознакою;

*удосконалено:*

– знаходження границь вимовлених слів за рахунок використання широкої структурної класифікації, що дає змогу виділення вокалізованих

фрагментів та слів із МС при наявності короткочасних високоамплітудних перешкод;

*одержали подальший розвиток:*

– методи ідентифікації особи, основані на частотно-часовому аналізі, в результаті урахування індивідуальних характеристик особи, що реалізують ідентифікацію незалежно від типу тексту й інтенсивності МС.

**Практичне значення одержаних результатів** роботи полягає в тому, що впровадження в процес ідентифікації мовного сигналу запропонованих у дисертації моделей, методів, алгоритму та програмного засобу сприяє підвищенню точності ідентифікації мовного сигналу.

На базі створених методів та моделі спроектовано та реалізовано ПС для ідентифікації особи у мовному сигналі.

Результати дисертації у вигляді методів, алгоритму та програмного засобу впроваджені в таких організаціях: Національному авіаційному університеті – № 29/09.01.02 «Онтології в інженерії програмного забезпечення», № 58/09.01.02 «Методологія підвищення ефективності процесів життєвого циклу розробки програмного забезпечення у гнучких підходах його розробки» (2019-2022р.) (акт впровадження від 21.05.2021) та в навчальному процесі при проведенні лабораторних робіт згідно з програмами навчальних дисциплін «Мультимедійні, ігрові та розважальні системи», «Інтелектуальні системи», «Людино-машинна взаємодія» за напрямом професійної підготовки 121 «Інженерія програмного забезпечення» (акт впровадження від 21.05.2021); ТОВ «ІСТ ТАУЕР» шляхом застосування методів та алгоритму, що дозволило підвищити точність ідентифікації мовної інформації при побудові систем мовної людино-машинної взаємодії (акт впровадження від 18.05.2021); ТОВ «КіберБіонік Систематікс» шляхом застосування методу, що показав значну ефективність роботи і можливість легкої програмної реалізації, та може бути використаний при побудові систем мовної ідентифікації клієнтів для поліпшення якості та швидкості



обслуговування в контакт-центрах, системах доступу та інших програмних системах (акт впровадження від 21.05.2021).

1. Розроблено метод сегментації МС, що дозволяє автоматизувати процедуру структурної розмітки мовних баз даних, а також може бути використаний при ідентифікації МС.

2. Запропоновано алгоритм виділення характеристик самоподібних структур та розроблені методи обробки МС дозволяють використовувати їх:

- для побудови систем ідентифікації МС;
- для створення інтелектуальних систем взаємодії користувача й комп'ютера.

3. Розроблено програмну систему для цифрової обробки МС та ідентифікації мовної інформації особи.

**Особистий внесок здобувача.** Особистий внесок здобувача полягає у самостійному виконанні теоретичної і експериментальної частин роботи та інтерпретації одержаних результатів. Всі основні положення, теоретичні і практичні результати дисертаційної роботи, що виносяться на захист, отримані здобувачем особисто й опубліковані в одноосібних підготовлених працях [5, 6, 10, 11]. У роботах [1-4, 7-9, 12, 13], написаних у співавторстві, автору належить: структура аналізу мовного файлу, метод ідентифікації заданих фрагментів на основі інформативного компонента [1]; метод представлення мовного сигналу та ідентифікації стійких самоподібних (мультифрактальних) структур [2]; метод сегментації мовного сигналу на основі використання фрактальної розмірності [3]; структура системи текстонезалежної ідентифікації мовного сигналу, критерії та алгоритми визначення основних параметрів мовного сигналу [4]; особливості побудови програмної системи та її практична реалізація [7]; критерії втрати точності при ідентифікації мовного сигналу [8]; метод ідентифікації мовного сигналу, мовна модель на основі мультифрактальних структур, можливість компонування фонем як атомарних структур на основі афінних перетворень [9]; критерії ідентифікації особи в мовному сигналі, практичні рекомендації

виділення ознак особи в мовному сигналі [12]; підхід до прийняття рішень в складних системах, що може використовуватися при побудові систем ідентифікації мовної інформації [13]. Аналіз результатів окремих досліджень та аспектів прикладних застосувань проведено у співавторстві.

**Апробація роботи.** Результати роботи доповідались і обговорювались на: III Міжнародній науково-технічній конференції «Комп'ютерні системи та мережні технології» (CSNT-2010, м. Київ, 2010); XVI міжнародній науково-практичній конференції «Современные информационные и электронные технологии» (м. Одеса, 2015); міжнародній науково-технічній конференції «Інтелектуальні технології лінгвістичного аналізу» (м. Київ, 2017); Всеукраїнській науково-практичній конференції «Безпека соціально-економічних процесів в кіберпросторі» (м. Київ, 2019); International Conference on Software Engineering (Kyiv, 2019); International Conference on Software Engineering «SoftEngine 2020» (Kyiv, 2020); International Conference on Software Engineering «SoftEngine 2021» (Kyiv, 2021); 13 Всеукраїнської науково-практичної конференції «Стан та удосконалення безпеки інформаційно-телекомунікаційних систем «SITS 2021»» (м. Миколаїв – Коблево, 2021), а також на наукових семінарах Національного авіаційного університету.

**Публікації.** Основні результати дослідження опубліковано в двадцяти двох наукових роботах [1-22], тринадцять з яких [1-13] – у наукових фахових виданнях і вісім [14-21] – у збірниках тез доповідей наукових конференцій. З них одна [13] – у виданні, що індексується наукометричною БД Scopus, три [6, 7, 8] – у виданні, що індексується наукометричною БД Index Copernicus та чотири [9-12] – у виданнях зарубіжних країн, що входять до ЄС, а також вісім [5, 6, 10, 11, 16-19] є одноосібними і один патент України на корисну модель.

**Структура і обсяг.** Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Обсяг роботи становить – 153 сторінки, включаючи 57 рисунків, 7 таблиць, 4 доповнення, список використаних джерел – 138 одиниць.

## РОЗДІЛ 1

# АНАЛІЗ СУЧАСНОГО СТАНУ І НАПРЯМКІВ РОЗВИТКУ ДОСЛІДЖЕНЬ ЗАДАЧ ІДЕНТИФІКАЦІЇ МОВНОЇ ІНФОРМАЦІЇ

### 1.1. Аналіз механізмів утворення мовного сигналу

Акустичне мовне коливання породжується у результаті руху органів артикуляційного апарату, а також сприймається органами слуху людини [1,2]. Функціонування артикуляційного апарату в процесі мовотворення є предметом глибоких фізіологічних та акустичних досліджень, результати яких наведено [3,4]. Проте слухове сприйняття мовних та інших коливань вивчається дуже пильно, а також їх обробка на периферії органів слуху та на вищих рівнях сприйняття, що зв'язані з комбінаторною діяльністю мозку [5,6]. Не будемо зупинятися на більш докладному трактуванні усіх теоретичних положень, що існують у теперішній час, а також виявлених шляхом досліджень фактах, наведемо невеличкий опис основних відомостей об утворенні і сприйнятті мови особою. За утворення звуків мови відповідають наступні органи: рот, ніс, язик, піднебіння, глотка, гортань, голосові зв'язки, трахея, бронхи, легені та діафрагма (рис. 1.1).

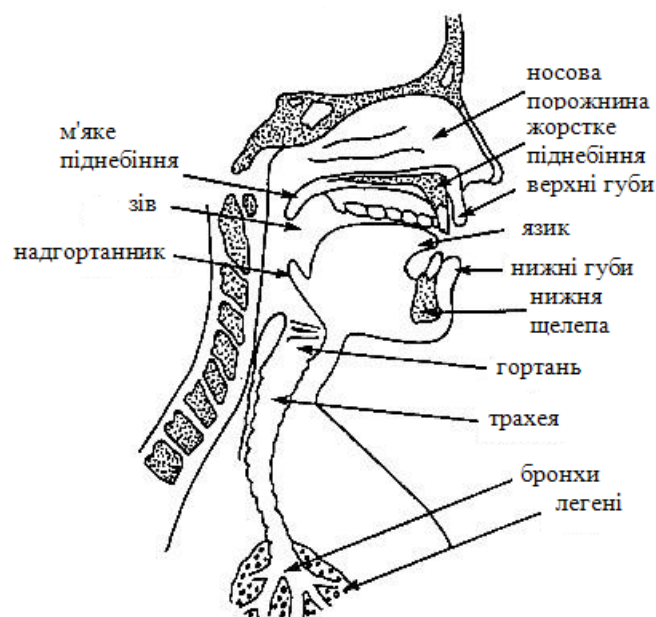


Рис. 1.1. Мовний тракт людини

Повітря, що зштовхається з легень, також проходить через трахею, гортань, порожнину глотки, рота і носа. Таким чином, мовне коливання являє собою акустичну хвилю [7], що розповсюджується по мовотворчій системі та випромінюється через губи і ніздрі.

Одну з головних ролей в утворенні звуків мови грають голосові зв'язки, що розташовані в гортані. При звичайному диханні голосові зв'язки розімкнуті і голосова щілина широко відкрита. Ця голосова щілина забезпечує прохід між зв'язками. При вимовленні деяких звуків мови зв'язки знаходяться в зімкнутому початковому стані і розмикаються лише під впливом тиску повітря, що подається з легень. Це повітря проривається через голосову щілину та розсовує зв'язки у поперечному напрямку. Під впливом м'язів, що звужують щілину, а також завдяки пружності та гідродинамічному ефекту Бернуллі, зв'язки знову повертаються у зімкнутий стан, а далі вже цей цикл повторюється [8]. Таких рух голосових зв'язок описує промову голосних та дзвінких приголосних звуків. Як результат цих коливань, повітря, що проходить через голосову щілину, носить імпульсний характер і далі вже надходить у глотку, ротову та носову порожнини. Гортань та ротова порожнина зазвичай має назву мовний тракт. Конфігурація мовного тракту в процесі виголошення звуків змінюється за часом. Ці зміни накладають на тракт потік повітря. Для утворення носових звуків до голосового тракту через піднебіння підключається носова порожнина.

Зміни конфігурації голосового тракту і коливання голосових зв'язок взаємозв'язані таким чином, що вся мовотворча система функціонує як єдиний складний об'єкт, а не як набір автономно існуючих органів, що певним чином між собою зв'язані [9]. Існує дві групи органів, що беруть участь у артикуляції. До першої відносяться – зуби, тверде піднебіння, задня стінка глотки і носової порожнини – вони беруть участь пасивно, тому що залишаються нерухомими. Друга група – нижня щелепа, губи, язик, м'яке піднебіння, голосові зв'язки – завжди залишаються активними, за рахунок виконання певних упорядкованих рухів при артикуляції.

Зміна конфігурації тракту уздовж його повздовжньої осі і за часом розраховується функцією площі поперечного перерізу. Голосовий тракт розглядається як система резонаторів, ознаки яких повільно змінюються за часом. Частоти та області резонаторів мають назву формантних частот та областей [10,11]. Найчастіше використовують термін форманта[12].

Під час проходження через мовотворчу систему у повітряному потоці можуть виникати турбулентності. Наприклад, вони можуть виникати при обтіканні повітрям перешкод різної форми, що створюються активною групою органів при артикуляції. Турбулентність породжує шумові (випадкові) складові акустичного коливання, призначення й опис яких слабо розкрито в акустичній теорії мовотворення. Шум, що викликає шиплячі звуки і може виникати у вільно поширеному потоці, має назву вихровий шум. А той, що викликає свистячі звуки і формується при обтіканні потоком перешкод, має назву краєвий тон.

Звуки, при формуванні яких голосові зв'язки виконують коливальні звуки, називають вокалізованими [13,14]. Таким чином, всі інші звуки можна віднести до невокалізованих. Більш того, невокалізовані звуки за способом утворення розрізняють фрікативні (виникають при збудженні голосового тракту турбулентним широкосмуговим шумом), вибухові (формується при створенні в тракті змички з подальшим раптовим звільненням повітря) та комбіновані (створюються при поєднанні голосних і приголосних звуків) [3,4]. Розподіл звуків за способом їх утворення на вокалізовані та невокалізовані з акустичної точки зору не мають чітко зазначеної межі, тому що коливання голосових зв'язок і утворення турбулентних шумів можуть супроводжувати один одного.

Існуючі концепції представлення мовотворчій системи [8,15], вводять припущення про функціонування і участі в створенні звука голосових зв'язок та мовного тракту, як незалежних одиниць. Припускаючи це, голосові зв'язки можна вважати самостійним джерелом, що створює послідовність імпульсів

повітряного потоку з періодичністю, що притаманна саме цим голосовим зв'язкам [16].

Згідно відомої моделі мовотворення [4], індивідуальні особливості мови можуть бути прогнозовані на основі характеристик мовного джерела та фільтра, яким виступає мовний тракт людини. Найбільшу варіативність серед цих елементів можна очікувати в функціонуванні мовного джерела. Виділяють основні параметри, що пов'язані з роботою мовного джерела та визначають його характерні особливості [5]: форма генеруючої хвилі мовного джерела; частотна модуляція періодів ОТ, амплітудна модуляція періодів ОТ, контур ОТ у вокалізованому фрагменті, середня частота ОТ, дисперсія ОТ, періодичність розподілу ОТ.

Всі ці параметри є основними ознаками МС особи, та успішно використовуються в більшості досліджень. Нажаль, ефективність їх використання без додаткових ознак вважається не достатньою та не перевищує 70-82 % [17,18].

Зважаючи на вищевикладене, при побудові системи ідентифікації МС важливим є використання більш широкого набору додаткових ознак окрім характеристик мовного джерела, оскільки точність подібної ідентифікації є не достатньою.

## **1.2. Аналіз механізмів сприйняття мовного сигналу**

Мовне коливання впливає на органи слуху особи, викликаючи певні слухові відчуття. Опишемо деякі відомості про механізм роботи периферійних органів слуху, в яких акустичне мовне коливання перетворюється у нервові збудження.

Людина при слуховому сприйнятті використовує так званий первинний акустичний перетворювач, який поділяють на три області (рис. 1.2): зовнішнє, середнє та внутрішнє вухо.

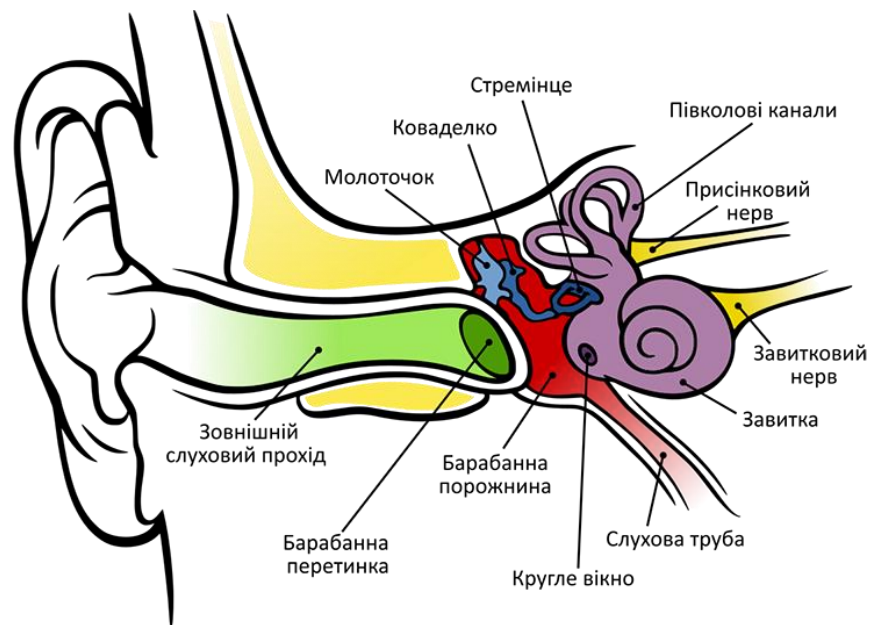


Рис. 1.2. Вухо людини

Зовнішнє вухо складається з вушної раковини та зовнішнього слухового ходу, який закінчується тонкою мембраною – барабанною перетинкою. Слуховий хід можна уявити як однорідну трубу, яка завжди відкрита з одного боку. Середнє вухо містить у собі слухові кісточки – молоточок, коваделко та стремінце, що забезпечують перетворення коливань барабанної перетинки у зміщенні об'єму рідини у внутрішньому вусі. Середнє вухо розташовано у порожнині, що заповнена повітрям, та знаходиться відразу за барабанною перетинкою. Перетворення акустичних коливань у середньому вусі описується функцією, що має частотні властивості нижчих частот із частотою зрізу від 1000 до 3000 Гц [8,10].

Внутрішнє вухо складається з завитки, вестибулярного апарату та закінчення слухового нерву. Завитка за формою має схожість з раковиною равлика, та саме в ній механічні коливання перетворюються у нервові збудження. Порожнина завитки заповнена рідиною, що не стискається. Площа поперечного перерізу завитки з боку середнього вуха (коваделка) складає приблизно 4 мм<sup>2</sup> і зменшується до 1 мм<sup>2</sup> з протилежного тонкого кінця (гелікотремі). Всередині завитка розділена поздовжньою перегородкою, таким чином порожнина з одного боку перегородки сполучується через овальне віконце з кісточками середнього вуха, які при руху визивають зсув рідини,

впливаючи на неї як на поршень. Роль поршня виконує стремінце. Порожнина з іншого боку має кругле віконце, що покрита пружною мембраною. При зсуві рідини її надлишок виходить через кругле віконце, за рахунок того, що обидві порожнини повідомляються між собою через прохід у гелікотремі завитки.

Всередині перегородки мається канал, так званий завитковий хід. Одна поздовж м'яка стінка завиткового хіда має назву мембрана Рейснера, інша стінка – це базиллярна мембрана, яка в міру наближення поступово звужується. На базиллярній мембрані розташований Кортіів орган, що містить чутливі клітини, які зв'язані з закінчення слухового нерву. Мембрана із боку середнього вуха значно товстіша, і виконує роль дисперсійної серед поширення коливань. Коли стремінце рухається, відбувається зсув рідини – це й призводить до поширення уздовж базиллярної мембрани біжучої хвилі. Параметри системи працюють таким чином, що відображення не відбувається і тому на мембрані не створюються стоячих хвиль. Високочастотні складові в міру наближення хвилі внаслідок дисперсійних властивостей мембрани поступово послаблюються. Механічні коливання тіла мембрани призводять до руху волосків (рис. 1.3) системи чутливих клітин Кортііва органу, при вигину яких виникають електричні потенціали в завитковій частині нерву.

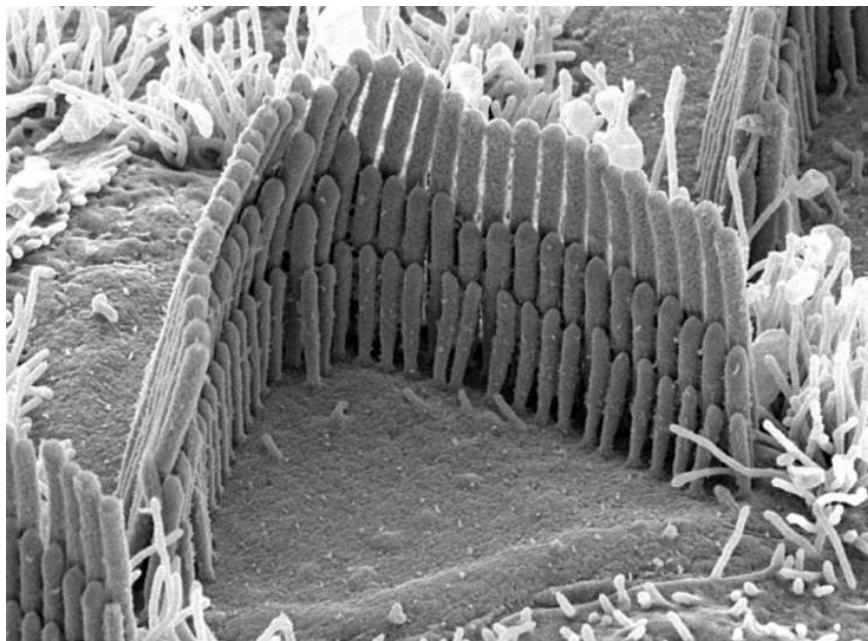


Рис. 1.3. Волоскові клітини внутрішнього вуха, що перетворюють коливання середовища у потенціал дії



Весь алгоритм, описаний вище, і є спрощеною схемою перетворення механічних коливань базилярної мембрани в електричні потенціали всередині клітин слухового нерву.

Нажаль, механізм перетворення зміщення мембрани у нервові збудження, як і спосіб кодування цих зміщень електричними потенціалами у клітинах нерву, вивчено не в повному обсязі. Проте, вважається встановленим, що локальні зміщення мембрани викликають збудження локалізованих в цієї області нейронів, які зберігають упорядкування і в слуховому нерві. Завдяки цієї обставини, можна висунути гіпотезу, об способі кодування висоти тону – за місцем максимального відхилення мембрани або по синхронному збудження певної групи нейронів. Також ця гіпотеза наводить на думки про те, що базилярну мембрану можливо уявити як частотно-вибіркову систему, в якій виконується спектральний аналіз вхідного акустичного коливання і яку можна описати частотними ознаками. Під описом частотними ознаками розуміють залежність амплітуди і фази зміщення мембрани в деякій точці від частоти гармонійного коливання стремінця. Роздільна здатність за часом більш найкраща у завитки, а за частотою - у гелікотреми.

Дослідження показали значно високу чутливість волокон слухового нерву до частоти імпульсу, що можливо було очікувати від властивостей внутрішнього вуха. Виміряні частотні ознаки волокон виявляються більш виборчими, ніж ознаки механічної системи вуха [10,11]. Виконаний аналіз результатів досліджень властивостей різних груп нейронів внутрішнього вуха дозволив зробити висновок, що у нейронні мережі слухової системи здійснюється досить тонкий аналіз, який дозволяє відобразити у нервових збудженнях детальні зміни акустичного коливання за часом і частотою [1,6,11,12]. Таким чином, проаналізувавши цей процес, можливо створити опис перетворення, що виконується слуховою системою людини та який дозволить врахувати такі особливості перетворень МС, які не були враховані раніше.

### 1.3. Огляд характеристик опису мовного сигналу

Для опису МС важливим є раціональний вибір його параметрів, що дозволяє найкращим чином представити характеристики мовного сигналу та індивідуальність особи. Представлення МС через його характеристики повинно забезпечувати: компактний опис ознак, надійну їх відмінність, мати часову стабільність, малу залежність від побічних факторів. Набір ознак оцінюється кожним дослідником окремо, а ефективність такого вибору визначається на основі критеріїв точності розпізнавання, але вважається, що найбільш точним є такий опис та набір ознак, що відповідає процесам сприйняття мови та мовотворення. Як підтверджуються багатьма дослідженнями, структурні елементи мови не можливо повністю класифікувати, оскільки неможливо підібрати такий набір ознак, щоб повністю описували елементи сигналу та виконував їх класифікацію.

Розглянемо існуючі ознаки опису мовного сигналу [19] (рис. 1.4). Їх можна класифікувати наступним чином:

- амплітудно-часові ознаки;
- кореляційні ознаки (у якості ознак моделі використовується міра зв'язку ознак та достовірність (значимість));
- спектрально-часові ознаки;
- кепстральні ознаки;
- фрактальні ознаки;
- ознаки лінійного передбачення (у якості ознак моделі використовуються коефіцієнт передбачення та порядок моделі лінійного передбачення. Також важливим параметром є похибка передбачення, як різниця між передбачуваною моделлю значенням та реальним значенням часового ряду (мовного ряду)).



Рис. 1.4. Класифікація ознак мовного сигналу

Щодо амплітудно-частотних ознак [20,21,22,23,24] (рис. 1.5), то вони дозволяють отримувати характеристики, значення яких можуть змінюватися.

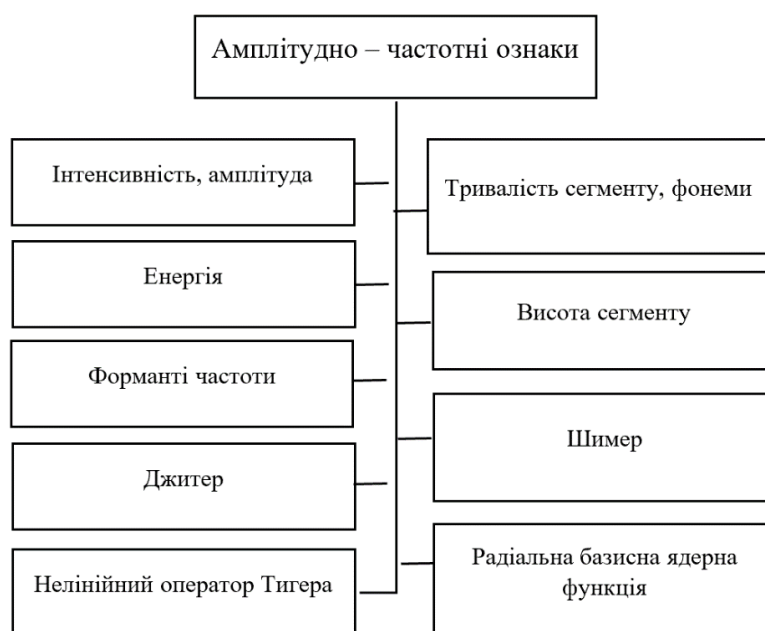


Рис. 1.5. Класифікація амплітудно – частотних ознак

Ці зміни відбуваються в залежності від параметрів дискретного перетворення Фур'є, за рахунок відображення та ширини вікна та при незначних зсувах вікна по вибірці. Якщо уявляти МС акустично, то він являє собою звукові коливання, які можна охарактеризувати за частотою (кількість

коливань в секунду), інтенсивністю (амплітуди коливань) і тривалістю. Щоб отримати більш-менш необхідну і достатню інформацію, амплітудно-частотні ознаки використовують при мінімальному часу обробки.

З поданої вище класифікації можна виділити декілька підходів: спектральні [25] і часові [26,27,28] (рис. 1.6).



Рис. 1.6. Класифікація спектрально – часових ознак

Спектральний підхід засновано на пошуку так званих спектральних піків. Часові підходи, зазвичай, використовують кореляційний підхід. Проведений аналітичний огляд показав, що за останні десятиріччя було запропоновано кілька методів, які за характеристиками перевершують традиційні. Деякі з них мають більшу точність, а інші мають більшу стійкість до шумів.

Основна задача кепстральних ознак [29,30] (рис. 1.7) – це виділення деяких параметрів МС, за допомогою дискретного конусного перетворення. У якості унікальності підходу, треба отримати якомога мінімальну кількість

параметрів, щоб пришвидшити процес порівняння з параметрами, що вже було збережено до бази даних. Але у той самий час ці параметри мають бути унікальними.



Рис. 1.7. Класифікація кепстральних ознак

Фрактальні ознаки (рис. 1.8) використовують для ідентифікації МС, більш за все, коли є певні компоненти хаотичного чи періодичного характеру. Всі ці методи базуються на статичній обробці.



Рис. 1.8. Класифікація фрактальних ознак

Проведений огляд дослідження показав, що не один з типів ознак не дозволяє достатньою мірою описати та класифікувати МС. Для оцінки сигналу необхідно створювати такі набори параметрів, які включають в себе ознаки

МС різних типів, щоб забезпечити ефективну ідентифікацію МС на усіх рівнях його декомпозиції.

Виходячи з зробленого огляду ознак МС, найбільш перспективними є методи, що дозволяють оцінити такі характеристики МС, як його самоподібність або фрактальність. В цьому аспекті, в якості ознак, можуть бути використані спектрально-часові та фрактальні методи. При цьому, передбачається, що фрактальні ознаки можуть бути отримані за допомогою спектрально-часових ознак та методів.

#### **1.4. Огляд методів визначення частоти ОТ**

На сьогоднішній день доцільно відомо, що вокалізовані фрагменти МС особи зберігають періодичність коливань голосових зв'язок. Мовний сигнал потрібно розглядати у вигляді пульсуючих коливань повітряного потоку, де період повторення імпульсів МС має назву частоти основного тону (ОТ). Досліджено, що ОТ визначає структуру МС, а також являє собою головний параметр мовного сигналу [31]. Інтонаційний контур особи є траєкторією зміни мелодії частоти ОТ. Просодика мовотворення, однією зі складових якої є інтонація, значно відрізняє акустичний сигнал мовлення від письмової мови. Тому частота ОТ носить в собі значну кількість інформації, що міститься в МС особи. З точки зору представлення цієї інформації, сам процес виділення частоти ОТ має незалежний інтерес [17]. При обчисленні частоти ОТ потрібно також враховувати як повільні зміни траєкторії частоти ОТ, так і швидкі її зміни та самі моменти включення/відключення МС. Але при побудові систем аналізу, синтезу, стиснення чи ідентифікації МС, необхідно враховувати, що частоту ОТ зазвичай використовують як одну з головних ознак, що необхідна для поліпшеного опису мовного сигналу. Коли потрібно виконувати розрахунки частоти ОТ з реального МС необхідно приділити увагу наступним моментам [17]:

- 1) можливість роботи при різних рівнях шуму;

- 2) оцінка частоти ОТ повинна виконуватися із мінімальною похибкою;
- 3) повинна дотримуватись мінімізація похибки в широкому діапазоні змін частоти ОТ, емоціональних та інших змінах мови особи, що приводять до варіації частоти ОТ;
- 4) має виконувати всі попередні перетворення сигналу;
- 5) ефективно розділяти сигнал на вокалізовані/невокалізовані фрагменти.

Щоб отримати максимальну кількість інформації, що міститься в контурі частоти ОТ, потрібно також звернути увагу на додатковий момент, коли потрібно виконувати миттєві розрахунки значення частоти ОТ.

Найбільш популярні методи виділення частоти ОТ розглянемо нижче.

Піковий метод, що вперше запропонував Доланський [32], визначає головні максимуми МС. Виконується цей метод за рахунок послідовного пошуку цих головних максимумів. Коли головний максимум знайдено, то накладають так званий інтервал блокування, у якому більше не виконується пошук максимумів. Після пошуку головного максимуму шукають наступні за величиною максимуми, повторюючи вказану процедуру. Виконані в роботі [33] дослідження різних варіантів пікових методів показали їх погану роботу в умовах шуму та складних змін форми акустичного сигналу, крім того ці методи дуже чутливі к подавленню низькочастотної компоненти МС.

Кепстральний метод, що вперше запропонував Нолл [34], розраховує частоту ОТ наступним чином. По-перше, з МС шляхом прямого перетворення Фур'є отримують широкосмуговий спектр сигналу. По друге, цей спектр логарифмується та виконується зворотне перетворення Фур'є. Все ще робиться, щоб очистити спектр сигналу від помилкових піків, які мають розташування в області допустимих значень періоду основного тону. З особливостей можна відзначити, що цей метод дуже чутливий до широкосмугових перешкод, а також вимагає складних обчислень. Щодо недоліків треба відзначити, що кепстральний метод має велику кількість помилок при різких змінах МС на переходах між вокалізованими і

невокалізованими фрагментами та не визначає моменти збудження мовного тракту [35].

Автокореляційний метод, що був запропонований Піроговим [36], обчислює частоту ОТ за рахунок автокореляційної функції, що шукає максимуми у МС. Цей метод дуже чуткий до певних моментів, а саме – коли змінюється форма обвідної часової функції МС, коли змінюється форма МС від періоду до періоду основного тону. Також, як і методи, що описано вище вимагає великої кількості обчислень та не визначає моменти збудження мовного тракту [37].

Метод Pitch, що був запропонований Рабінером - Голдом [38], полягає у наступному – спочатку МС обмежують фільтром низьких частот в діапазоні 600 – 800 Гц. Далі, вже в відфільтрованому МС, визначають відстані між усіма максимумами сигналу, знаходять відстані між усіма мінімумами сигналу та знаходять відстані між мінімумів і максимумів. І вже після всіх цих кроків визначають період за найбільшою кількістю збігів виділених відстаней. Нажаль, цей метод також не визначає моменти збудження мовного тракту та має збільшену кількість помилок за рахунок швидких змін МС та не дозволяє відстежувати миттєві значення частоти ОТ [39].

Фільтровий метод, що вперше запропонував Сапожков [4], полягає у наступному – на мовний сигнал накладаються обмеження фільтром нижніх частот таким чином, що після цієї, свого роду, фільтрації, МС містить лише основну гармоніку сигналу. Далі, у фільтрованому МС, знаходять максимуми – саме вони будуть відповідати певним моментам збудження мовного тракту. Щодо недоліків цього методу, то він не може відслідкувати швидкі зміни частоти ОТ, саме на переходах між вокалізованими та невокалізованими фрагментами, за рахунок того, що вимагає більш детального встановлення частоти зрізу фільтра низьких частот.

Метод ЛЛК (Лобанова - Левіна - Коваля) [40] обчислює частоту ОТ наступним чином. Визначається автокореляційна функція МС, знаходять найбільш ймовірне значення періоду основного тону, далі піковим методом



обчислюють значення максимумів МС. При підрахунках частоти ОТ цим методом, маємо значну кількість помилок, особливо на ділянках між вокалізованими та невокалізованими фрагментами.

Методи лінійного передбачення. Принципова різниця цієї групи методів полягає у наступному – всі вони використовують зворотній лінійний фільтр. Весь механізм зводиться до обчислення параметрів МС з метою отримання певних характеристик його спектральних властивостей. При наближенні частоти повторення в зворотному фільтрі до частоти ОТ спостерігається більше і сильніше вирівнювання спектра. Складність цього методу полягає в постійному розрахунку спектра МС та підстроювання частоти повторення в зворотному фільтрі. Загалом, робота методу має задовільні оцінки, але лише до тих пір, поки спектр вирівнюється добре. Також достовірно відомо, що при частотах ОТ вище 200 Гц методи цієї групи показують погані результати [41].

Таким чином, точність методів визначення частоти ОТ можна вважати недостатньою у зв'язку з наявністю недоліків кожного з розглянутих методів. Тому розробка методів визначення частоти ОТ, що більш точно дозволяють слідкувати за її траєкторією та використовують такі унікальні ознаки МС, що характеризують його особливості, є актуальною науковою задачею.

### **1.5. Огляд методів сегментації мовного сигналу**

Сегментацію мовного сигналу можна визначити як процес знаходження меж (з конкретною характеристикою) в розмові між словами, складами або фонемами [42,43].

Основною метою сегментації мовного сигналу є обслуговування інших проблем аналізу мовлення, таких як синтез мовного сигналу, навчання даних для розпізнавання мовного сигналу, визначення вокалізованих фрагментів для ідентифікації особи або для виготовлення та маркування мовних баз даних. Зважаючи на це, сегментація мовного сигналу є важливою підзадачею для різних галузей мовного аналізу[44,45].

Традиційний підхід до вирішення цього питання полягає в ручній сегментації мовного сигналу, яку найчастіше виконують спеціалізовані фонетики. Однак цей метод заснований на прослуховуванні фрагментів та створення судження за рахунок візуальних образів різних характеристик мовного сигналу, що робить його досить трудомістким, але високоточним в порівнянні з методами автоматичної сегментації мовного сигналу [46,47].

Системи розпізнавання мовного сигналу потребують розкладу мовного сигналу на деякі основні одиниці такі як слова, фонемі або склади. Слово є найбільш природною одиницею сегментації, воно з більшою вірогідністю несе в собі характеристики людини на відміну від фонем або складів. Саме тому, ідентифікація особи шляхом прослуховування окремих фонем або складів у теперішній час не знайшла можливостей для використання.

Фонемі - це найменша сегментарна одиниця звуку, яка використовується для формування значення. Та сама фонема в різних словах має відмінне значення. Існує надмірне узагальнення фонем. Тож суміш фонем і слова породжують наступний рівень основної одиниці мови, що називається складами [48].

На реалізацію фонемі сильно впливають сусідні фонемі. Фонемі дуже залежать від контексту. Отже, акустична мінливість основних фонетичних одиниць внаслідок контексту надзвичайно велика і погано розуміється на багатьох мовах. [49,50]

Стратегії автоматичної сегментації мовного сигналу можна згрупувати за різними перспективами, однак однією загальною класифікацією є поділ на сліпі та допоміжні методи сегментації.

1) Сліпа сегментація: Термін сліпа сегментація відноситься до методів, де не існує попередньо існуючих або зовнішніх знань щодо мовних властивостей, такі як орфографія чи повна фонетична анотація сигналу для сегментації. Вирішення цієї проблеми складається з методів, які не потребують жодної довідки знання про фонетичний зміст і базуються переважно на статистичному аналізі сигналів [51].

Через відсутність зовнішньої інформації, перша фаза сліпої сегментації спирається повністю на акустичні особливості, наявні в сигналі. Другий етап або обробка знизу вгору як правило, побудований на параметризації інтерфейсу мовного сигналу, часто використовуючи MFCC, LP-коефіцієнти або ШФП [52].

2) Самостійна сегментація: Методи допоміжної сегментації використовують якесь зовнішнє знання мовного потоку, щоб сегментувати його на відповідні сегменти. При такому типі сегментації використовується додаткове орфографічне або фонетичне представлення сигналу. До цього типу, що використовують розпізнавання із прихованими марковськими моделями (СММ) [53], [54], [55], динамічним викривленням часу (DTW) [56] або штучні нейронні мережі (ШНМ) [57].

Одним з найпоширеніших методів використання самостійної сегментації є використання СММ [58]. Методи, засновані на СММ, домінують у більшості програм розпізнавання мови з минулого століття через їх високі показники розпізнавання та відносно невелику обчислювальну складність у цій галузі розпізнавання мови [59] [60] [61].

Розглянемо існуючі методи сегментації мовного сигналу.

*Використання вейвлет-аналізу.* В. Zioko, S. Manandhar, Richard, C. Wilson і М. Zioko розробили метод сегментації мовного сигналу, виділення меж фонем на основі вейвлет-аналізу (DWT) за допомогою спектрального аналізу з достатньо високою ефективністю [62].

С. Ратсамеевай, Н. Теера-Умпон, Дж. Віласдечанон, С. Уатронджіт та К. Лікіт-Анурукс, використовуючи вейвлет-аналіз розділили мовний сигнал на низькочастотні та високочастотні компоненти, та використовуючи енергетичний контур визначили, принципи виявлення меж слів та фонем. Проведений ними експеримент на основі даних 1000 складів 10 мовців показав точність 96,0, 89,9, 92,7 та 98,9 % для початкових приголосних, голосних, кінцевих приголосних і тиші відповідно [63].

М. Ф. Толба, М. Е. Гадалла, Т. Назмі та А. А. Абдельхамід представили метод сегментації на основі вейвлет-перетворення та спектрального аналізу, орієнтованого на пошук меж між приголосними та голосними частинами. Виконаний експеримент показав точність близько 88,3 % для сегментації [64].

*Штучні нейронні мережі.* Дж. Камараускас представив метод, який використовує перцептрон і модель зворотного розповсюдження для розпізнавання окремих слів, фонем, використовуючи різні особливості мовного сигналу, що використовується в мові, або динаміку зміни звукового сигналу (лінійне кодування передбачення, коефіцієнти Честраля та коефіцієнти Фур'є перетворення). Він також визначив, що штучні нейронні мережі також можуть бути використані при встановленні меж вокалізованих фрагментів, слів та окремі голосових та приголосних фонем [65].

Ю. Сух та Ю. Лі запропонували техніку використання методу сегментації фонем з використанням багат шарового перцептрону. Структура складалася з трьох елементів препроцесора на основі багат шарового перцептрону. Під час подальшої обробки вони визначили положення меж фонем, використовуючи вихідні дані багат шарового перцептрону. Отримано для 5 мс точність 84 % і точність 87 % для 15 мс. При зменшенні порога на 0,4, вони отримали точність 92 % для 5 мс [66].

*Метод блокування чорної області.* Пані Міджанур Рахман, Ф. Хатун та Ал-Амін Бхуян запропонували метод автоматичної сегментації під назвою блокування чорної області для сегментації вокалізованих і невокалізованих фрагментів, використовуючи метод Осту [12] для динамічного порогу, Щоб перевірити працездатність системи, використовувалась база даних, що містить 500 речень з 3280 словами. Запропонована система досягла середньої точності 90,58 % [67].

*Метод зміни короткострокової енергії.* Е. А. Каур та Е. Т. Сінгх використовували короткотермінову енергію для сегментації мови на паузи, слова і склади. Запропонований метод був впроваджений та проаналізований для мовних сигналів різних мов [10].

С. Хоссейн, Н. Нахид, Н. Нузхат Хан, Д. Гомес, С. Мохаммад Мугаб на основі запропонованого вище методу отримали точність сегментації мовного сигналу 85 % [68].

*Гібридний метод сегментації мовного сигналу.* М. Каламані, С. Валарматі та С. Аніта запропонували метод гібридної сегментації для автоматичного розпізнавання мови. Метод сегментації базується на особливостях частотно-часового представлення мовного сигналу. Для виявлення меж промови використовується метод порогових значень. Цей метод досягає точності сегментації 95,33 % [69].

*Марковські моделі.* П. Бансал, А. Прадхан, А. Гоял, А. Шарма, М. Арора запропонували фонетичну сегментацію та аналіз мовлення на фонетичному рівні. Отриманий показник точності становив близько 78,14 %. Особливість реалізації полягала в поліпшенні точності при моделюванні [70].

Дж. Дайнс, С. Східхаран та М. Муді запропонували метод сегментації на основі стратегій навчання, оптимізованих для завдання сегментації. Вони проілюстрували побудову системи сприйняття мовного сигналу та виявили відмінності в оцінці парадигм сегментації мови. Оцінка системи показала високу точність сегментації мовного сигналу до 95,4 % [71].

А. Столке, Н. Райант, В. Мітра, Дж. Юань, В. Ванг та М. Ліберман представили методи для підвищення точності автоматичної фонетичної сегментації на основі акустико-фонетичних моделей СММ. Вони показали покращені результати випробувань, використовуючи більш потужні статистичні моделі для корекції меж, які зумовлені фонетичним контекстом. Вони дійшли висновку, що поєднання декількох акустичних інтерфейсів дає додатковий вигравш у точності сегментації. Крім того, використовуючи додаткову інформацію про структуру мовного сигналу, вони змогли підняти точність сегментації з 93,9 % до 96,8 % при аналізі часових інтервалів розміром 20 мс [72].

Зважаючи на виконаний огляд методів сегментації МС можна зробити висновок в наявності цілого ряду обмежень цих методів та «плаваючі»

характеристики точності сегментації, що ставить завдання на основі методів аналізу МС розробити метод його сегментації.

### 1.6. Прийняття рішення в задачах ідентифікації мовної інформації

Для задач ідентифікації мовного сигналу завжди встає окреме питання прийняття рішення [73]. Це питання повинно встановити ступінь відносин заявленої моделі і характеристик мовного сигналу, що перевіряється. Система ідентифікації по поданому мовному сигналу з параметрами мови  $Z$  особи  $A$ , має прийняти один із наступних варіантів прийнятих гіпотез:

$H_0$ :  $Z$  визначається як  $A$  (приймається як нульова гіпотеза)

$H_1$ :  $Z$  не визначається як  $A$

Висновок відносно вибору певної гіпотези базується на критерії правдоподібності на основі оцінки ймовірності отримання відмінностей між зразками:

$$\Delta(Z) = \frac{p(Z|H_0)}{p(Z|H_1)} \quad (1.1)$$

де  $p(Z|H_0)$  і  $p(Z|H_1)$  - функції щільності ймовірності (звані також правдоподібністю), пов'язані з особою  $A$  («свій») і «не особа»  $A$  («чужий»).

В дослідженні систем ідентифікації мовної інформації відомого Інституту Стандартів США (NIST) замість критерія (1.1) останнім часом використовується критерій [74].

$$\Delta(Z) = \log \left( \frac{p(Z|H_0)}{p(Z|H_1)} \right) \quad (1.2)$$

Існує безліч методів для опису моделі «не особа», але більш за все нас будуть цікавити наступні два методи. Перший метод буде ґрунтуватися на обранні для кожної особи  $A$  певного стандарту шаблонів  $\overline{A}_1, \dots, \overline{A}_N$  [75]. Завдяки тому, що ці шаблони будуть створюватися для кожної особи, можна зробити висновок, що шаблони відносяться до моделі «не особа». Другий метод базується на обранні в якості еталону випадуючих з загального

розподілення осіб, що будуть відповідати до моделі «не особа». Даний метод потребує тривалої підготовки на тестових даних, але за рахунок підлаштування під великій набір вхідних даних, він вважається найбільш ефективним та частіше за інші методи використовується при побудові систем ідентифікації мовної інформації [75,76,77].

При розробці практичних систем ідентифікації мовної інформації найчастіше визначають деяке порогове значення  $\theta$  для виконання вибору про прийняття чи відхилення особи. Коректний вибір  $\theta$  завжди є непростюю задачею, що потребує проведення цілого ряду експериментальних досліджень.

Перед використанням систем мовної ідентифікації мовного сигналу необхідно оцінити точність ідентифікації. Для подібних систем існує три основних коефіцієнти точності:

1. коефіцієнт помилки перехрещення (CER);
2. коефіцієнт помилкового відхилення (FRR);
3. коефіцієнт помилкового прийняття (FAR).

Помилкове відхилення відбувається при відбракуванні системою ідентифікації «правильного» варіанту. Помилкове відхилення прийнято називати «похибкою 1-го роду». Подібні помилкові відхилення досить важливі в реальній ідентифікації бо призводять до втрати довіри до систем та до необхідності додаткового часу на авторизацію. Помилкове прийняття відбувається при прийнятті «помилкового» варіанту як «правильного». Помилкове прийняття теж є дуже важливим, тому що при цьому варіанті неавторизована особа отримає доступ до ресурсів. Помилкове прийняття прийнято називати «похибкою 2-го роду». Вважається, що «похибки 2-го роду» важливіші ніж «похибки 1-го роду», тому що під час прийняття рішень краще не допустити справжню особу, ніж допустити помилкову особу.

Коефіцієнт помилки перехрещення (рис. 1.9) являє собою точку перетинання кривих похибок 1-го (FRR) та 2-го (FAR) роду [78].

Коефіцієнт помилкового прийняття також має назву коефіцієнта рівних помилок (ERR). Зважаючи на те, що цей коефіцієнт акумулює значення

коефіцієнтів помилкового відхилення та помилкового прийняття, то найчастіше саме він зустрічається в описі систем ідентифікації мовного сигналу [79].

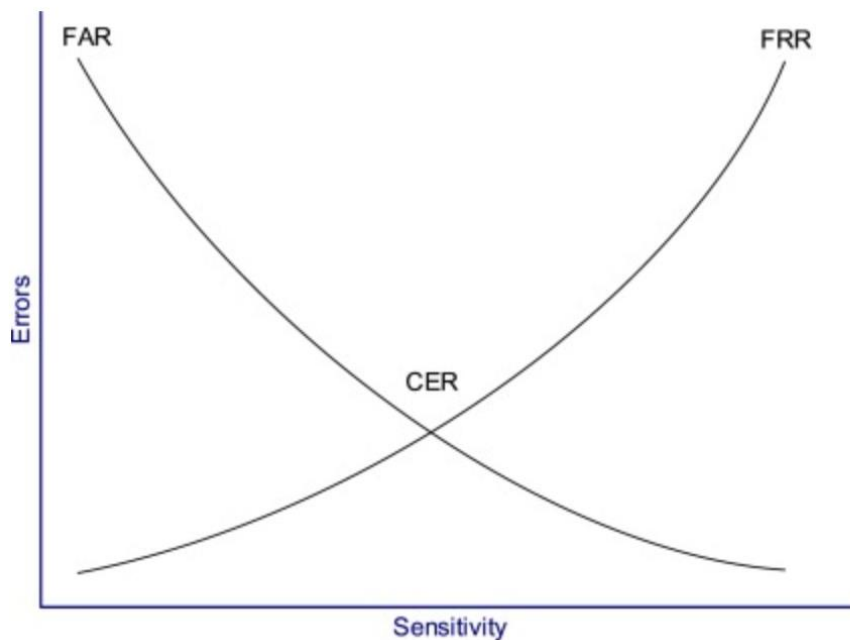


Рис. 1.9. Графік залежностей помилок 1-го та 2-го роду [58]

В змаганнях серед систем ідентифікації систем мовного сигналу також приймає участь функція витрат, яка визначає зважену суму ймовірностей помилкового прийняття і помилкового відхилення [54]:

$$C_{Det}(\theta) = C_{Miss} \times P_{Target} \times P_{Miss}(\theta) + C_{FalseAlarm} \times (1 - P_{Target}) \times P_{FalseAlarm}(\theta) \quad (1.3)$$

де параметрами функції витрат є  $C_{Miss}$  (коефіцієнт вартості пропущеного виявлення=1) і  $C_{FalseAlarm}$  (коефіцієнт вартості помилкового виявлення=1), і  $P_{Target}$  (апріорна ймовірність вказаної цільової особи=0,05),  $P_{Miss}$  (ймовірність пропущеного виявлення).

Прийняття рішення по ідентифікації МС є основним показником та результатом досліджень при порівнянні МС осіб. Тому особливу увагу треба приділити правильному підбору параметрів системи ідентифікації, щоб таким чином забезпечити необхідні значення похибок 1 та 2 роду.



## 1.7. Постановка мети і задач дослідження

Аналіз сучасного стану і напрямків розвитку досліджень задач ідентифікації мовної інформації показав, що існуючі підходи до розв'язання задач мовної ідентифікації особи не забезпечують необхідну точність ідентифікації, достатньої для ефективного практичного використання.

Основними недоліками існуючих підходів є:

1. неможливість описати мовний сигнал та визначити його унікальні характеристики, що дозволять однозначно ідентифікувати МС особи;
2. складності при прийнятті рішення у зв'язку з високою варіативністю мовних характеристик особи;
3. відсутність ефективного математичного інструменту для аналізу мовної інформації;
4. відсутність надійних акустико-фонетичних ознак для проведення ідентифікації МС особи.

Саме тому, для підвищення точності ідентифікації МС особи потребує вирішення наукового завдання удосконалення методів і алгоритмів обробки МС та підвищення точності роботи систем ідентифікації мовного сигналу особи (рис. 1.10).

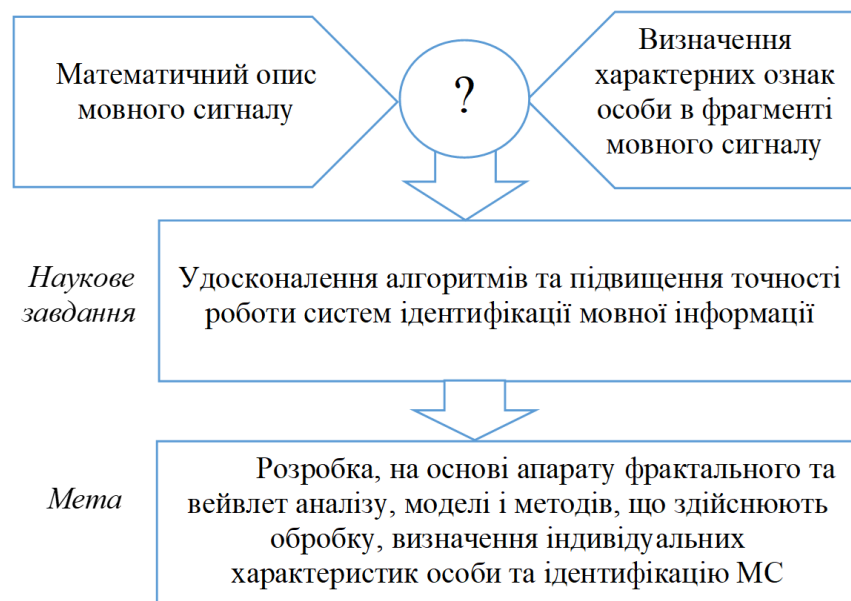


Рис. 1.10. Постановка наукового завдання дослідження

Отже, актуальною метою дослідження є розробка, на основі апарату фрактального та вейвлет аналізу, моделі і методів, що здійснюють обробку, визначення індивідуальних характеристик особи та ідентифікацію МС.

Для досягнення поставленої мети необхідно вирішити наступні наукові задачі:

1. Провести аналіз математичних моделей формування та сприйняття МС для задач їх цифрової обробки.

2. Використовуючи фрактальний аналіз МС, розробити метод його сегментації.

3. На базі вейвлет-аналізу розробити алгоритм виділення характеристик самоподібних структур в МС, що можуть використовуватись в якості унікальних ознак при ідентифікації особи.

4. На основі методу сегментації МС та алгоритму виділення характеристик самоподібних структур в МС розробити метод ідентифікації особи.

5. Провести експериментальні дослідження методів та алгоритму, для мовної ідентифікації особи, проаналізувати та порівняти результати з відомими.

6. Створити програмну систему мовної ідентифікації особи та провести порівняльний аналіз з існуючими аналогами.

Схема проведення дисертаційного дослідження представлена на рис 1.11.

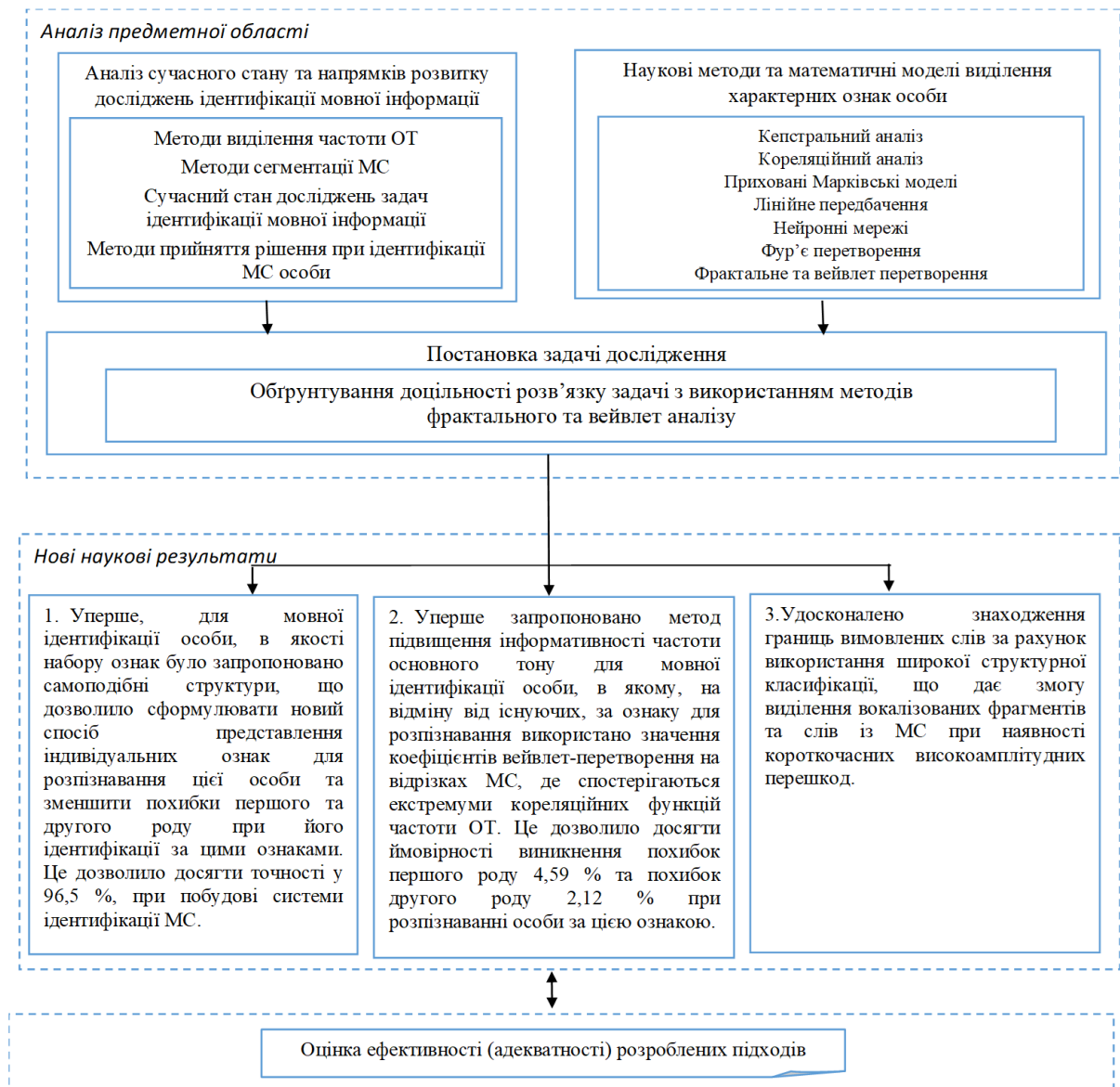


Рис. 1.11. Схема проведення дисертаційного дослідження

## Висновки за 1 розділом

У першому розділі проведено аналіз вітчизняної та зарубіжної літератури з питань, пов'язаних з темою дисертації. Створювані системи розпізнавання особи пройшли етапи від простого порівняння МС на слух до багатокomпонентних моделей, що використовують сучасні математичні методи представлення МС, ймовірнісного розкладання, передбачення та інше.

1. Проведено аналіз методів параметризації МС, методів попередньої обробки МС і його сегментації, основні підходи до ідентифікації МС особи, які застосовуються в існуючих інтелектуальних системах подібного типу.

2. Проведений огляд характеристик опису МС показав, що перспективними є методи, що дозволяють оцінити такі характеристики МС, як його самоподібність (фрактальність). В якості ознак опису цих характеристик, можуть бути використані спектрально-часові та фрактальні методи.

3. Проведений огляд методів визначення частоти ОТ показав недостатню точність цих методів, тому необхідним є розробка методу підвищення інформативності частоти ОТ, що більш точно дозволяє слідкувати за її траєкторією та використовує унікальні ознаки МС, що характеризує особу.

4. Проведений огляд методів сегментації МС показав цілий ряд обмежень цих методів та «плаваючі» характеристики точності сегментації, що потребує аналізу та розробки методу сегментації мовного сигналу.

5. Виділено, що основними причинами, які перешкоджають використанню мовних технологій у практику, є низька якість ідентифікації особи; неможливість описати мовний сигнал та визначити його унікальні характеристики, що дозволять однозначно ідентифікувати МС особи; відсутність ефективного математичного інструменту для аналізу мовної інформації; відсутність надійних акустико-фонетичних ознак для проведення ідентифікації МС особи.

6. Проведений аналіз дозволив визначити в якості основних інформаційних критеріїв ідентифікації МС особи частоту ОТ та частоти максимумів формант, як найбільш рекомендовані експертами в сфері мовних технологій.

7. Проаналізовано метод прийняття рішення в задачах ідентифікації мовної інформації та визначено коефіцієнти точності, що дозволяють оцінити ефективність системи ідентифікації.

8. На основі виконаного огляду і аналізу сформульовані задачі та наукове завдання дисертаційного дослідження, що полягає в удосконаленні алгоритмів та підвищенні точності роботи систем ідентифікації мовної інформації.

## РОЗДІЛ 2

### РОЗРОБКА МОДЕЛІ СЛУХОВОЇ СИСТЕМИ ТА МОДЕЛІ ПРЕДСТАВЛЕННЯ МОВНОГО СИГНАЛУ

#### 2.1. Модель слухової системи людини

Враховуючи притаманну слуховій системі людини властивість розпізнавати осіб, в якості моделі потенційної системи для мовного розпізнавання особи розроблено модель слухової системи пересічної людини.

Існуючі дослідження показують [80,81], що основним елементом перетворення зовнішнього МС в слуховій системі людини є завитка. Вібрації, що надходять від кісточок у середньому вусі, передаються в завитку де вона поширюються і оброблюються волосковими клітинами (сенсорами). Ці сенсори топологічно розташовані вздовж спірального ходу завитки, причому сенсори, що розташовані біля основи завитки, є більш чутливими до низькочастотних звуків, а сенсори біля верхньої частини завитки, є більш чутливими до верхньочастотних звуків. Таким чином, створювальна первинна спектрограма МС передається наступним елементом слухової кори людини. Математичне представлення цього процесу полягає в описі мовного сигналу  $s$ , як частотно-часової форми  $S$ . Частотно-часове представлення  $S$  представлено у вигляді Фур'є – перетворення функції  $s$  визначається як

$$S(\tau, \omega) = \int_R s(t)W(\tau - t)e^{2\pi it\omega} dt \quad (2.1)$$

де  $W$  – компактне вікно,  $\tau$  – час,  $\omega$  – частота.

В роботі [82] пропонується наступне представлення розкладення МС в завитці. В якості моделі вокалізованого сигналу, будемо використовувати [83] наступне представлення.

$$s(\tau) = \sum_{i=1}^n A_i \sin(\omega_i \tau) \quad (2.2)$$

Якщо мовний сигнал має тільки одну частоту, що змінюється з часом:

$$s(\tau) = A \sin(\omega(\tau)\tau) \quad (2.3)$$

При досить великому вікні перетворення Фур'є та не значному коливанні частоти приблизно отримаємо [61]

$$S(\tau, \omega) \sim \frac{A}{2i} \left( \delta_0(\omega - \omega(\tau)) - \delta_0(\omega + \omega(\tau)) \right) \quad (2.4)$$

де  $\delta_0$  – центрована відносно нуля функція Дірака.

Внесемо нову змінну

$$v = d\omega/d\tau \quad (2.5)$$

Таким чином, параметризація опису МС буде тривимірною, з параметрами  $q = (\tau, \omega, v)$ .

Отже, фактично опис МС буде представляти собою  $(\tau(t), \omega(t), v(t))$ , представив  $u(t) = dv/dt$  отримаємо наступне представлення управляючої системи

$$\frac{d}{dt} q(t) = X_0(q(t)) + u(t)X_1(q(t)) \quad (2.6)$$

де

$$X_0 = \begin{pmatrix} 1 \\ v \\ 0 \end{pmatrix} X_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (2.7)$$

Для загального МС  $s$  необхідно виконати перетворення для кожного рівня  $S$ . Використовуючи теорему неявної функції [61], це дає наступну множину просторів

$$\Sigma = \{(\tau, \omega, v) \in R^3 \mid v \partial_\omega |S|(\tau, \omega) + \partial_\tau |S|(\tau, \omega) = 0\} \quad (2.8)$$

Для нормальних умов представлення МС набір  $\Sigma$  представляє собою поверхню. Таким чином, зовнішній вихід завитки буде мати вигляд

$$I(\tau, \omega, v) = S(\tau, \omega) \delta_\Sigma(\tau, \omega, v) \quad (2.9)$$

де  $\delta_\Sigma$  – центрована відносно  $\Sigma$  функція Дірака.

Розглядаючи отриманий вираз (2.9) можна зробити висновок, щодо алгоритмів обробки МС в завитці. Таким чином, основними алгоритмами можна вважати наступні:

1.  $S(\tau, \omega)$  – потребує використання частотно-часового (спектрального) аналізу МС;

2.  $\delta_\Sigma$  – може використовувати різні варіанти аналізу розподілення, в частоті будемо використовувати фрактальний аналіз.

Згідно з вищевикладеного, структурну схему узагальненої моделі слухової системи людини, можна представити наступним чином (рис.2.1).



Рис. 2.1. Структурна схема узагальненої моделі слухової системи

Блок попередньої обробки сигналу, що відображає роботу барабанної перетинки, молоточка та коваделка, виконує функції фільтрації шуму МС, що поступає на його вхід.

Сегмент, що реалізує роботу завитки, представляється двома взаємозв'язаними блоками: блоком фрактального аналізу та блоком спектрального аналізу. Блок фрактального аналізу відображає дію завитки, що сама по собі представляє собою фрактальну самоподібну структуру, яка виконує сегментацію сигналу на вокалізовані сегменти та паузи. Блок вейвлет аналізу реалізує пошук самоподібних структур з метою ущільнення інформації в сигналі та визначення основних мовних характеристик особи.

Блок сприйняття відображає дію слухових нейронів, з'єднаних з волосковими клітинами базилярної мембрани вуха. Вона враховує такі слухові ефекти, як однонапівперіодне випрямлення і регулювання підсилення.

Сумісне використання визначених методів розкладання і перетворення МС в завитці дозволить визначити найважливіші характеристики особи для її

подальшої ідентифікації. Розгляд математичного представлення основних блоків, що виконують обробку МС, буде описано в наступних підрозділах.

## **2.2. Фрактальні структури, як елементарні складові мовного сигналу**

Будь-які системи і моделі в області розпізнавання мовних сигналів, повинні використовувати шаблони в узагальненому сенсі для розпізнавання «атомарних» складових мови. Це можуть бути накопичені у вигляді мовних фрагментів шаблони слів, речень, фонем, або просто параметри математичних моделей, що враховують передісторію процесу зміни мовного сигналу, які зіставляються з якимись аналогічними шаблонами-параметрами. Без наявності шаблонів в узагальненому сенсі при вирішенні задач ідентифікації і розпізнавання, в тому числі і при моделюванні на основі нейронних мереж, безглузда навіть постановка задач розпізнавання мови.

Будемо розглядати задачу ідентифікації мовного сигналу особи в рамках контекстно-незалежної мови з необмеженим мовним словником і довільними характеристиками мови. При побудові основної фізичної концепції будемо розглядати задачу на рівні фонемічних складових української мови або структур еквівалентних фонемічним складовим. В даному дослідженні будемо розглядати тільки голосні звуки української мови - [a], [i], [o], [y], [e]. Як буде видно з подальшого, це обмеження не є принциповим для аналізу основ індивідуальних характеристик мови особи.

Розглянемо фрагмент промови у вигляді графіка мовного сигналу (рис. 2.2).

Вид графіка залежить від індивідуальних характеристик мови особи, темпу мови і ряду інших чинників. Встановимо відповідність між голосними фонемами фрагмента мови і часовими інтервалами мовного сигналу шляхом прослуховування фрагментів.



Основою подальших результатів досліджень і розробок є критерії відповідності моделі, що розробляється, і об'єктивно існуючих характеристик мови особи. Основою вирішення цього питання є наступна методика дослідження.

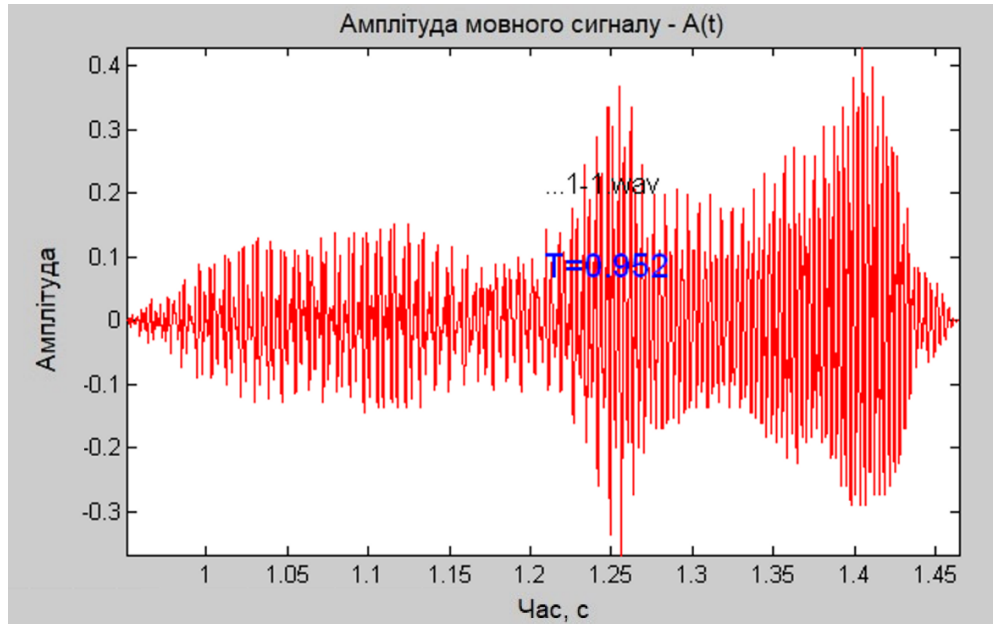


Рис. 2.2. Вокалізований фрагмент мовного сигналу

Фрагменти мовного сигналу еквівалентні при прослуховуванні гласним звукам української мови цілеспрямовано модифікуються, відповідно до розробленої фізичної моделі (в часовій і в частотній областях). Після модифікацій ці фрагменти прослуховуються групою осіб з метою виявлення ідентичності як фонемічного звучання первинного фрагмента і модифікованого, так і індивідуальних характеристик мови особи. Ідентичність звучання фонемі і характеристик мови особи до і після модифікації визначається на групі осіб випробовуваних в разі 95 % розпізнавання. Ця підхід має певні обмеження і деяку частку суб'єктивізму. Але в даний час, органи слуху людини є практично єдиним «еталонним приладом», який можна застосувати в даному дослідженні.

Дослідження проводилися як з голосними звуками, що вимовляються окремо, так і зі звуками в рамках різних текстів і характеристик мови особи.

Далі викладаються результати досліджень з наступною постановкою завдання: потрібно виявити стійкі самоподібні структури з певними параметрами, які повністю характеризують певну фонему, індивідуальні характеристики мови особи і є контекстно-незалежними.

У дослідженнях будемо розглядати фрагменти мовного сигналу, як дискретний часовий ряд амплітуд МС. Розглянемо задачу виявлення самоподібних структур в мовному фрагменті часового ряду, на основі яких можливе формування фонем і характеристик мови особи. Самоподібність будемо розглядати як самоподобу мультифрактальних структур у відповідності з концепціями Мандельброта [84,85,86,87]. Це зразкова геометрична подібність, що візуально спостерігається у фрагментах мовного сигналу. На рис. 2.3 представлений фрагмент мовного сигналу еквівалентний фонемі [i] (частина фрагмента фонемі).

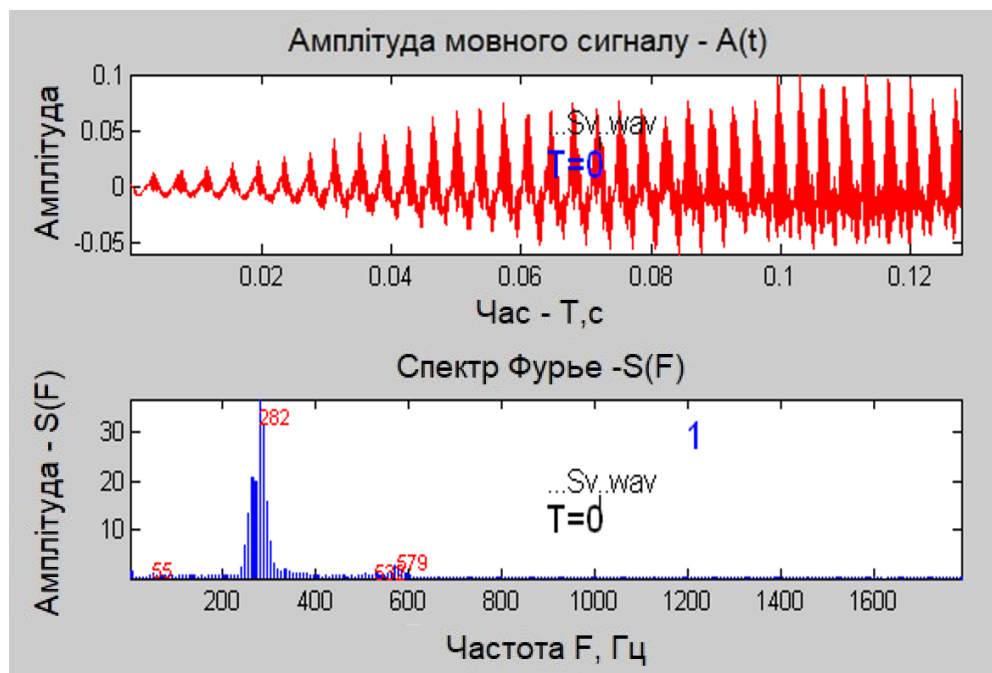


Рис. 2.3. Фрагмент мовного сигналу (фонема [i])

На рис. 2.3 представлено спектр Фур'є виділеного фрагмента фонемі. Спектри окремо виділених певних голосних звуків мають деяку геометричну подобу між собою. Ця зразкова подоба спостерігається при візуальному перегляді, незалежно від характеристик мови особи і характеру вимови фонемі. Численні роботи по встановленню відповідності між спектром МС і

конкретною гласною фонемою вказують на певну ступінь взаємозв'язку. Однак досить суворі науково обґрунтовані моделі виявлення відповідності між голосними фонемами і спектрами не існують.

Розглянемо подібність фрагментів амплітуди мовного сигналу фонемі [i] на малих інтервалах часу (близько 5-20 мс) рис. 2.4. У часовій області мовного сигналу (рис. 2.4) спостерігаються періодично повторювані самоподібні структури. Спектр Фур'є фрагмента на малих інтервалах часу дещо відрізняється від спектра рис. 2.3, як по положенню локальних екстремумів в частотній області, так і по амплітуді екстремумів.

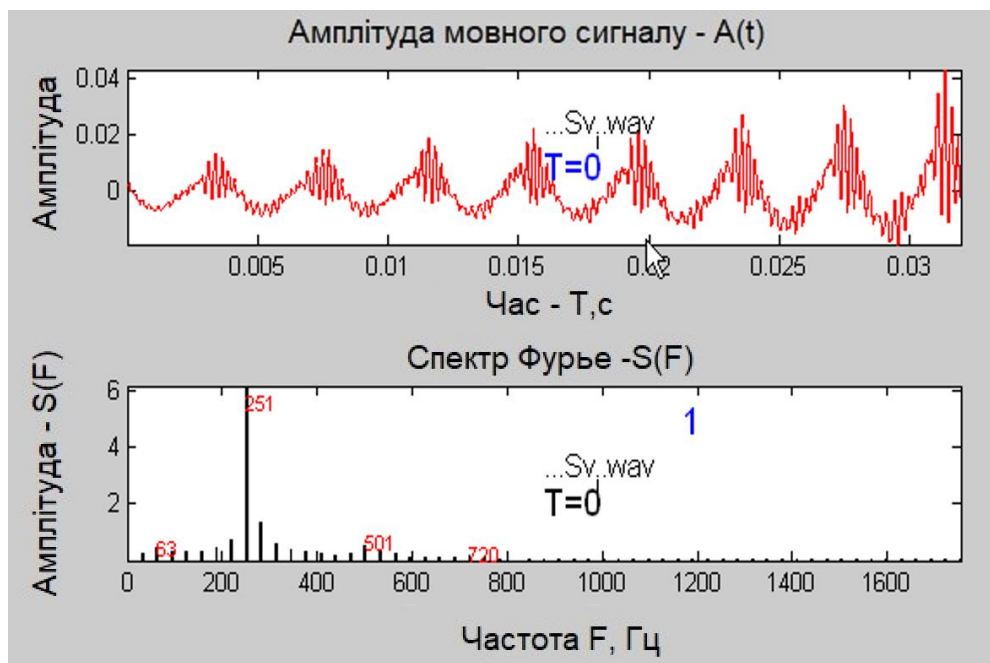


Рис. 2.4. Фрагмент мовного сигналу (фонема [i])

Виділимо фрагмент фонемі в області  $T = 0.015 - 0.02$  с (рис. 2.5). Спектр даного малого фрагмента істотно відрізняється від спектрів з великими часовими інтервалами для фонемі [i]. Також він відрізняється і по частотній роздільній здатності, зважаючи на зменшення часового інтервалу. Розглянемо задачу «конструювання» фонемі [i] з малого «атомарного» фрагмента фонемі шляхом афінних перетворень розтягування-стиснення.

Складемо з фрагмента рис. 2.5 акустичний фрагмент з декількох десятків повністю аналогічних фрагментів, що відрізняються один від одного розтягуванням-стисненням по амплітуді рис. 2.6.

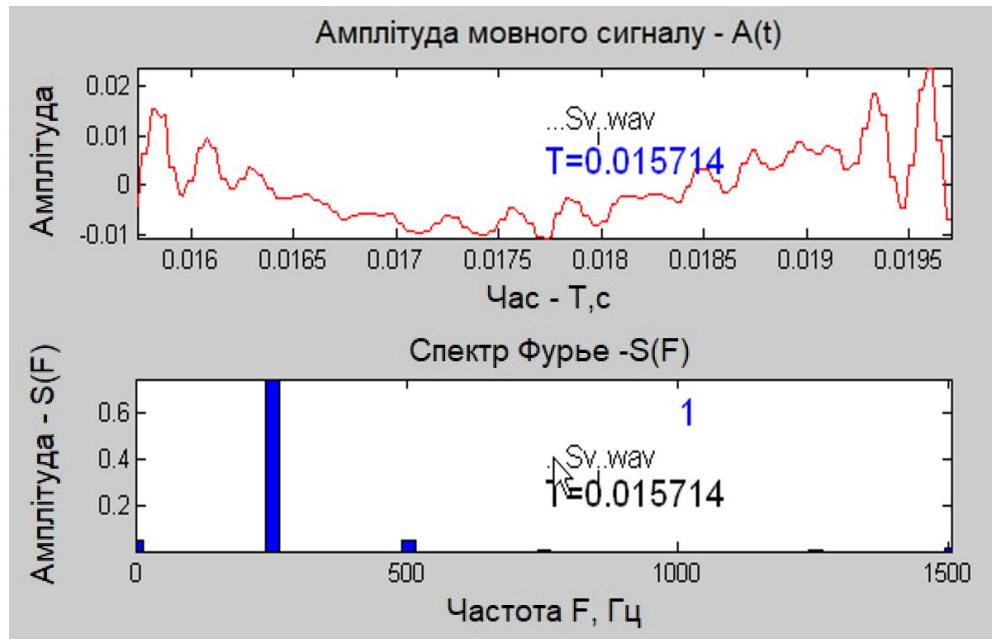


Рис. 2.5. Фрагмент мовного сигналу (фонема [i])

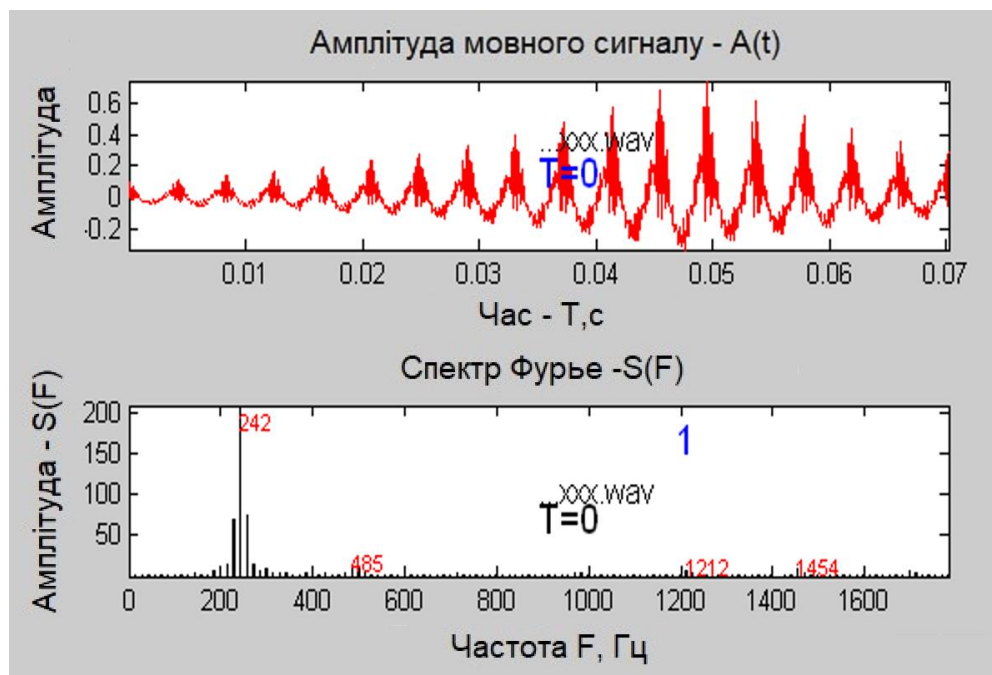


Рис. 2.6. Фрагмент модифікованої амплітуди мовного сигналу (фонема [i])

При прослуховуванні модифікованого фрагмента дослідниками чітко ідентифікується фонема [i]. В експериментах варіювалися характеристики голосу і часові розміри «атомарної» структури (від 5 до 15 мс). У широких межах також змінювалася тривалість звучання сконструйованої фонемі. При прослуховуванні після перетворень абсолютно чітко ідентифікується штучно

сконструйована фонема [i]. При цьому в більшості випадків індивідуальні характеристики мови особи або відсутні, або виражені дуже слабо.

Отримані результати побічно фігурують у безлічі досліджень. Однак у представленому дослідженні ці результати є лише вихідними передумовами для побудови моделі мультифрактальної структури звуків мовного сигналу і індивідуальних характеристик мови особи.

Таким чином, наведені вище фрагменти результатів досліджень показують, що голосна фонема [i] може бути представлена (сконструйована) на основі певних «атомарних» фрагментів цієї ж фонемі шляхом серії афінних перетворень розтягування-стиснення по амплітуді «атомарного» фрагмента.

Повністю аналогічні експерименти для фонемі «а» також показують можливість конструювання цієї фонемі з «атомарного» фрагмента фонемі розміром від 5 до 15 мс. Голосні фонемі української мови - [o], [y], [e], як показують дослідження, також мають мультифрактальну структуру. Ці фонемі можливо «конструювати» з «атомарних» фрагментів кожної фонемі.

При цьому необхідно зазначити, що в ряді випадків для «конструювання» фонемі може знадобитися не одна атомарна структура.

Як впливає з виконаного дослідження, характеристики голосних фонем більшості мов можливо вивчати і визначити за характеристиками «атомарних» складових мультифрактала. Однак, в більшості довільних варіантів конструювання фонемі, індивідуальні характеристики вихідного сигналу мови особи пропадають. Розглянемо задачу «конструювання» гласної фонемі з мультифрактальних компонентів зі збереженням індивідуальних характеристик мови особи.

Будемо досліджувати індивідуальні характеристики мови особи для голосних фонем, вивчаючи модифікації фрагментів МС як в часовій області, так і в частотній. Це пов'язано в першу чергу з відомими на сьогоднішній день нейрофізіологічними закономірностями обробки мовної інформації людиною [11-21].

Необхідно відзначити за результатами досліджень, викладеним вище, що характеристики звучання «сконструйованих» фонем залежать як від розмірів «атомарної» структури, так і від структурної геометрії фонем при візуальному перегляді в часовій області. Цей факт є очевидним. Але індивідуальні характеристики вихідної мови особи при подібному моделюванні практично повністю відсутні. Крім можливості ідентифікації чоловічого і жіночого мовних сигналів.

Розглянемо задачу більш складних афінних перетворень «атомарних» структур, які враховують ряд важливих параметрів фрагментів мови. Розглянемо фрагмент фонем [i] на менших часових масштабах - рис. 2.7.

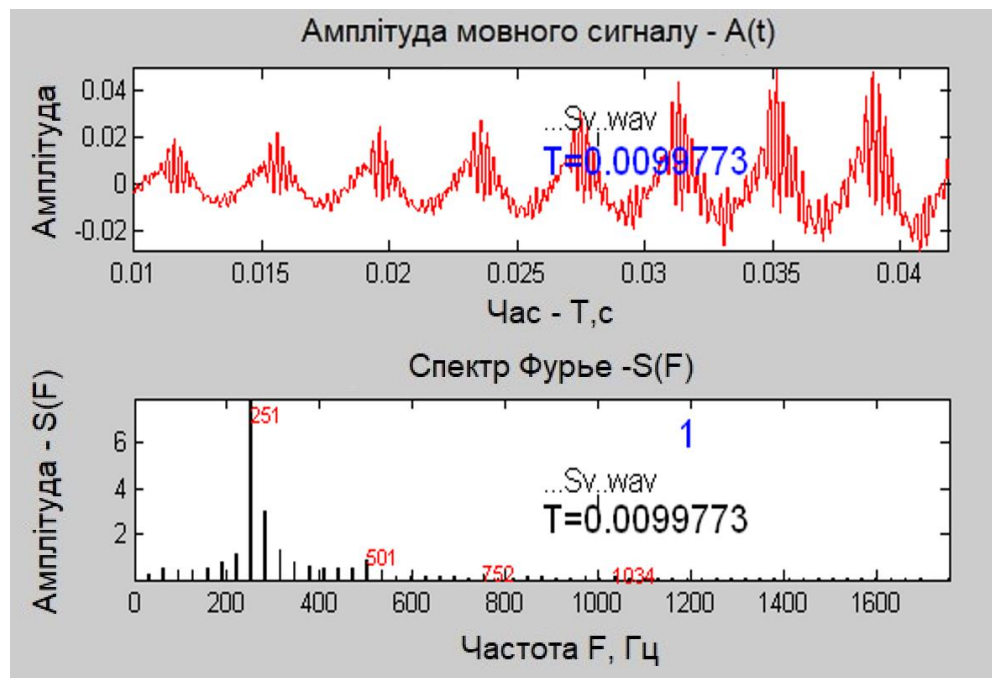


Рис. 2.7. Фрагмент мовного сигналу (фонема [i])

На основі мультифрактальної моделі можна сформулювати фонему [i] з будь-якого обраного «атомарного» фрагменту фонемі тривалістю від 5 до 15 мс. Однак при цьому зникають індивідуальні характеристики вихідної мови особи. Будемо «конструювати» атомарну структуру на часовому інтервалі еквівалентному частоті основного тону для фонемі рис. 2.7. При цьому розглянемо для даної фонемі дві базові «атомарні» структури. Одну візьмемо з часового інтервалу 0.019-0.024 мс, іншу структуру з інтервалу 0.031-0.036 мс.

Сконструємо на основі афінних перетворень розтягування-стиснення фонему [i] з двох послідовних складових частин. Перша частина фонemi - на основі першої атомарної структури, друга - на основі другої атомарної структури. Частку кожної атомарної структури в сконструйованій фонемі візьмемо пропорційно відповідним часовим інтервалам рис. 2.7. Отримана мультифрактальна конструкція представлена на рис. 2.8.

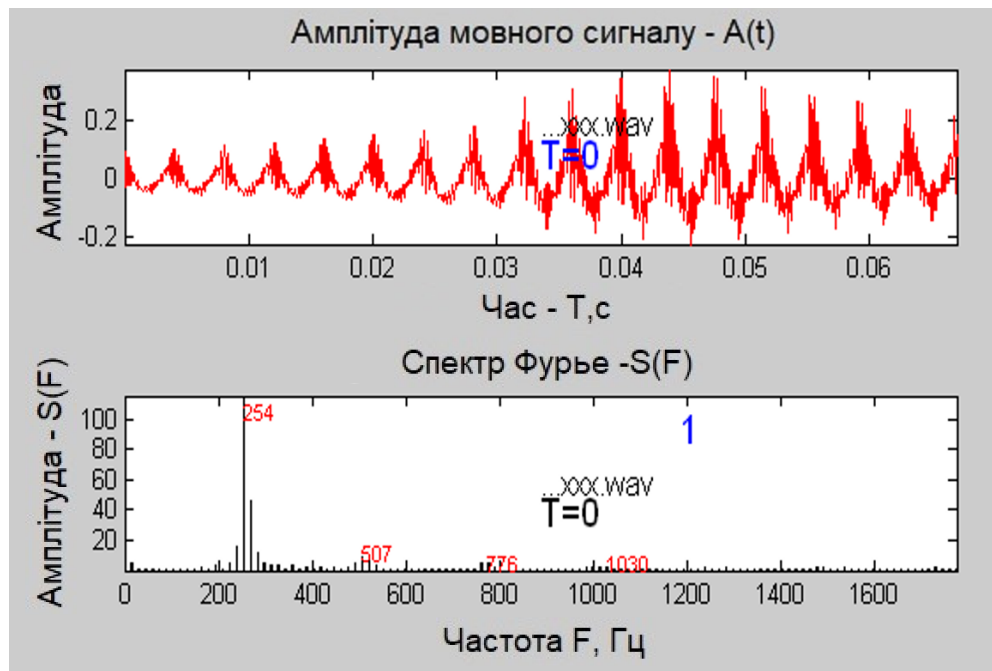


Рис. 2.8. Фрагмент мовного сигналу фонemi [i], сконструйованої з атомарних структур

При прослуховуванні «сконструйованої» фонemi дослідники ідентифікують характеристики немодифікованої мови особи. Деякі відмінності по сприйняттю характеристик мови особи після модифікації все-таки спостерігаються.

При прослуховуванні можна досягти більшої ідентичності характеристик мови особи. Для цього можливо розбиття фонemi на три або чотири базові «атомарні» структури. Однак, при цьому зростає і складність практичного використання розроблюваної моделі.

Таким чином, викладені результати досліджень показують можливість подання голосних фонем української мови у вигляді мультифрактальних структур. При цьому побудова голосних фонем на основі мультифрактального



підходу дозволяє моделювання індивідуальних характеристик мови особи, а мультифрактальний підхід для моделювання мовного сигналу може бути використаний при побудові систем ідентифікації мовного сигналу.

### 2.3. Обґрунтування вибору моделі представлення мовного сигналу

Генерація людиною мовного сигналу формує деякий ступінь турбулентності повітря під час промови. Цей процес досліджувався в багатьох роботах, які фокусуються на випадковому характері основної складової мовного сигналу, що пов'язуються з високим ступенем геометричної складності та фрагментації МС. Існує певна кількість робіт, що вважали МС таким, що не можна змоделювати, тому більшість його компонентів трактується як шум [88]. Будемо використовувати теорію фракталів [89] для оцінки геометричної складності МС та визначення його характерних особливостей.

Розглянемо обґрунтування використання теорії фракталів для оцінки ступеня мінливості МС на основі аналізу динаміки поведінки МС. Під час промови збереження імпульсу мовного потоку підкоряється рівнянню Нав'є – Стокса [90]

$$\rho \left( \frac{\partial u}{\partial t} + u \cdot \nabla u \right) = -\nabla p + \nabla \mu^2 u \quad (2.10)$$

де  $p$  – повітряний тиск,  $u$  – швидкість повітря,  $\rho$  – щільність повітря і  $\mu$  – коефіцієнт в'язкості повітря.

Іншою важливою характеристикою потоку є число Рейнольдса [91]

$$Re = \frac{\rho u D_{\Gamma}}{\mu} \quad (2.11)$$

де  $D_{\Gamma}$  – це діаметр, що характеризує мовний тракт особи.

Зважаючи на те, що коефіцієнт в'язкості повітря дуже малий, число Рейнольдса має велике значення, а зв'язку з цим, сили інерції мають бути значно більше, ніж сили в'язкості, а отже це призводить до формування вихрових потоків уздовж мовного тракту [92]. Ці вихрові потоки достатньо



вивчені та описані в [93,94], причому перехід до турбулентності під час мовлення відбувається для значно нижчих значень числа Рейнольдса [95]. Існуючі теорії турбулентності передбачають існування завихрень у багатьох масштабах, та з урахуванням теорії енергетичного каскаду [96] вважається, що вихри великого масштабу створювану енергію передають дрібним вихорам, які вже розсіюють цю енергію. Кількісну оцінку подібної багатомасштабної турбулентності можна описати за допомогою фракталів. Все це, приводить до фрактального аспекту оцінки МС, що виробляється мовною системою людини, яка створює МС на вищеописаних принципах.

Турбулентні потоки повітря мовної системи людини формують МС, створюючи вкладення в його структуру. Згідно з теоремою вкладення вектор

$$X(n) = [s(n), s(n + T_D), \dots, s(n + (D_E - 1)T_D)] \quad (2.12)$$

складений із зразків вихідного сигналу і сповільнюється кратно постійної затримці за часом  $T_D$  визначає рух в реконструйованому  $D_E$  - вимірному просторі, який поділяє спільні аспекти з вихідною фазою - простором  $Y(n)$ . Зокрема, інваріантні величини допустимої динамічної системи, такі як фрактальні розмірності з  $Y(n)$  зберігаються в реконструйованому просторі тобто простежується в  $X(n)$ .

Таким чином, вивчаючи конструктивну динамічну систему  $X(n) \rightarrow X(n + 1)$ , ми можемо розкрити корисну інформацію про складності, які пов'язані з цими інваріантними величинами про оригінал невідомої динамічної системи  $Y(n) \rightarrow Y(n + 1)$ . Вказане можливо за умови, що розкриття динаміки успішно, наприклад, розмірність вкладення  $D_E$  досить велике.

Однак, теорема вкладення не вказує методи визначення необхідних параметрів  $(T_D, D_E)$ , але тільки встановлює обмеження на їх значення. Наприклад,  $D_E$  повинно бути більше, ніж у два рази від вікна підрахунку розмірності багатовимірного набору.

Менші значення  $T_D$  дають більш скорельовані значення, ніж повинні бути у послідовних елементів. Навпаки, чим більше  $T_D$ , тим більш випадковою

буде послідовність елементів і будь-який існуючий раніше порядок зникне. Середнє взаємної інформації  $I$  для сигналу  $s(n)$ , по-перше оцінюється як

$$I(T) = \sum_{n=1}^{N-T} P(s(n), s(n+T)) \cdot \log_2 \left[ \frac{P(s(n), s(n+T))}{P(s(n)) \cdot P(s(n+T))} \right] \quad (2.13)$$

де  $P(x)$  функція щільності вірогідності оцінена з гістограми  $s(n)$ .  $I(T)$  є мірою нелінійної кореляції між парами зразків сегмента сигналу, які відділяються на  $T$  позиції один від одного. Тоді, часова затримка  $T_D$  визначається

$$T_D = \min \left\{ \arg \min_{T \geq s_0} I(T) \right\} \quad (2.14)$$

Таким чином, найважливішим елементом є оцінка фрактальної розмірності фрагментів МС, яка зберігає основні його характеристики при будь-якому перетворенні.

Припустимо, що можливість впізнавати особу за мовою є здібністю виділяти у фрагменті мови характерних ознак, притаманних особі, що упізнається. Причому, ідентифікація особи, є дією по виділенню ознак, що повторюються в сигналі (є постійно присутніми в сигналі), а отже можна зробити висновок, що в сигналі повинні бути постійні самоподібні структури, які формуються при мовленні кожної конкретної особи.

Подібність структур самим собі можлива за рахунок їхнього масштабування в мовних фрагментах. Тому найбільш вдалим інструментом виявлення самоподібних структур можна вважати вейвлет аналіз. Зважаючи на це, в створюваній моделі слухової системи людини присутній блок саме такого аналізу.

#### **2.4. Модель представлення мовного сигналу на основі вейвлет аналізу та визначення характеристик його самоподібних структур**

Представимо МС у вигляді складових з різною структурою:

$$f(t) = x_1 f_1(t) + x_2 f_2(t) + \dots + x_s f_s(t) \quad (2.15)$$

У випадку кореляції коефіцієнтів  $x_1, x_2, \dots, x_s$ , висновок про тип функцій апроксимації зробити досить важко. Основним шляхом вирішення такого варіанту сигналу є подання його у вигляді складових:

$$\sum_i g_i f_j(t_i) f_k(t_i) = \delta_{jk} \quad (2.16)$$

де  $g_i = 1/\sigma_i^2$ .

Зважаючи на те, що функції  $f_i$  в (2.15) мають різну структуру, яка схильна до змін у випадкові моменти часу, то найбільш ефективним способом їх опису є застосування методів апроксимації, заснованих на базисній декомпозиції:

$$f_i(t) = \sum_n c_{in} \varphi_{in}(t) \quad (2.17)$$

де  $f_i \in L^2(R)$ ,  $\varphi_{in}$  – базисні функції простору  $L^2(R)$ .

Для створення моделей, що адаптуються до структури сигналу, запропоновано використовувати нелінійні схеми апроксимації. У цьому випадку наближення  $f$  виконується  $M$  векторами, залежними від структури сигналу:

$$f_M = \sum_{m \in I_M} \langle f, \varphi_m \rangle \varphi_m \quad (2.18)$$

де  $I_M$  – множина індексів, яка визначається властивостями функції  $f$ .

Математична конструкція (2.15) з урахуванням введених властивостей (2.15) - (2.18) щодо опису структури сигналу названа *структурною моделлю МС*.

З огляду на пошук самоподібних структур в МС, їх різну форму та протяжність, найбільш відповідним простором для їх подання є простір вейвлет-базисів. Вейвлет-коефіцієнти  $c_{j,n} = \langle f, \psi_{j,n} \rangle$ , де  $\{\psi_{j,n}\}_{(j,n) \in \mathbb{Z}^2}$ , розглядаються як результат відображення функції  $f$  в простір з роздільною здатністю  $j$ .

Розглянемо процес виділення самоподібних структур в МС. Природньо, для отримання характеристик сигналу та виділення будь яких структур в

сигналі використовуються алгоритми вейвлет перетворення, що дозволяють розкласти сигнал шляхом операції зсуву та розтягнення вейвлета  $\psi$ . Особливістю вейвлета є те, що його середнє значення дорівнює нулю, а його інтеграл має вигляд

$$Wf(a, b) = \int f(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt \quad (2.19)$$

Та дозволяє оцінювати поведінку  $f$  в околиці точки  $b$ , що є у прямої залежності з  $a$ . Особливістю вейвлет коефіцієнтів, що отримуються з (2.19), є ефективне представлення властивості функції  $f$  при близькості масштабу  $a$  до нуля в околиці точці  $b$ . Запишемо многочлен Тейлора для деякої околиці  $v$ , при виконанні умови диференційованості функції  $f$   $m$  раз на інтервалі  $[v-l; v+l]$

$$\rho_v(t) = \sum_{k=0}^{m-1} \frac{f^{(k)}(v)}{k!} (t-v)^k \quad (2.20)$$

При такому вигляді многочлену та оцінки якості її опису  $\varepsilon_v(t) = f(t) - \rho_v(t)$  повинна виконуватись наступна умова

$$\forall t \in [v-l; v+l] |\varepsilon_v(t)| \leq \frac{|t-v|^m}{m!} \sup_{u \in [v-l; v+l]} |f^{(m)}(u)| \quad (2.21)$$

Оцінка максимальної похибки  $\varepsilon_v(t)$  при прагненні  $t$  до  $v$  визначається порядком диференційованості  $f$  в околиці  $v$ . Для уточнення максимального значення похибки скористаємося гладкістю Ліпшиця шляхом додавання наступного показника [97]:

Функція  $f$  задовольняє умові Ліпшиця  $\alpha \geq 0$  в точці  $v$ , якщо існує  $K > 0$  та многочлен  $\rho_v$  ступеня  $m = \lfloor \alpha \rfloor$  такі, що

$$\forall t \in R; |f(t) - \rho_v(t)| \leq K|t-v|^\alpha \quad (2.22)$$

Якщо функція  $f$  задовольняє (4) для всіх  $v \in [a, b]$  з константою  $K$ , незалежною від  $v$ , то вважається, що вона задовольняє умові Ліпшиця  $\alpha$  на  $[a, b]$ . Відомо, що показники Ліпшиця дають найбільш точне представлення гладкості, яке застосовується на всьому інтервалі.

В роботі [98] показано, якщо вейвлет  $\psi$  має  $n$  нульових моментів, тобто

$$\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0, k = \overline{0; n-1} \quad (2.23)$$

та  $n$  похідних, то існує  $A > 0$  для  $f \in L^2(R)$ , що задовольняє рівномірному показнику Ліпшиця  $\alpha$ ,  $\alpha \leq n$  на  $[a, b]$

$$\forall (s, u) \in R^+ \times [a, b] |Wf(s, u)| \leq As^{\alpha+1/2} \quad (2.24)$$

Можна зробити зворотній висновок, що  $f$  буде задовольняти умові Ліпшиця  $\alpha$  на  $[a + \varepsilon, b - \varepsilon]$  при будь-якому  $\varepsilon > 0$ , якщо  $f$  є обмеженою, та вейвлети коефіцієнти  $Wf(s, u)$  будуть задовольняти (2.24) для  $\alpha < n$ .

Таким чином умова (2.24) дає можливість сказати, що спадання амплітуди вейвлет – перетворення МС в залежності від масштабу пов'язано з рівномірною і точковою гладкістю Ліпшиця.

Оцінювати самоподобу  $f$  в точці  $v$  на основі показників Ліпшиця може бути достатньо складно, у зв'язку з тим, що вони можуть довільно змінюватися в околицях точки  $v$ . Скористаємося теоремою Жаффара [80] для накладення умови достатності на вейвлет - перетворення для оцінки гладкості Ліпшиця функції  $f$  в точці  $v$ .

Нехай вейвлет  $\psi$  має  $n$  нульових моментів та  $n$  похідних. Якщо  $f \in L^2(R)$  відповідає умові Ліпшиця  $\alpha \leq n$  в точці  $v$ , то існує  $A$  таке, що

$$\forall (s, u) \in R^+ \times R |Wf(s, u)| \leq As^{\alpha+1/2} \left( 1 + \left| \frac{u-v}{s} \right|^\alpha \right) \quad (2.25)$$

Зворотно, якщо  $\alpha < n$  — не ціле, та існує  $A, \alpha < \alpha$  такі, що

$$\forall (s, u) \in R^+ \times R |Wf(s, u)| \leq As^{\alpha+1/2} \left( 1 + \left| \frac{u-v}{s} \right|^\alpha \right) \quad (2.26)$$

то  $f$  задовольняє умові Ліпшиця  $\alpha$  в точці  $v$ .

Тому, при убиванні масштабу  $s$ , розраховані амплітуди вейвлет – коефіцієнтів на основі МС мають швидке убивання до нуля в областях, де сигнал є гладким та не має самоподібних структур.

Таким чином, якщо  $|Wf(s, u)|$  не має локальних максимумів на малих масштабах, то вважається, що функція  $f$ , що описує МС, є локально гладкою

та процес виділення самоподібних структур функції МС  $f$  може бути побудований за рахунок визначення максимальних значень функцій  $|Wf(s, u)|$  на малих масштабах. Враховується, що параметри масштабу обмежуються параметрами сегментації МС та його кроку.

Після виділення самоподібної структури наступною задачею є її класифікація. Використаємо підхід, запропонований в роботах [81,99]. Вираз (6) можна записати в подібному вигляді

$$\log_2 |Wf(s, u)| \leq \log_2 A + \left( \alpha + \frac{1}{2} \right) \log_2 s \quad (2.27)$$

Зважаючи на це, параметри гладкості в точці  $v$  визначаються нахилом функції  $\log_2 s$  (а відповідно  $\log_2 |Wf(s, u)|$ ) уздовж лінії максимумів. Особливістю ліній максимумів є побудова її на основі точок максимумів модуля, що є кривою  $s(u)$  в координатах  $(s, u)$ .

Проведення класифікації самоподібних структур МС будемо виконувати за допомогою (2.27) наступним чином.

Введемо позначення  $O_v(s, u)$  – як лінію максимумів вейвлет перетворення МС, що сходиться к точці  $u = v$ , при  $s \rightarrow 0$ . Для кожної точці  $v$  визначимо нахил  $\log_2 O_v(s, u)$  як функції  $\log_2 s$  при  $s \rightarrow 0$ :

$$\log_2 O_v(s, u) = \log_2 A + \left( \alpha' + \frac{1}{2} \right) \log_2 s \quad (2.28)$$

Будемо вважати, що в точці  $u = v$  маємо самоподібну структуру  $\alpha'$ .

Ефективним рішенням з задачі класифікації самоподібної структури МС залежить від особливостей базисної функції вейвлет перетворення  $\psi$ .

Наприклад, якщо вейвлет – базис  $\psi$  має  $n$  нульових моментів, то існує функція  $\theta$ , така, що

$$\psi = (-1)^n \theta^{(n)} \int_{-\infty}^{+\infty} \theta(t) dt \neq 0 \quad (2.29)$$

А вейвлет перетворення визначається у вигляді

$$Wf(s, u) = s^n \frac{d^n}{du^n} (f * \bar{\theta}_n)(u) \quad (2.30)$$

де  $\theta_s(t) = s^{-1/2}\theta\left(\frac{-t}{s}\right)$  [81]. Зважаючи на це, якщо вейвлет  $\psi$  представлений тільки одним нульовим моментом, то

$$\psi = -\theta', Wf(s, u) = s \frac{d}{du} (f * \bar{\theta}_s)(u) \quad (2.31)$$

В (2.31) максимуми  $|Wf(s, u)|$  – представляють собою згладжені функцією  $\bar{\theta}_s$  максимуми першої похідної функції МС  $f$ . Ці багатомасштабні максимуми визначають розташування точок розриву і перепаду функції МС  $f$ , а відповідно дозволяють описати розташування самоподібних структур. При наявності у вейвлета  $\psi$  двох нульових моментів представлення максимумів модуля функції МС  $f$  буде мати вигляд

$$W_2f(s, u) = s^2 \frac{d^2}{du^2} (f * \bar{\theta}_s)(u) \quad (2.32)$$

А самі вони будуть відповідати локальним особливостям самоподібних структур МС. Таким чином, як було встановлено, виконання вейвлет перетворення дозволяє отримати набір вейвлет-коефіцієнтів, що описують МС. Найбільш популярним методом представлення результатів вейвлет-перетворення є скейлограма (рис. 2.9), що дозволяє візуально представити та оцінити розташування екстремумів поверхні, що побудовані на основі вейвлет-коефіцієнтів  $W(a, b)$ .

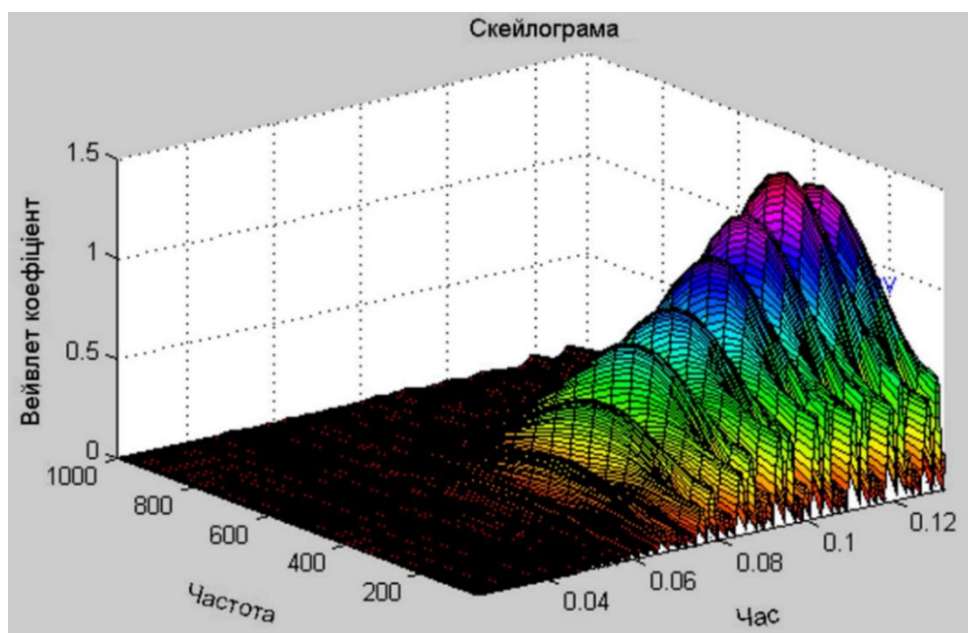


Рис. 2.9. Скейлограма фрагмента промови

Особливий інтерес представляють локальні екстремуми поверхні коефіцієнтів. Теоретично, аналіз самоподібних структур можливо виконувати на основі параметрів скейлограми, але існує цілий ряд статистичних функцій, що дозволяють виконувати оцінку характеристик спектру більш ефективно.

Загальний вид подібних статистичних функцій міри можна представити у вигляді

$$M(q, a) = \sum_{l \in L(a)} |W(a, t_l(a))|^q \quad (2.33)$$

де  $l$  – лінія локального максимуму,  $L$  – набір ліній максимумів модулів вейвлет-коефіцієнтів,  $t_l(a)$  – максимуми вейвлет-коефіцієнтів, що відносяться до ліній  $l$  масштабу  $a$ .

Згідно [100] виконується залежність

$$M(q, a) \sim a^{\tau(q)} \quad (2.34)$$

де  $\tau(q)$  визначається для значення  $q$  шляхом обчислення нахилу  $\ln(M(q, a))$  від  $\ln a$ , що називається скейлінговою експонентною. Задаючи значення  $q$  в (9) отримуємо залежність  $\tau(q)$ . Залежність  $\tau(q)$  дозволяє отримати мультифрактальний спектр МС на основі вейвлет перетворення [101], який дозволяє описати основні характеристики самоподібних структур. Для отримання мультифрактального спектру використовується наступна залежність

$$\begin{cases} D(h) = \min_q [qh - \tau(q)] \\ h = \partial\tau/\partial q \end{cases} \quad (2.35)$$

Стійкість такого метода отримання характеристик самоподібних структур полягає в використанні частотно-часового вікна, що автоматично виконує операції усереднення, а також у отриманні модулів вейвлет-коефіцієнтів при розрахунку. Необхідно відзначити, що з точки зору енергетичного підходу з питань аналізу самоподібних структур найбільш значимими є саме максимуми коефіцієнтів вейвлет-перетворення на різних рівнях декомпозиції.



Представлений підхід дозволяє розширити можливості аналізу МС за рахунок використання фрактального та вейвлет аналізу. На відміну від існуючих методів це дозволяє виконувати аналіз не стаціонарних та коротких за часом сигналів.

## **Висновки за 2 розділом**

В 2 розділі виконано основні теоретичні розробки, обґрунтовано використання фрактального та вейвлет аналізу в задачах ідентифікації МС.

1. На основі огляду роботи слухової системи людини, запропоновано її модифіковану модель, яка дозволила сформулювати принципи оброблення інформативних ознак МС, що можуть бути використані для ідентифікації МС особи.

2. Експериментальні дослідження по частотно-часовій модифікації голосних фонем української мови, показали явно виражену мультифрактальну структуру голосних фонем. Показана можливість «конструювання» голосних фонем з «атомарних» фрагментів фонем шляхом афінних перетворень. Показана можливість моделювання індивідуальних характеристик мови особи на основі мультифрактальної структури МС.

3. Обґрунтовано використання теорії фракталів для оцінки ступеню мінливості мовного сигналу на основі аналізу динаміки його поведінки. Показано, що подібність структур в мовному сигналі можливе за рахунок їх масштабування в МС. Зважаючи на це, обґрунтовано, що основними підходами в задачі ідентифікації МС можуть бути фрактальний та вейвлет аналіз.

4. Описана структурна модель МС, розглянута та узагальнена послідовність математичних перетворень інформаційного каналу мовного спілкування, на основі якої запропоновано підхід до виділення особливостей самоподібних структур.

## РОЗДІЛ 3

### РОЗРОБКА МЕТОДУ ТА АЛГОРИТМІВ, НЕОБХІДНИХ ДЛЯ ПОБУДОВИ СИСТЕМИ ІДЕНТИФІКАЦІЇ МОВНОЇ ІНФОРМАЦІЇ

#### 3.1. Вибір методу розрахунку фрактальної розмірності

Згідно існуючих досліджень, процес формування МС є процесом постійного колювання окремих частин мовного тракту. Зважаючи на те, що в процесі промови ці частини практично не змінюються, в МС формуються самоподібні або мультифрактальні структури [102]. Таким чином, при початку промови, будемо спостерігати різкі зміни, що описують фрактальність МС. Для розробки методу сегментації МС необхідно виконати оцінку варіативності цих характеристик при переході з мови на паузи та прийняти рішення щодо можливості їх використання. В якості фрактальної характеристики, що дозволяє описати характеристики МС такого типу, будемо використовувати фрактальну розмірність.

Існує багато методів розрахунку фрактальної розмірності. Для використання в задачах ідентифікації мовної інформації необхідно оцінити ці методи та обрати раціональний підхід для її обчислення. Розглянемо більш детально кожен з них.

*Метод мірила (Yard Stick Method)*. Спочатку обирається фіксований розмір вікна  $r$ , який розділяє мовний сигнал на фрагменти, що покривають його профіль. Таким чином, сигнал буде представлений у вигляді  $N(r)$ . При обчисленні фрактальної розмірності за цим методом для кожного фрагменту  $n$  розраховується довжина кривої  $L_i$ , і кожен  $r_i$  ставиться у відповідність з отриманою довжиною. Отриманий ряд  $(r_i, L_i)$  будується на логарифмічній шкалі за обома координатами, та виконується підлаштування лінії до графіку за методом найменших квадратів. На основі отриманого графіку обчислюється фрактальна розмірність, як  $D = 1 - \alpha$ , де  $\alpha$  - коефіцієнт регресії.

*Метод Хаусдорфа.* В основі метода лежить покриття профілю сигналу квадратними сітками. Коли ширина сітки  $r$  змінюється, кількість елементів сітки  $N(r)$  буде змінюватися теж. Відносини між  $N(r)$  і  $r$  представляються у вигляді:  $N(r) = kr^{-D}$ . Зважаючи на те, що ширина квадратної сітки буде  $r_1, r_2, r_3, \dots, r_k$ , кількість елементів буде відповідно  $N(r_1), N(r_2), N(r_3), \dots, N(r_k)$ . Обчислення фрактальної розмірності за цим методом виконується наступним чином. Будується графік отриманого ряду  $(r_i, N(r_i))$  на логарифмічній шкалі за обома координатами та використовується лінійно-регресійний метод для аналізу даних. В результаті чого отримуємо коефіцієнт регресії  $\alpha$ . Фрактальна розмірність обчислюється як  $D = -\alpha$ .

*Варіаційний метод.* Для розрахунку за цим методом профіль покривається прямокутниками з шириною  $r$ . Для кожного прямокутника обирається еталонна точка та розраховується відхилення  $H_i$  між найвищою та найнижчою позицією. Якщо ширина  $r$  дуже мала,  $H_i$  приблизно дорівнює довжині кривої  $V(r)$ . Таким чином, міра дорівнює  $V(r) = \sum \frac{rH_i}{r^2} = \sum H_i/r$ . Будується графік отриманого ряду  $(r_i, V(r_i))$  на логарифмічній шкалі за обома координатами та виконується лінійний підбір. Фрактальна розмірність обчислюється як  $D = 2 - \alpha$ , де  $\alpha$  - схил отриманої функції.

*Метод структурної функції.* Метод структурної функції також називаються підходом збільшення. Профіль розглядається як послідовність функції висоти  $z(x)$ . Для будь-яких двох точок з відстані  $r$  у послідовності  $z(x)$  визначається функція структури  $S(r)$ , що є середнім арифметичним значенням квадрата різниці висот. Відносини між послідовністю  $z(x)$  і функцією структури  $S(r)$ :  $S(r) = E[z(x+r) - z(x)]^2 = cr^{4-2D}$ , де  $r$  називається інтервальним масштабом. При обчисленні фрактальної розмірності вибираються різні масштаби  $r$ , та отримуються значення функції  $S(r)$ . Будується графік отриманого ряду  $(r_i, S(r_i))$  на логарифмічній шкалі за обома координатами. Фрактальна розмірність обчислюється як  $D = 2 - \alpha/2$ , де  $\alpha$  - схил отриманої функції.

*Середньоквадратичний метод.* Основний принцип подібний до методу структурної функції. Дослідження показало, що масштабному співвідношенню функції  $z(x)$  з фрактальними характеристиками відповідає наступна формула

$$z(x) - z(x_0) = \zeta |x - x_0|^{2-D} \quad (3.1)$$

Нехай  $x_0 = 0$  к  $z(0) = 0$ , Тоді ми можемо обчислити дисперсію або кореляційний момент послідовності функції  $z(x)$

$$S(r) = D(r)^{1/2} = cr^{2-D} \quad (3.2)$$

де  $r$  є інтервальним масштабом, і  $r = x - x_0 = x$ . Рівняння (3.2) показує, що відносини між кореляційним моментом  $S(r)$  і масштабом інтервалу  $r$  є експонентою потужності, а потужність є функцією фрактальної розмірності. Обчислення фрактальної розмірності виконується наступним шляхом для кожного інтервального масштабу  $r_i (i = 1, 2, \dots, n)$ . Розраховується дисперсія  $S(r)$ . Будується графік отриманого ряду  $(r_i, S(r_i))$  на логарифмічній шкалі за обома координатами. Фрактальна розмірність обчислюється як  $D = 2 - \alpha$ , де  $\alpha$  це нахил лінії.

*Метод Херста.* Для функції висоти  $z(x)$ , при заданому масштабі  $r$ , обчислюється середнє значення у вигляді:  $\bar{z}_r = \frac{1}{r} \sum_{x=0}^r z(x)$ , накопичувальне відхилення:  $z(x, r) = \sum_{x=0}^r [z(x) - \bar{z}_r]$ , максимальна різниця  $R(x) = \max_{0 \leq x \leq r} z(x, r) - \min_{0 \leq x \leq r} z(x, r)$ , і стандартне відхилення

$$S(r) = \sqrt{\frac{1}{r} \sum_{x=0}^r [z(x) - \bar{z}_r]^2}. \text{ Дослідження Херста виявило, що статистичний}$$

закон  $R/S$  дорівнює

$$R/S = cr^H \quad (3.3)$$

де,  $c$  це постійна,  $H$  це індекс Херста. Беручи логарифм з обох сторін до рівняння (3.3), отримаємо

$$\ln(R/S) = \ln(c) + H * \ln(r) \quad (3.4)$$

Після визначення максимальної різниці  $R(r)$  та стандартного відхилення  $S(r)$  з різним масштабом  $r$ , на підставі рівняння (3.4), індекс Херста  $H$  можна

розрахувати за допомогою методу найменших квадратів. Обчислення фрактальної розмірності має вигляд:  $D = 2 - H$ .

*Метод Хігучі.* Оригінальне формулювання методу обчислення пояснюється Хігучі в [4]. Фрактальна розмірність Хігучі (HFD) з  $X$  обчислюється таким чином: Для кожного  $r \in \{1, \dots, r_{max}\}$  та  $m \in \{1, \dots, r\}$  визначається довжина  $L_m(r)$  на

$$L_m(r) = \frac{N-1}{\left[\frac{N-m}{k}\right] r^2} \sum_{i=1}^{\left[\frac{N-m}{r}\right]} |X_N(m+ir) - X_N(m+(i-1)r)| \quad (3.5)$$

Довжина  $L(r)$  визначається середнім значенням  $r$  довжин  $L_m(r)$ .

$$L(r) = \frac{1}{r} \sum_{m=1}^r L_m(r).$$

Нахил лінійної функції, яка найкраще підходить, через точки даних  $\left\{ \left( \log \frac{1}{r}, \log L(r) \right) \right\}$  визначено як фрактальну розмірність Хігучі.

Визначимо, який з методів розрахунку фрактальної розмірності є найкращим для оцінки фрактальної розмірності мовного сигналу. Для попереднього налаштування методів було використано 2 тестових мовних сигнала, представлених на рис.3.1. Причому рис.3.1а) представляє собою вокалізований фрагмент, а рис.3.1б) – невокалізований фрагмент або паузу.

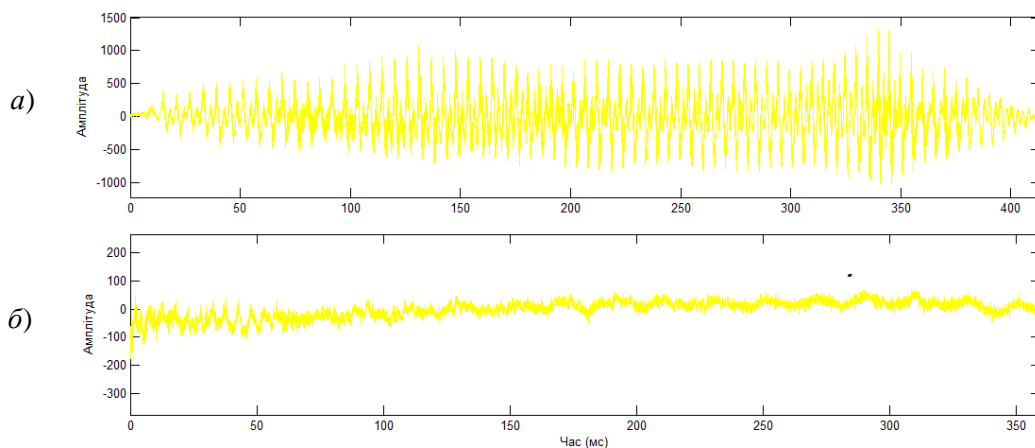


Рис. 3.1. Тестові мовні сигнали

а) вокалізований фрагмент, б) невокалізований фрагмент

Для розрахунку кожного з варіантів підбиралось раціональне значення показника  $r$  – воно обиралось таким чином, щоб забезпечити рівень значущості коефіцієнта кореляції 0.03. Це досягалось за рахунок змінення показника  $r$  з кроком 0.1 та обчислення значення фрактальної розмірності відповідного методу  $D(r)$ , з урахування покриття кривої мовного сигналу згідно встановленого показника коефіцієнта кореляції.

Результати обчислення фрактальної розмірності кожним з методів, що представлені вище, наведено в таблиці 3.1.

Таблиця 3.1

Порівняння методів розрахунку фрактальної розмірності для мовного сигналу (представленого на рис.3.1)

Метод обчислення фрактальної розмірності	Вокалізований фрагмент		Невокалізований фрагмент	
	Фрактальна розмірність	Кореляційний коефіцієнт	Фрактальна розмірність	Кореляційний коефіцієнт
Хігучі	1.75	0.995	1.221	0.981
Метод Херста	1.72	0.968	1.181	0.979
Середньоквадратичний метод	1.657	0.979	1.119	0.982
Метод структурної функції	1.701	0.997	1.199	0.998
Метод «мірила» (Yard Stick Method)	1.422	0.993	1.011	0.992
Варіаційний метод	1.353	0.996	1.154	0.999
Метод Хаусдорфа	1.68	0.998	1.197	0.995

Попередні дослідження для тестових фрагментів мови та паузи показали погані показники кореляції при обчисленні фрактальної розмірності методами Хігучі, Херста та середньоквадратичним методом, найбільш високі показники кореляції забезпечуються при використанні варіаційного метода та метода Хаусдорфа.

Наступним етапом дослідження було проведення аналізу роботи методів для більшого набору тестових даних з налаштуваннями згідно тестових фрагментів. В дослідженні використовувалось по 300 підготовлених

вокалізованих і невокалізованих фрагментів. Результати дослідження представлені на рис. 3.2.

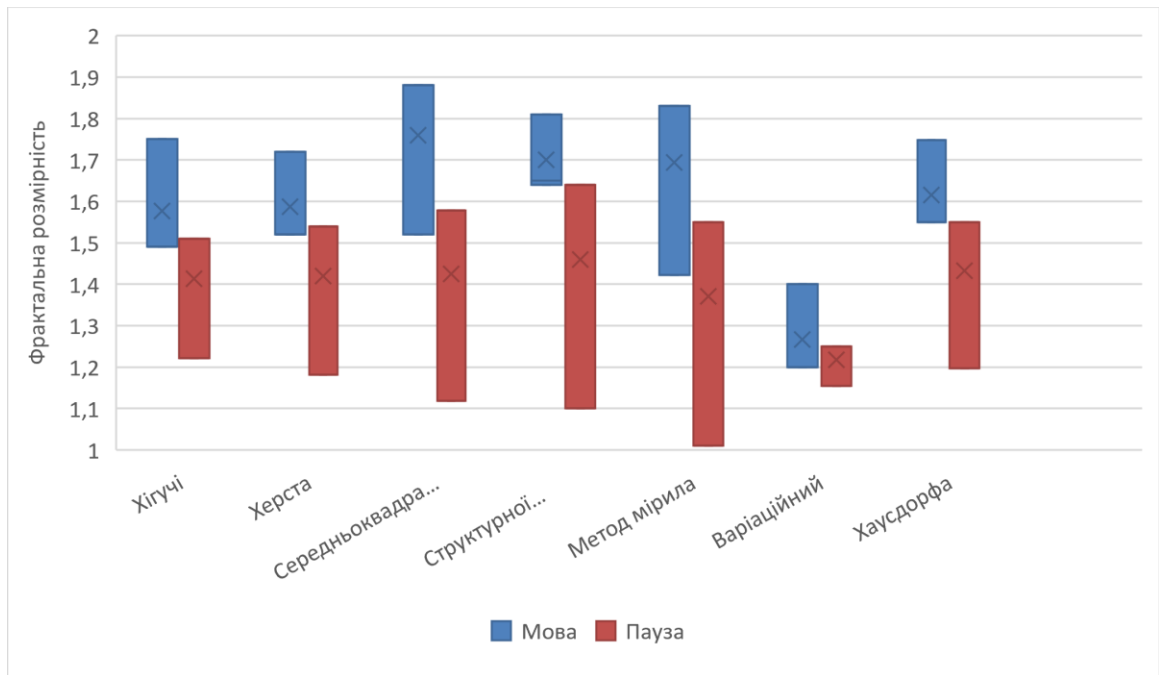


Рис. 3.2. Порівняння методів розрахунку фрактальної розмірності для фрагментів мовного сигналу

На основі аналізу результатів можна зробити висновок, що найбільш прийнятними для використання в сегментації мовного сигналу є методи Хаусдорфа та структурної функції, бо вони забезпечують менше значення перекриття діапазонів фрактальної розмірності для вокалізованих і невокалізованих фрагментів, варіаційних метод не може розглядатися в задачах сегментацію у зв'язку з малим діапазоном варіації фрактальної розмірності в заданих умовах, методи «Yard Stick» та середньоквадратичний мають широкий діапазон перекриття значень фрактальної розмірності, що призведе для зниження точності сегментації при використанні цих методів. Методи Хігучі та R/S аналізу задовільно можуть використовуватись для сегментації мовного сигналу, але їх якість буде значно нижче ніж у методів Хаусдорфа та структурної функції.

Підводячи підсумок, методи Хаусдорфа та структурної функції є більш придатними для обчислення фрактальної розмірності в рамках задачі сегментації мовного сигналу. Зважаючи на більшу обчислювальну складність

методу структурної функції для сегментації МС обираємо метод розрахунку фрактальної розмірності за Хаусдорфом.

### 3.2. Розробка методу сегментації на основі фрактальних характеристик сигналу

Для дослідження мовних сигналів і проведення апроксимації графіків часових рядів амплітуд звукової хвилі певними сумами (розбивками типу клітина), звернемося до розмірності Хаусдорфа. Фрактальна розмірність  $Dx$  по Хаусдорфу визначається наступним чином

$$S(p) \sim p^{2-Dx} \text{ при } p \rightarrow 0 \quad (3.6)$$

де  $S(p)$  - повна площа комплексу, з масштабом розбиття  $p$ .

З практичного боку, при спробі обчислення  $Dx$  на підставі (3.6), виникає ряд проблем. Це обумовлено тим, що реальні часові ряди завжди матимуть мінімальний масштаб  $p_0$  і одночасно з цим перехід к асимптотичному поданню в (3.6), відбувається досить повільно.

На відміну від звичайних часових рядів, які розглядаються в більшості завдань обробки сигналів, мовні сигнали мають значні відмінності. Одне з головних відмінностей полягає в тому, що амплітуда звукової хвилі добре описується в вигляді суми гармонійних коливань (як з фізичної, так і з математичної точки зору). Цей підхід дозволяє істотно знижувати кількість відліків, які необхідні для проведення справжнього аналізу фрактальної розмірності.

Наступним, важливим моментом є можливість оцінок в ряді завдань мовних технологій порядку мінімального фрактального масштабу. На основі раніше проведених досліджень, відомо, що вся важлива мовна інформація міститься в певному діапазоні частот, а саме до 4500 Гц завдяки чому існує можливість виконувати якісну оцінку мінімально раціональних фрактальних масштабів мовних фрагментів.



Як приклад, розглянемо мовний сигнал з частотою дискретизації  $F_s$  і розрядністю  $r$  біт. Це означає, що  $F_s$  знаходиться в діапазоні від 8000 Гц до 44100 Гц, а  $r$  - від 8 до 24 біт. Отже, щоб планомірно покрити графік досліджуваної звукової хвилі, необхідно використовувати деякі мінімальні значення, схожі з розміром прямокутника  $a * b$ . Де  $a$  буде визначатися як мінімально здійсненне зміна амплітуди звукової хвилі для конкретної розрядності звукового файлу. Як приклад, для 8-розрядного звукового файлу  $a = 2/256$ . Мається на увазі, що всі звукові файли перетворені в wav. формат, а значить значення амплітуди розглядається в діапазоні від -1 до +1 і представлена у вигляді числа з плаваючою комою. Для загального випадку:

$$a = 2/2^r \quad (3.7)$$

Мінімальне значення сторони прямокутника для часової осі  $b_0$  буде рівним  $1/F_s$ . І так як розмірність не суттєва для даного випадку, то приймемо мінімальний розмір  $b_0 = 1$ . Використовуємо такий метод розрахунку фрактальної розмірності по Хаусдорфу: кожне часове вікно МС представимо у вигляді набору прямокутників розміру  $a \times b$ , що покривають графічне представлення МС. Нехай масштаб представлення  $p$ :

$$p = k \cdot b \quad (3.8)$$

де  $k = 1, 2, 3, \dots$  – це коефіцієнт масштабу представлення.

Відомо, що обчислення фрактальної розмірності по Хаусдорфу, виконується, як

$$D = 2 - \lim[\ln(N(p))/\ln(p)] \quad (3.9)$$

де  $\ln(N(p))$  – натуральний логарифм залежного від масштабу представлення числа прямокутників  $N(p)$ , які включають хоча б одне значення амплітуди мовного сигналу,  $\ln(p)$  – натуральний логарифм масштабу представлення.

Визначаємо фрактальну розмірність  $D$  на підставі [3,4]

$$D = 2 - Dx \quad (3.10)$$

Залежно від частоти дискретизації цей часовий інтервал відповідає числу відліків  $N = F_s * 2/100$ .

Введемо мінімальний фрактальний масштаб  $k \geq 3$ . Після численних досліджень різних фрагментів на різних аудіо файлах була оцінена величина фрактальної розмірності, яка показала велику варіабельність величини  $Dx$  при мінімальних значеннях  $k \geq 2$ . Поряд з цим, при значенні оцінки фрактальної розмірності, будуть слабо змінюватися в межах деяких часових інтервалів по 20мс.

Для певної реалізації часового вікна паузи аудіо файлу будемо змінювати масштаб. Для цього побудуємо графік залежності  $\ln(N(p)) = f(\ln(p))$ . Наступним кроком, після побудови графіка, зробимо апроксимацію перших точок графіка за допомогою лінійної залежності

$$f = c * \ln(p) + c_0 \quad (3.11)$$

де  $c, c_0$  - це коефіцієнти апроксимації.

Попередні проведені дослідження показали необхідність апроксимації перших декількох точок МС, тому було прийнято розмір апроксимації в 10 точок, яке буде корегуватися при більш детальному дослідженні.

В рамках даного методу величина дорівнює  $Dx$

$$Dx = abs(c) \quad (3.12)$$

Також необхідно відзначити, що при обчисленні фрактальної розмірності на основі методу, описаного вище, з використанням мінімального фрактального масштабу ( $k = 1$ ), значення фрактальної розмірності істотно відрізняються. Отже, в пропонованій методиці оцінки фрактальної розмірності на основі Хаусдорфа, що розраховується її величина не відповідає визначенню розмірності Хаусдорфа. Але, після численних досліджень звукових файлів на основі модифікованого підходу, була доведена ефективність використання параметра (3.11) в задачах мовних технологій.

Проведений розрахунок фрактальної розмірності дозволив побудувати рис. 3.3, на якому представлена величина  $D$  - фрактальної розмірності для фрагмента МС для всього часового вікна. Огляд графіка показує достатньо високий рівень слідкування фрактальної розмірності за зміною амплітудних

значень мовного сигналу, що може використовуватися для сегментації мовного сигналу.

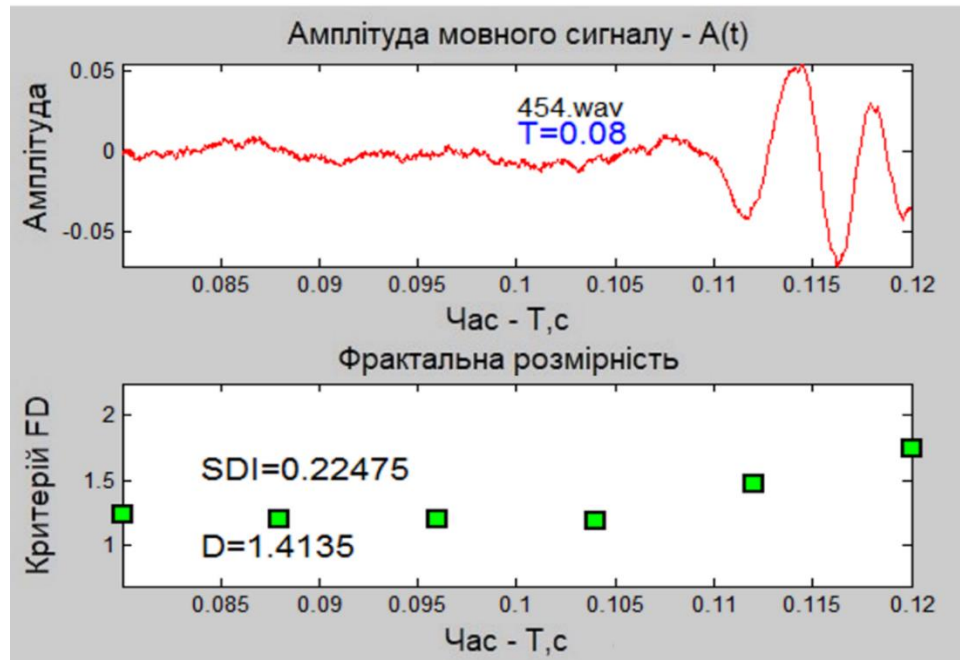


Рис. 3.3. Фрактальна розмірність при переході з паузи на мову

На основі виконаного дослідження пропонується наступний метод сегментації мовного сигналу з використанням фрактальної розмірності:

1. розбиття на часові фрагменти;
2. попередня апроксимація мовного сигналу (залежність 3.11);
3. визначення фрактальної розмірності з заданим часовим вікном в кожному часовому фрагменті (залежність 3.10);
4. розподіл часових фрагментів на вокалізовані та невокалізовані фрагменти згідно встановленого порогу величини фрактальної розмірності.

### 3.3. Розробка алгоритму виділення характеристик самоподібних структур в якості ідентифікаційних ознак мовного сигналу

Будемо розглядати фрагменти мови в сигналі, як дискретні часові ряди амплітуди звукової хвилі. Поставимо задачу визначення характеристик самоподібних структур в часовому ряді. Це можуть бути різні геометрично подібні структури, візуально спостерігаються при розгляді графіків зміни

амплітуди звукової хвилі. Введемо поняття самоподоби, як наявності близької геометричної форми об'єктів МС, що можуть відрізнятися часовими або амплітудними характеристиками при збереженні форми [103]. Одним з варіантів представлення цих структур є використання вейвлет перетворення МС. В роботі обґрунтовано використання комплексного вейвлету Морле для опису МС:

$$C_{mor}(t_i, T_k, F_b, F_c) = (\pi F_b)^{0,5} \exp(2j\pi F_c t_i) \exp(-(t_i - T_k)^2 / F_b) \quad (3.13)$$

де  $F_b$  – параметр ширини вейвлета,  $F_c$  – центральна частота вейвлета,  $t_i$  – дискретні часові відліки,  $T_k$  – часовий відлік відповідний центральній частині часового вікна,  $j$  - комплексна одиниця.

Нехай  $A(t_i)$  – величина амплітуди звукової хвилі фрагмента промови аудіо файлу в момент часу  $t_i$ . Розглянемо часове вікно фрагмента промови з інтервалом  $\delta T$ . Цей інтервал, як зазначено вище, в дослідженнях приймався в діапазоні 10-30 мс. Параметр ширини комплексного вейвлета Морле  $F_b$  виберемо постійним для всіх перетворень. Його величина вибиралася з умови практичного загасання абсолютних значень вейвлета Морле при значеннях  $t_i - T_k$  рівних  $\delta T / 2$ . Будемо для кожного фрагмента промови обчислювати згортку вейвлета Морле з фрагментом часового ряду амплітуди звукової хвилі у вигляді

$$C(T_k, F_b, F_c) = (1/N) \text{abs} \left( \sum_{t_{ij}=0}^{N_m} C_{mor}(t_i, T_k, F_b, F_c) A(t_i) \right) \quad (3.14)$$

де  $C(T_k, F_b, F_c)$  – значення модуля коефіцієнта вейвлет-перетворення,  $N$  – кількість дискретних відліків на інтервалі  $\delta T$  часового вікна.

При фіксованому параметрі ширини  $F_b$  комплексного вейвлета Морле значення модуля є функцією частоти  $F_c$  вейвлета Морле і положення часового вікна за часом -  $T_k$ . Типовий графік просторової скейлограми  $C(T_k, F_b, F_c)$  у функції  $F_c$  та  $T_k$  представлений на рис. 3.4.

Розглянуте подання фрагмента мови у вигляді просторової скейлограми володіє важливими особливостями, що дозволяють істотно підвищити

ефективність виявлення самоподібних структур. Зокрема, локальні максимуми вейвлет-перетворення є досить інформативними для аналізу атомарних складових мультифракталів в мовному сигналі.

На рис. 3.5 наведений рисунок, на якому поєднані фрагменти аудіозапису мови і скейлограма, побудована на основі розглянутої моделі.

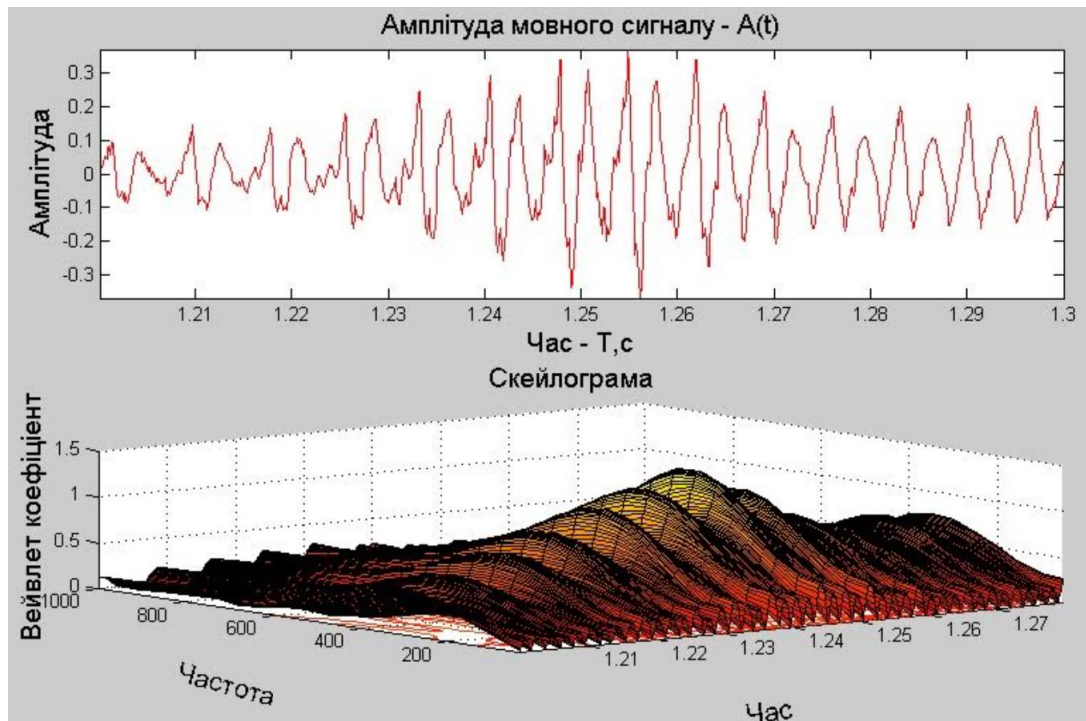


Рис. 3.5. Ілюстрація виявлення самоподібних структур

Важливо відзначити для подальшого аналізу наступні моменти дослідження. Крок за частотою  $F_c$  далі розглядається незалежним від розмірів вікна перетворень. Цей крок буде рівний 1 Гц. Дозвіл по частоті відповідно до принципу Гейзенберга [35, 104] не накладає істотних обмежень на аналіз формантної структури на малих часових інтервалах. Справа в тому, що відстань між максимумами формант в спектрі мови на малих часових інтервалах відповідає частоті основного тону мови (близько 85-600 Гц). А дозвіл по частоті близькості двох максимумів на інтервалі перетворення 20 мс - близько 50 Гц. При кроці по частоті перетворень 1 Гц точність оцінок відстаней між максимумами формант на малих часових інтервалах цього ж порядку. Дана точність оцінок не суперечить принципу невизначеності Гейзенберга для інформаційних технологій.

Необхідно також відзначити, що застосування вейвлета Морле не створює проблем фізичної трактування частотних перетворень. Оскільки цей вейвлет є віконне перетворення Фур'є з функцією Гауса [35].

Аналіз показує, що розташування «гребнів» скейлограм по часовому параметру на рис. 3.5 строго відповідає локальним максимумам амплітуди звукової хвилі в часовій області. При цьому локальні максимуми відповідають сплескам амплітуди звукової хвилі, зумовленими частотою ОТ.

Важливим фактором високої спорідненості базису Морле з самоподібними структурами в мовних фрагментах є більш високий ступінь гладкості скейлограми в порівнянні, наприклад, з аналогічним перетворенням Фур'є. Більш висока ступінь гладкості функцій забезпечує досить ефективну можливість математичного аналізу параметрів «гребнів».

При цьому максимуми формант по частоті істотно можуть відрізнятися від їх максимумів в спектрі Фур'є.

Відстані між локальними частотними максимумами скейлограм є в даному підході оцінкою частоти основного тону мови. Важливим фактором стійкості і достовірності оцінок частоти ОТ для цього методу є можливість оцінки частоти ОТ не тільки по локальних максимумах, але і по кореляції між фрагментами областей максимумів. На малих інтервалах ці області є приблизно самоподібними структурами. При аналізі такої самоподібності можливо довільним чином «вирізати» однакові за розмірами часового вікна структури не орієнтуючись на максимуми скейлограм [18].

Так, наприклад, розглянемо мовної фрагмент тривалістю 1 с, при частоті дискретизації 44100 Гц. Будемо аналізувати його спектральні характеристики і параметри частоти основного тону з використанням часового вікна в 20 мс на основі даної моделі. Прийmemo мінімально можливий крок по дискретності переміщення вікна аналізу рівним величині зворотного частоті дискретизації. Тоді, кількість оцінок віконних спектрів і параметрів частоти основного тону за 1 с. буде порядку 44080. Ці значення корельовані між собою. Але подібна кількість статистик для фрагментів мови невеликої тривалості відкриває

шляхи для отримання достовірних і стійких оцінок характеристик мови для повідомлень невеликої тривалості.

Суттєвою особливістю характеристик «гребнів» є форма «гребня». Дослідження показують, що після раціонального нормування функції  $C(T_k, F_b, F_c)$  при фіксованих  $T_k$  та  $F_b$  на вершині «гребня» ці функції мають високий ступінь геометричної подоби. При цьому форма нормованих «гребнів» індивідуально відмінна при відмінності мовних характеристик.

Причиною подібності структур типу «гребінь» в частотній області на скейлограмі для одних і тих же мовних характеристик є специфіка спорідненості вейвлета Морле часовій структурі амплітуди звукової хвилі в областях локальних максимумів. Вейвлет-перетворення Морле ефективно при даному підході виділяє ці структури. Розглянута подоба має вельми прозоре фізичне трактування. Структури амплітуди звукової хвилі в області локальних максимумів, відповідних частоті  $OT$ , мають достатньо явно виражену геометричну симетрію щодо амплітуди локального максимуму. При цьому вейвлет Морле в силу спорідненості дозволяє виявити цю симетрію у вигляді явно виражених екстремумів скейлограм.

Перший рівень вейвлет декомпозиції МС дозволяє отримати перший рівень самоподібних структур, але зважаючи на те, що наявність особливостей сигналу, що характеризують МС конкретної людини, може лежати в наступних рівнях розкладання сигналу. Тому, пропонується розглянути також наступні рівні декомпозиції МС для визначення характеристик самоподібних структур на основі аналізу мультифрактального спектру.

Розглянемо принципи визначення максимумів вейвлет коефіцієнтів, що будуть використовуватися для отримання характеристик самоподібних структур. Часова локалізація вершин вейвлет максимумів потребує отримання вейвлет коефіцієнтів з використанням компактних вейвлетів. Для кількісного визначення локальних особливостей самоподібних структур раціонально використовувати експоненту Гельдера. Вейвлет спектр особливостей буде

вказувати на величину набору показників Гельдера в мовному сигналі. Характерні максимуми самоподібних структур (рис.3.6) визначаються як [105]

$$L_x(j, k) = \sup_{\lambda' \in 3\lambda} |d_x(j, k)| \quad (3.15)$$

де  $j$  – ступінь вейвлет декомпозиції,  $k$  – час,  $d_x(j, k)$  – вейвлет коефіцієнти. При визначенні максимумів сусідніми елементами вважаємо  $3\lambda_{j,k} = \lambda_{j,k-1} \cup \lambda_{j,k} \cup \lambda_{j,k+1}$ .

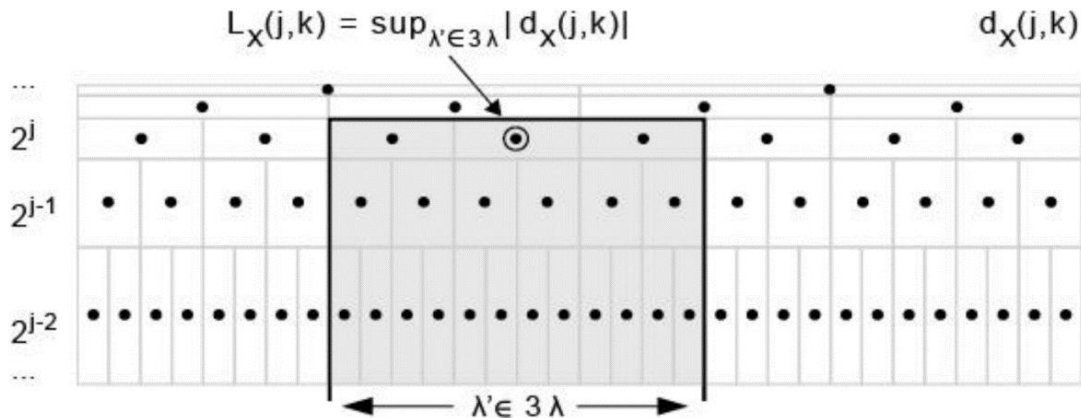


Рис. 3.6. Визначення характерних вейвлет максимумів

Для виділення самоподібних структур  $L_x(j, k)$  використовуємо наступний алгоритм відповідно до теорії мультифрактального спектру [106]:

1. Використовуючи дискретне вейвлет перетворення обчислимо вейвлет-коефіцієнти  $d_x(j, k)$  та зберігаємо їх значення для кожної шкали. Особливістю проведення вейвлет перетворення є подвоєння кількості коефіцієнтів між шкалами (рис.3.7).

6		2		7		5	
4	3	5	2	1	0	4	3

Рис. 3.7. Приклад коефіцієнтів вейвлет перетворення на двох рівнях декомпозиції

2. Починаємо розгляд зі шкали, яка на 1 рівень грубіша ніж шкала самої точної декомпозиції.



3. Порівнюємо перше значення з усіма інтервалами і отримуємо максимальне значення (рис.3.8).

	6		2		7		5
4	3	5	2	1	0	4	3
	6		5		7		5

Рис.3.8. Приклад визначення максимуму для двох рівнів вейвлет декомпозиції

4. Переходимо до наступного значення та порівнюємо це значення з усіма його сусідніми значеннями вейвлет декомпозиції. (рис.3.9).

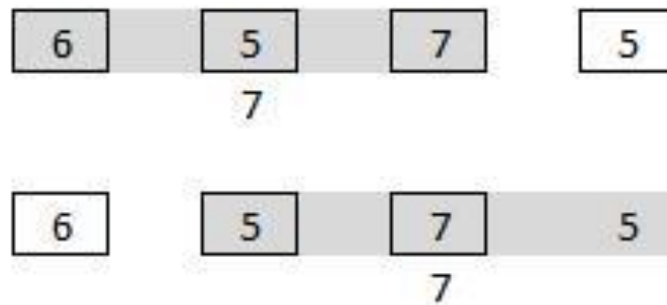


Рис. 3.9. Приклад покрокового визначення максимумів уздовж вейвлет-коефіцієнтів

5. З максимальних значень, отриманих для однієї шкали, визначаються максимум перших 3 значення. Це максимальне значення є особливістю самоподібної структури на цьому рівні вейвлет декомпозиції.

6. Переходимо до наступного рівня декомпозиції, та повторюємо процес визначення максимумів самоподібних структур.

Для визначення характерних особливостей самоподібних структур скористаємося описом МС у вигляді мультифрактального спектру. Алгоритм визначення характеристик самоподібних структур МС представлено на рис. 3.10.

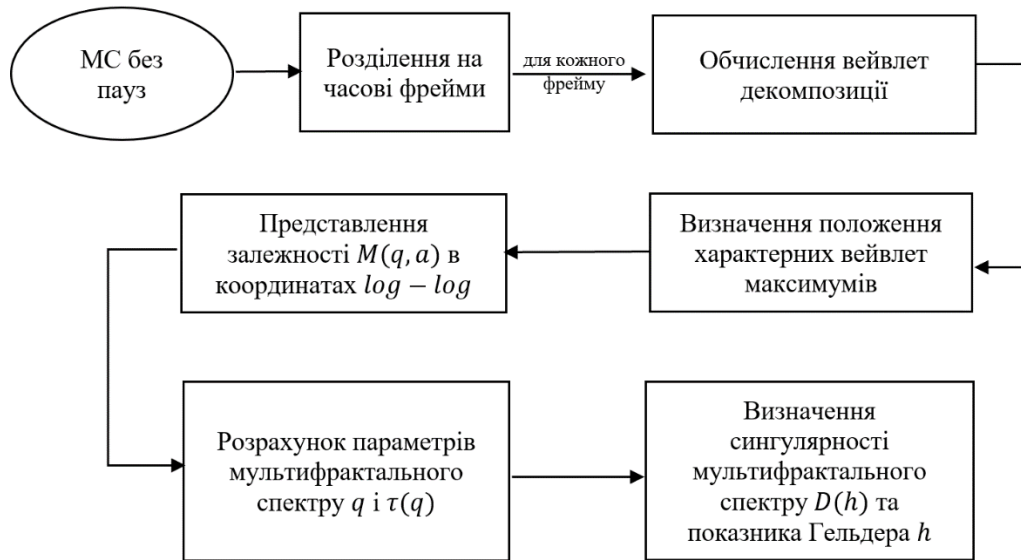


Рис. 3.10. Алгоритм визначення характеристик самоподібних структур МС

В більшості літератури мультифрактальним спектром називається залежність  $\tau(q)$ , але існують роботи, в яких мультифрактальним спектром також називають залежність показника Гельдера  $h$  від сингулярності мультифрактального спектру  $D(h)$ . В результаті обчислення представленим алгоритмом отримані залежності сингулярності мультифрактального спектру МС будуть виглядати наступним чином (рис. 3.11). Ці залежності представляють собою характеристики самоподібних структур МС.

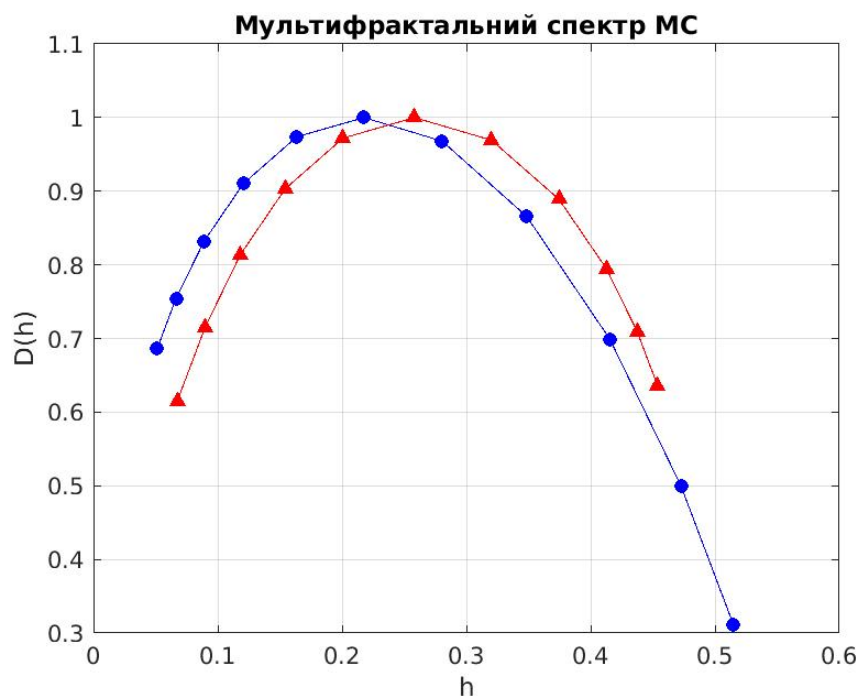


Рис.3.11. Залежність сингулярності мультифрактального спектру двох МС

Згідно виконаного в розділі 1 аналізу сучасного стану досліджень задач ідентифікації мовної інформації визначено, що основними ознаками, що більш точно описують мовні характеристики особи є частота ОТ та частоти формант. Відомо, що мовні експерти в якості ідентифікації мовних сигналів осіб виконують пошук ударних голосних та порівняння їх спектрограм. Тому, основну, найбільш важливу інформацію, несуть в собі вершини піків спектрограм.

Зважаючи на це, запропонуємо наступний метод підвищення інформативності частоти ОТ, що враховує характеристики самоподібних структур:

1. Сегментуємо вокалізований МС на часові інтервали по 20 мс.
2. Визначаємо мультифрактальний спектр сегменту МС.
3. Обчислюємо коефіцієнти вейвлету Морле для кожного сегменту МС на основі (3.13).

4. Для визначення положення кожної окремої вершини піку тривимірного спектру Морле на даному інтервалі виконуємо порівняльний аналіз площини під кривою кожного зрізу тривимірного спектру Морле для кожного інтервалу:

$$S_{i-1} < S_i < S_{i+1} \quad (3.16)$$

де  $S$  – площа зрізу тривимірного спектру Морле на відповідному часовому вікні  $t$ .

5. На визначеному зрізі (з максимальним значенням площини на інтервалі) тривимірного спектру Морле, знаходимо координати 7 максимумів спектрограми та обчислюємо різницю між сусідніми максимумами за частотною шкалою, і обчислюємо мультифрактальний спектр.

6. Зберігаємо в базі для кожного сегменту розбиття значення різниць між сусідніми максимумами та мультифрактального спектру.

7. На етапі порівняння двох МС визначаються два типи залежностей частоти ОТ:

7.1. перша залежність - по результатам отриманих значень різниць максимумів будуємо нормовану гістограму розподілу частоти ОТ.

7.2. друга залежність - по результатам отриманих значень різниць максимумів будуємо нормовану гістограму розподілу частоти ОТ тільки для тих фрагментів, які мають близькі значення сингулярності мультифрактального спектру двох МС.

Більш детально побудова залежностей частоти ОТ буде описано в наступному розділі.

На основі розроблених вище методів та алгоритму визначимо метод мовної ідентифікації особи в наступному вигляді:

1. Сегментація досліджуваного мовного файлу на вокалізовані та невокалізовані фрагменти запропонованим методом сегментації на основі фрактальної розмірності;

2. Виділення вокалізованих фрагментів мовного файлу, згідно п.1 та окремий розгляд кожного з них в наступних кроках;

3. На основі алгоритму виділення характеристик самоподібних структур отримання з фрагментів п.2 характерних ознак особи і збереження їх в базі даних для подальшого дослідження;

4. Визначення формантних частот;

5. Для порівняння формуються наступний набір критеріїв:

– залежності частоти ОТ на основі різниць максимумів на частотній шкалі з урахуванням та окремо залежностей без урахування характеристик мультифрактального спектру.

– сукупність залежностей формантних частот.

6. Порівняння методом ранжування, згідно набору критеріїв, характерних ознак особи, виділених в п.3, зі збереженими в базі наборами інших осіб;

7. Відображення найбільш близького набору характерних ознак особи, згідно ранжирування з п.4, у вигляді ймовірнісної відповідності з набором ознак особи, що перевіряється.

### 3.4. Визначення частотних характеристик формант

У системах ідентифікації мовного сигналу отримання характерних ознак мови особи, є одним з основних запорук успіху роботи системи ідентифікації мовної інформації. У завданнях аналізу мовного сигналу з метою визначення найважливіших характеристик, як правило, використовують методи його частотно-часового або спектрального уявлення, одними з найбільш ефективних є методи вейвлет перетворення мовного сигналу. Найважливішим параметром, що характеризує спектр (розподіл енергії або амплітуди по частотах) мовного сигналу є форманти [107], які визначають як концентрацію енергії в обмеженій частотній області. Форманта характеризується частотою, шириною частотної смуги і амплітудою. Зважаючи на складність визначення і опису формантної частотної смуги часто в дослідженнях під частотою форманти мають на увазі частоту максимальної амплітуди в межах форманти. Тому часто можна зустріти визначення форманти, як амплітудний сплеск на графіку спектра [108], що має частоту, відповідну до частоти піку цього сплеску. Форманти прийнято позначати F0-F6. Форманта F0 називається також частотою основного тону мовного сигналу. Форманти F1 - F6 визначають концентрацію енергії мовного сигналу по частоті і характеризують вокалізовані (як правило, голосні) звуки.

Існує безліч методів виявлення формант [109,110,111], основними з яких є кепстральний аналіз і метод лінійного передбачення. Кепстральний аналіз є математичною основою нелінійних методів виділення сигналів. Кепстр - це математичне перетворення, що полягає в тому, що спектральному перетворенню піддається спектр функції. Якщо взяти спектр від спектра, то вийде сама початкова функція. Природно, що для отримання більш гладкої функції в результаті спектрального перетворення необхідно згладити початковий спектр. Для цього найбільш часто використовується логарифмування початкового спектру або його модуля. Такий варіант

перетворення прийнято називати кепстр. В отриманому кепстрі виявляються гармонійні складові [112].

Метод лінійного передбачення (метод, який використовує LPC-коефіцієнти) полягає в пошуку комплексних коренів полінома з LPC-коефіцієнтами (коефіцієнти фільтру мовного тракту) і надалі їх наступні перетворення. Вважається, що метод лінійного передбачення [9] забезпечує високу обчислювальну швидкість, незначну складність і максимальну точність оцінювання формант.

Зазвичай використовується ефективний метод обчислення коефіцієнтів лінійного передбачення, заснований на автокореляції, що використовує алгоритм Левінсона-Дарбина. У задачі розпізнавання мови часто застосовуються коефіцієнти лінійного передбачення, обчислені на основі статичного закону людського слуху [113,114]. Відмінності від простого обчислення коефіцієнтів лінійного передбачення полягають в наступному. По-перше, застосовується шкала Баркова і логарифмічна компресія амплітуди до застосування алгоритму Левінсона-Дарбина. По-друге, для відповідності ступеневим законом людського слуху потужність спектральних компонент зводиться до ступеня 0.33. Ця модифікація застосовується в частотній області, що призводить до того, що автокореляційні коефіцієнти не можуть бути обчислені безпосередньо, отже, необхідно застосувати додаткове перетворення Фур'є [115].

Описані методи використовуються в різних комбінаціях в чотирьох основних формант-трекерів PRAAT [116], SNACK [117], ASSP [118] та DEEP [119].

PRAAT [92] виділяє форманти наступним чином: для кожного фрагменту, що аналізується, застосовується вікно подібне до вікна Гауса. Обчислюються коефіцієнти LPC за алгоритмом Бурга [120]. Кількість максимумів, яку обчислює цей алгоритм вдвічі перевищує максимальну кількість формант, тому рекомендується встановлювати максимальну кількість формант кратну до 0,5. Спочатку алгоритм знаходить максимальну

кількість формант у всьому діапазоні від 0 Гц та вище. Тому знайдені форманти можуть мати аномально низькі або високі частоти, що пов'язано з артефактами алгоритму LPC.

SNACK [93] застосовує той самий підхід, що й PRAAT. Цей метод використовує відбір формантів за мінімальними значеннями вагових коефіцієнтів в основі яких лежать відмінності між фрагментами, частота та ширина форманти. На відміну від PRAAT використовується лише один параметр, пов'язаний з відстеженням функції. Цей параметр дозволяє визначити значення першої форманти, вищі форманти визначаються на основі заданого значення для першої форманти F1. Таким чином метод вважає, що всі інші форманти пов'язані з F1 коефіцієнтами прямого зв'язку. Цей метод не дає можливості змінити поведінку обчислення або вагові коефіцієнти для інших формант.

В методі ASSP [94] значення формант отримують шляхом визначення резонансних частот за рахунок розв'язання коренів поліному LPC, використовуючи метод автокореляції на основі алгоритму Спліт-Левінсона (SLA) [121]. Потім виконується класифікація резонансних частот як формант, використовуючи частоти Писаренко [122] та встановлення границі діапазону частот формант, що визначаються з номінальної частоти першої форманти F1. В методі вказується, що форманті частоти необхідно збільшити на 12 % для більш точного обчислення формант жіночих голосів.

DEEP [95] використовує нейронні мережі для визначення формант. Вхідними даними є отримання на основі LPC метода кепстральні коефіцієнти, які корегуються частотою основного тону, отриманої зі спектрограми. Навчання нейронної мережі виконувалось на основі тестової частини мовного датасету VTR-TIMIT (датасет 8 жінок та 16 чоловіків) [123]. VTR-TIMIT є відкритим датасетом з 516 видами записів від 186 осіб, що є носіями англійської мови. В датасеті виконана ручна корекція та маркування формант спеціальної групою експертів, тому саме цей датасет найчастіше використовується для оцінки точності роботи формант-трекерів.

Існують ряд досліджень, що стосуються порівняння формант-трекерів, однак серед них не можна виділити такий, що має найкращу ефективність [124,125].

На основі вищевикладеного, можна зробити висновок в деякій складності, а частково, навіть у відсутності фізичної аргументації механізмів визначення формант, а також відсутності достатньої точності формант-трекерів, що потребує побудови методу визначення формант, маючого достатнє фізичне обґрунтування та достатню точність в порівнянні з аналогами.

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

Дослідження формантних атрибутів найчастіше виконується шляхом:

1. проведення порівняння спектру формант однакових фонемічних елементів (ударних голосних, голосних в кінці або на початку слів та ін.);
2. проведення порівняння спектрів формант для зрізів спектрограм (кожного елементу спектрограми або всередині слів, складів та ін.);
3. проведення порівняння динамічних змін частот формант уздовж усього МС чи у важливих його компонентах.

Виділення формант супроводжує цілий ряд проблем, пов'язаних з їх динамічної зміною в процесі мови. Навіть однакові голосні змінюють формантний набір в залежності від свого розташування у складі слів, складів та ін. Складність також визивають цілий ряд проблем, пов'язаних з близьким розташуванням піків при аналізі спектрограм та проблемами правильного визначення піків максимумів формант на спектрограмі. Визначення розташування формант на спектрограмах МС достатньо легко виконується людиною, але автоматизація цього процесу визиває деякі складнощі. Типове представлення спектрограми з розміченими людиною розташування формант



представлено на рис. 1.12, що відповідає зонам розташування формант причому характеризується не тільки середньою частотою форманти, а також її шириною.

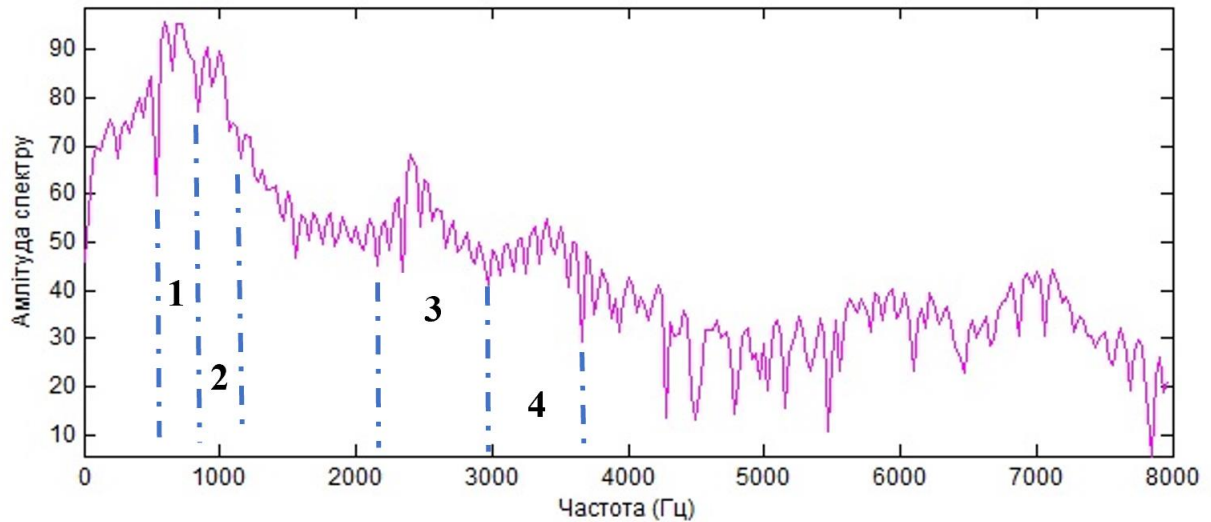


Рис. 3.12. Визначення формант людиною

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 данні, 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 з них.

Існує декілька підходів для визначення положень формант на частотній шкалі, але всі вони базуються на аналізі та перетвореннях спектрограми мовного сигналу. При виділенні формантних частот першим етапом дослідження завжди є побудова спектрограми за визначеними дослідником критеріями. Серед них є ширина фрейму, тип вейвлет-базису, частотний діапазон та ін.

Наступним етапом є таке представлення спектрограми, що дозволить провести сегментацію формантих частот на основі різних математичних методів, найчастіше всі вони базуються на алгоритмах кластеризації або на побудові огинаючої спектру. Найбільш відомими з алгоритмів, що використовують огинаючу спектру є:

1. метод лінійного передбачення – коли огинаюча спектру будується на алгоритмі LPC [126];
2. апроксимації спектру кубічними сплайнами або іншими видами функцій [127].

Однак згідно досліджень обидва алгоритми мають практично однакові точність при різній обчислювальній складності [128].

Приклад огинаючої на основі алгоритму лінійного передбачення виділеного спектру для стандартного фрагменту дослідження (20мс) представлений на рис. 3.13. На основі цього рисунку виділяються локальні максимуми огинаючої для спектру мовного сигналу, які розглядаються як центри формантих частот.

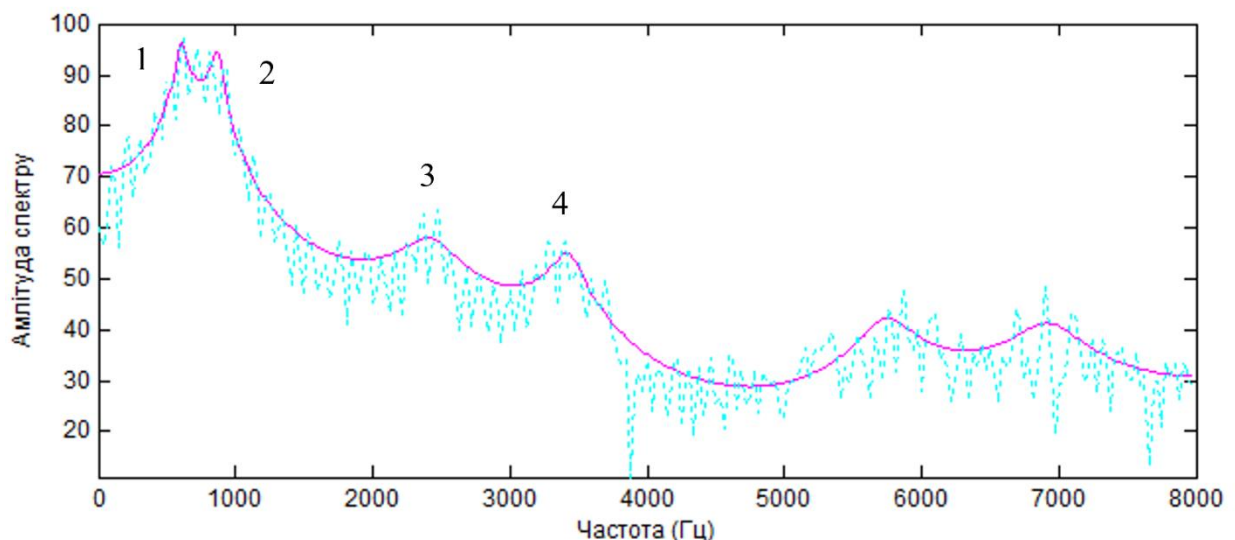


Рис. 3.13. Побудова огинаючої для спектру мовного сигналу  
(числами зображено формантні діапазони 1-4)

Проведені попередні дослідження вказують на відповідність розподілу частот відносно голосних звуків, які вносять найбільшу вагу в формування

формант в мовному сигналі. Встановлено, що частина голосних звуків в більшості мов, розташована в частотному діапазоні 200 – 500 Гц, а інша частка голосних звуків лежить в діапазоні від 500 до 1500 Гц. Зважаючи на це, раціональним є окремий розгляд цих частотних діапазонів, при формуванні характерних ознак мовного сигналу, що дозволяє підвищити кількість параметрів, та набирати більшу статистику при визначенні максимумів формантних частот.

Результатом проведеного огляду підходів до визначення формантних частот став алгоритм (рис. 3.14), що складається в наступному:

1. сегментація мовного сигналу на вокалізовані фрагменти та паузи;
2. розбиття вокалізованих фрагментів МС на часові фрейми;
3. для кожного фрагменту отримання спектру на основі вейвлет-перетворення;
4. побудова огинаючої лінії;
5. знаходження всіх максимумів;
6. визначення положень формантних діапазонів;
7. отримання максимумів формантних діапазонів;
8. побудова графіків траєкторії положення формант;
9. розрахунок залежності щільності вірогідності розподілу кожної з чотирьох формантних частот (максимумів формантних частот).



Рис. 3.14. Алгоритм визначення формантних частот

Сегментація МС на вокалізовані фрагменти та паузи виконується методом оцінки змін фрактальної розмірності [129]. В роботі визначено, що фрактальна розмірність  $D$  для невокалізованих фрагментів в 99 % перебуває в межах  $1,04 \leq D \leq 1,45$ , а фрактальна розмірність вокалізованих фрагментів не спостерігалася менш  $D = 1,55$  для часового вікна розміром 20 мс.

Операцію розбиття виділених вокалізованих фрагментів МС на часові фрейми виконувалось фреймами 10-20 мс для тестування роботоспроможності методу.

Отримання спектру мовного сигналу виконувалось з використанням комплексного вейвлету Морле на основі віконної функції Гауса.

Згідно розглянутого підходу до визначення формантних частот необхідно побудувати огинаючу спектру для кожного з фрагментів МС. Проведені в багатьох роботах дослідження при побудові огинаючої пропонують використання безлічі підходів, але при визначенні формантних частот саме кубічні сплайни дозволяють найточніше визначити вид цієї функції огинаючої за заданими значеннями спектру МС. Фактично побудова функції огинаючої представляє собою задачу інтерполяції. Визначимо, як ця задача вирішується у нашому випадку.

Розглянемо спектр МС, як функцію визначену в точках  $x_0, x_1, x_2, \dots, x_N$ , де відомі значення деякої функції  $f(x)$ , а саме  $y_0, y_1, y_2, \dots, y_N$ .

Інтерполяція графіку функції  $f(x)$  є побудова функції  $F(x)$ , такої, що приймає в зазначених точках ті ж значення, тобто

$$F(x_0) = y_0, F(x_1) = y_1, \dots, F(x_N) = y_N \quad (3.17)$$

З геометричної точки зору стоїть задача пошуку певного типу кривої  $y = F(x)$ , що проходить через набір представлених точок. Подібна задача може мати багато рішень або рішення можуть бути відсутні. При використанні кубічного сплайну наша функція  $F(x)$  представляє собою набір фрагментів, що визначені на кожному інтервалі  $[x_{k-1}; x_k]$  та може мати наступний вигляд:

$$F_k(x) = a_k + b_k(x - x_k) + c_k(x - x_k)^2 + d_k(x - x_k)^3$$

$$\begin{aligned}
 F &= F_1 \text{ для } [x_0; x_1] \\
 &\dots \\
 F &= F_N \text{ для } [x_{N-1}; x_N]
 \end{aligned}
 \tag{3.18}$$

де  $N$  – це кількість точок, за якими проводиться інтерполяція.

Звичайно що коефіцієнти поліному  $a_k, b_k, c_k, d_k$  будуть відрізнятися для кожного з відрізків  $[x_{k-1}; x_k]$ .

Для визначення коефіцієнтів накладається ряд умов, серед яких:

1. рівність других похідних функції на кінцях відрізка  $[x_0; x_N]$ :

$$F''(x_0) = 0, F''(x_N) = 0 \tag{3.19}$$

2. безперервності першої і другої похідної функції  $F(x)$ , а також тобто

$$\left. \begin{aligned}
 F_{k-1}(x_{k-1}) &= F_k(x_{k-1}) \\
 F'_{k-1}(x_{k-1}) &= F'_k(x_{k-1}) \\
 F''_{k-1}(x_{k-1}) &= F''_k(x_{k-1})
 \end{aligned} \right\} \tag{3.20}$$

при  $k = 2, 3, \dots, N$

Запишемо похідні функції  $F_k$  у вигляді

$$\left. \begin{aligned}
 F'_k(x) &= b_k + 2c_k(x - x_k) + 3d_k(x - x_k)^2 \\
 F''_k(x) &= 2c_k + 6d_k(x - x_k)
 \end{aligned} \right\} \tag{3.21}$$

Зважаючи на описані вище умови, отримуємо наступну систему рівнянь:

$$\left. \begin{aligned}
 a_1 - b_1 h_1 + c_1 h_1^2 - d_1 h_1^3 &= y_0 \\
 a_k &= y_k, k = 1, 2, \dots, N \\
 a_{k-1} &= a_k - b_k h_k + c_k h_k^2 - d_k h_k^3, k = 2, 3, \dots, N \\
 b_{k-1} &= b_k - 2c_k h_k + 3d_k h_k^2, k = 2, 3, \dots, N \\
 c_{k-1} &= c_k - 3d_k h_k, k = 2, 3, \dots, N \\
 c_1 - 3d_1 h_1 &= 0 \\
 c_N &= 0
 \end{aligned} \right\} \tag{3.22}$$

де  $h_k = x_k - x_{k-1}$ ,  $k = 1, 2, \dots, N$ ,  $l_k = (y_k - y_{k-1})/h_k$ ,  $c_0 = 0$

Існує досить багато варіантів розв'язання представленої системи рівнянь. Для розв'язання будемо використовувати метод прогонки [130,131].

В основі метода лежить використання коефіцієнтів, що будуть корегуватися в процесі підлаштування. Визначимо коефіцієнти в наступному вигляді:

$$\left. \begin{aligned} \delta_1 &= -h_2/(2(h_1 + h_2)) \\ \lambda_1 &= 3(l_2 - l_1)/(2(h_1 + h_2)) \\ \delta_{k-1} &= -h_k/(2h_{k-1} + 2h_k + h_{k-1}\delta_{k-2}), k = 3, 4, \dots, N \\ \lambda_{k-1} &= (3l_k - 3l_{k-1} - h_{k-1}\lambda_{k-2})/(2h_{k-1} + 2h_k + h_{k-1}\delta_{k-2}) \end{aligned} \right\} \quad (3.23)$$

На основі коефіцієнтів прогону отримуємо  $c_k$  за алгоритмом зворотного прогону

$$c_{k-1} = \delta_{k-1}c_k + \lambda_{k-1}, k = N, N - 1, N - 2, \dots, 2 \quad (3.24)$$

Все це дає можливість визначити коефіцієнти  $b_k$  і  $d_k$  за формулами

$$\left. \begin{aligned} b_k &= l_k + (2c_k h_k + h_k c_{k-1})/3, k = 1, 2, \dots, N \\ d_k &= (c_k - c_{k-1})/(3h_k), k = 1, 2, \dots, N \end{aligned} \right\} \quad (3.25)$$

Таким чином, результатом описаного алгоритму будуть коефіцієнти  $b_k, c_k, d_k$  для кожного з розглянутих інтервалів графіку спектру МС.

Результатом побудови огинаючої буде сумісний графік спектру МС, в заданому фрагменті, та огинаючої спектру для цього ж інтервалу (рис. 3.15).

Виконуючи визначення максимумів огинаючої спектру (рис. 3.15) отримуємо максимуми чотирьох перших формант. Набір формант для кожного фрейму відображається на графіку відповідно до частоти форманти та номеру фрейму, з якого вона була отримана (рис. 3.16). Аналіз даного графіку повинен показувати стабільність визначення формантих частот для розглянутого при попередньому дослідженні мовного сигналу.

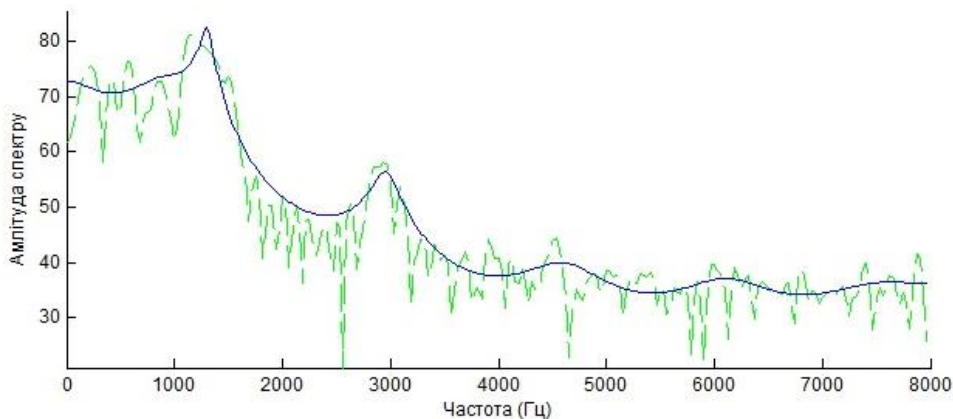


Рис. 3.15. Огинаюча спектру МС

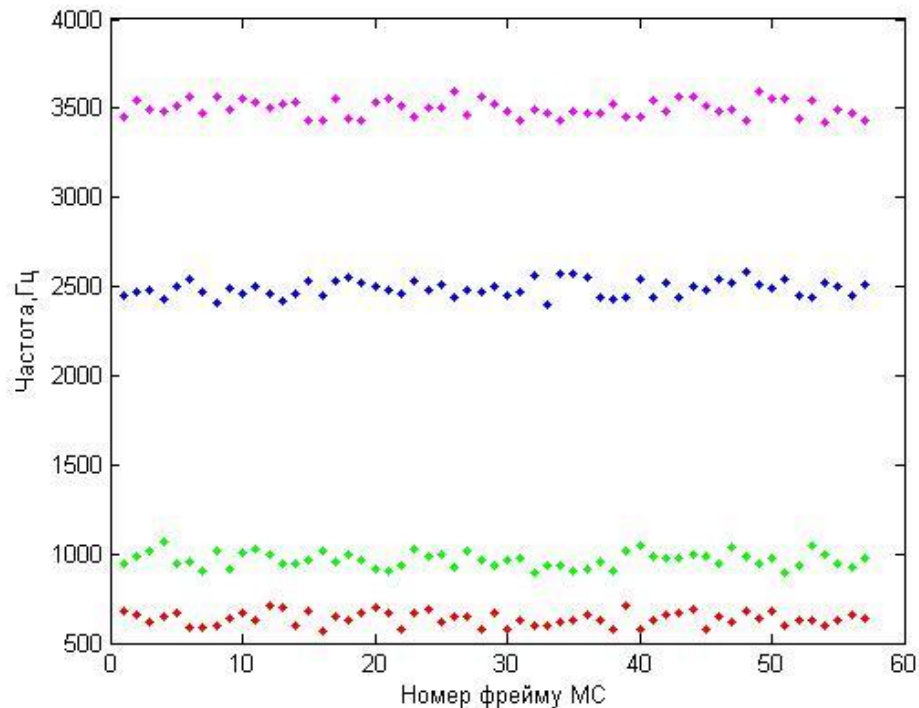


Рис. 3.16. Траекторія положення формант за номерами фреймів

### 3.5. Порівняння характерних ознак мовного сигналу

Ідентифікація особи на основі отриманих ознак мовного сигналу повинна проводитися шляхом їх порівняння між собою. В даному дослідженні прийняття рішення щодо відповідності двох мовних сигналів одній особі виконується на основі порівняльного аналізу кривих щільності вірогідності отриманих статистичних характеристик самоподібних структур.

При виконанні порівняння, по отриманим значенням максимумів вейвлет розкладання  $C_1(T_k, F_b, F_c)$  та  $C_2(T_k, F_b, F_c)$ , будуються скейлограми  $S_1(F)$  та  $S_2(F)$ . Згідно описаних вище методу підвищення інформативності частоти ОТ та алгоритму визначення частот максимумів формант розраховуються щільності вірогідностей у вигляді

$$p(F) = \int S(F)d(F) \quad (3.26)$$

Тобто отримуємо набір щільностей вірогідності для частоти ОТ  $F_0, F_{0MF}$  та максимумів частот формант  $F_1, F_2, F_3, F_4$  для кожного МС –  $p_1^{F_0}(F), p_1^{F_1}(F), p_1^{F_2}(F), p_1^{F_3}(F), p_1^{F_4}(F)$  та  $p_2^{F_0}(F), p_2^{F_1}(F), p_2^{F_2}(F), p_2^{F_3}(F), p_2^{F_4}(F)$ .

Введемо для їх порівняння наступну міру подоби, узагальнив коефіцієнти (1,2) як  $(i, j)$ .

$$\begin{aligned} \Delta(i, j) = & \gamma \left( \sum_{t=1}^N |p_i^{F_{0MF}}(F_t) - p_j^{F_{0MF}}(F_t)| + \sum_{t=1}^N |p_i^{F_0}(F_t) - p_j^{F_0}(F_t)| \right) + \\ & + \sum_{t=1}^N |p_i^{F_1}(F_t) - p_j^{F_1}(F_t)| + \sum_{t=1}^N |p_i^{F_2}(F_t) - p_j^{F_2}(F_t)| + \\ & + \sum_{t=1}^N |p_i^{F_3}(F_t) - p_j^{F_3}(F_t)| + \sum_{t=1}^N |p_i^{F_4}(F_t) - p_j^{F_4}(F_t)| \end{aligned} \quad (3.27)$$

де  $\gamma$  це коефіцієнт, що регулює ступень впливу частоти ОТ відносно максимумів частот формант,  $N$  – кількість точок розбиття щільності вірогідності,  $p_i^{F_{0MF}}(F), p_i^{F_0}(F), p_i^{F_1}(F), p_i^{F_2}(F), p_i^{F_3}(F), p_i^{F_4}(F)$  – значення щільності вірогідності для частоти ОТ з урахуванням та без урахування характеристик мультифрактального спектру і чотирьох формант, що відповідає особі  $i$ ,  $p_j^{F_{0MF}}(F), p_j^{F_0}(F), p_j^{F_1}(F), p_j^{F_2}(F), p_j^{F_3}(F), p_j^{F_4}(F)$  – значення щільності вірогідності для частоти ОТ з урахуванням та без урахування характеристик мультифрактального спектру і чотирьох формант, що відповідає особі  $j$ .

Тоді порівняння осіб  $i$  і  $j$  може здійснюватись за допомогою визначального правила при заданому пороговому значенні  $\Delta_{lim}$

$$g(i, j) = \begin{cases} 1, & \text{якщо } \Delta(i, j) < \Delta_{lim} \\ 0, & \text{інакше} \end{cases} \quad (3.28)$$

Підбір параметрів  $\gamma, \Delta_{lim}$  буде описано більш детально у розділі 4.

Проведені попередні дослідження статистичних характеристик особи, отримані на основі самоподібних структур, показали стійку залежність



щільності вірогідності від індивідуальних особливостей МС особи (рис. 3.17, 3.18).

Представлені залежності щільності вірогідності на рис. 3.17 показують представлення цієї залежності для двох МС різних осіб. Видно, що центральні піки частоти ОТ приходяться відповідно на 248 та 398 Гц. Зазвичай існуючі методи використовують саме ці одиничні параметри, але зважаючи на високу варіативність МС особи необхідно враховувати розподіл частоти ОТ вздовж всього сигналу, саме тому в цьому дослідженні використовуюся щільності вірогідності.

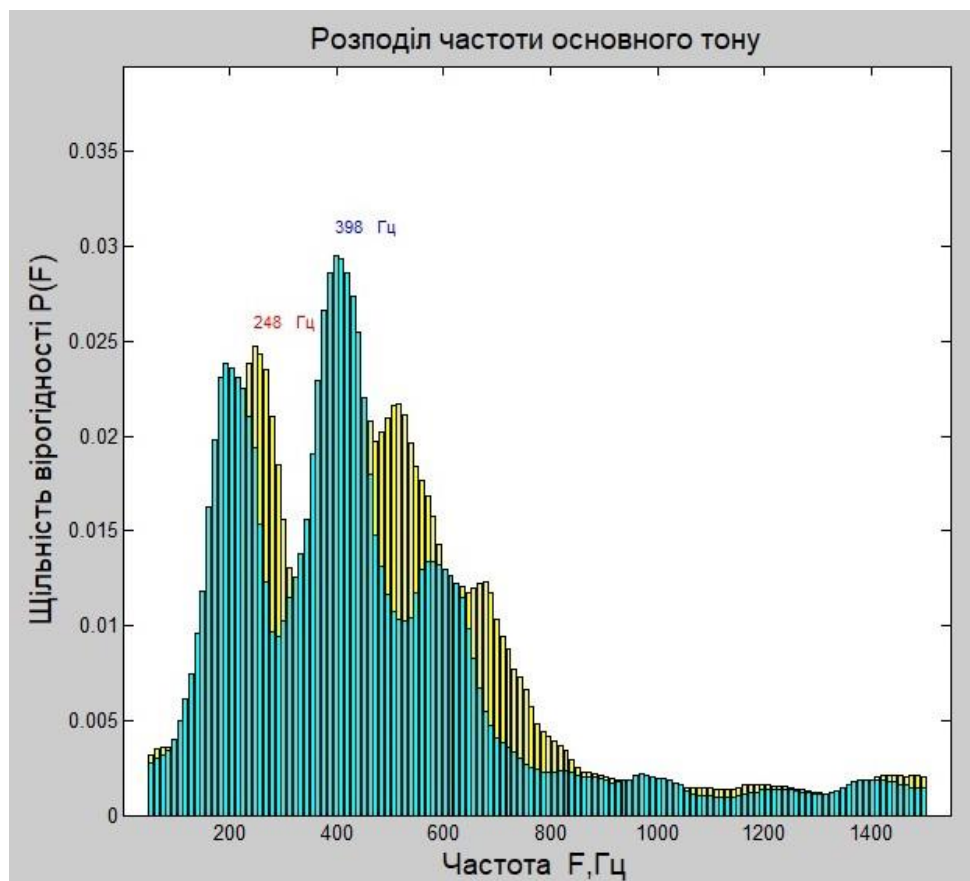


Рис. 3.17. Розподіл щільності вірогідності максимумів самоподібних структур для частоти ОТ для двох МС різних осіб

За необхідність використання щільності вірогідності також говорить рис.3.18, де представлені МС однієї і тієї ж особи з двома пиками розподілу в 237 та 248 Гц відповідно (такий розподіл не дозволить ідентифікувати МС однієї особи, бо частота ОТ на них відрізняється).

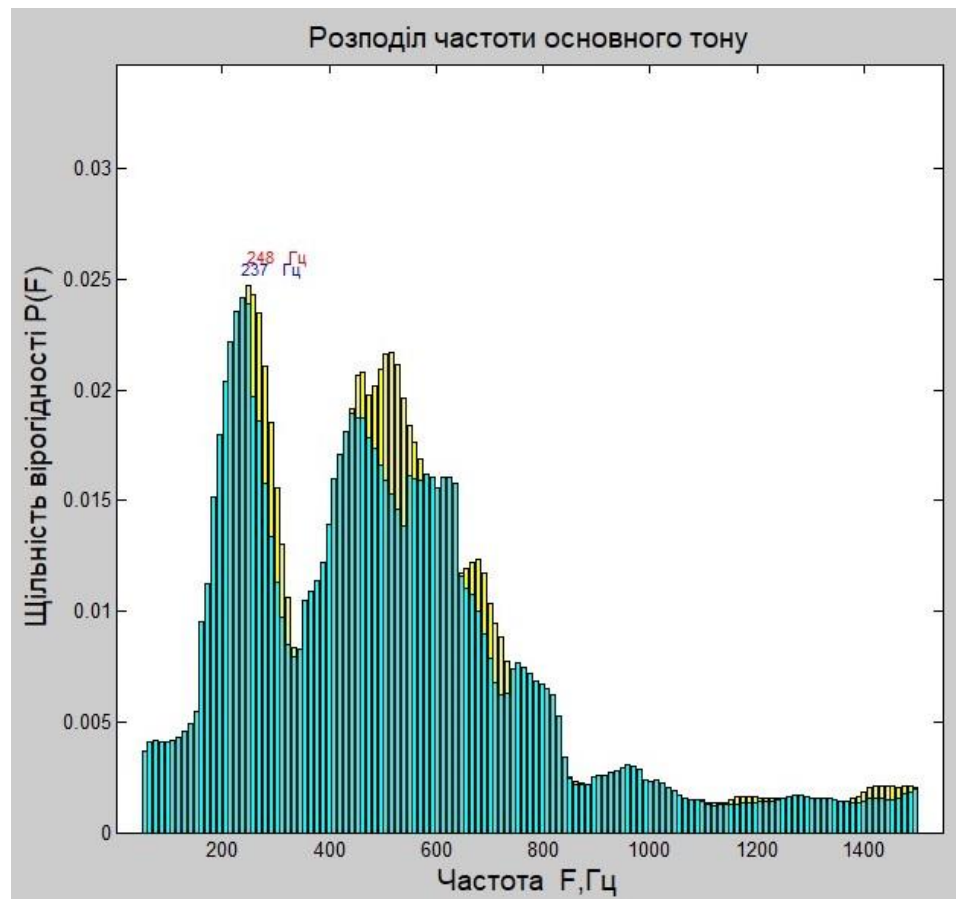


Рис. 3.18. Розподіл щільності вірогідності максимумів самоподібних структур для частоти ОТ для двох МС однієї особи

Таким чином, представлена закономірність розподілу щільності вірогідності максимумів самоподібних структур для частоти ОТ та чотирьох формант може бути використана в методі ідентифікації МС.

### Висновки за 3 розділом

В розділі 3 розроблено принципи використання фрактального та вейвлет аналізу в задачі ідентифікації особи, метод сегментації МС на вокалізовані фрагменти та паузи, метод підвищення частоти ОТ, а також алгоритм виділення характеристик самоподібних структур.

1. Виконано порівняльний аналіз методів обчислення фрактальної розмірності та визначено, що для задач сегментації мовного сигналу на вокалізовані та невокалізовані фрагменти, найбільш ефективним є метод

обчислення фрактальної розмірності по Хаусдорфу, на основі чого запропоновано метод сегментації МС.

2. Розроблено алгоритм виділення характеристик самоподібних структур МС, на основі максимумів вейвлет перетворення. При виділенні використовується комплексний вейвлет Морле, що найбільш ефективно описує самоподібні структури МС на першому рівні декомпозиції. Для отримання характеристик самоподібних структур на інших рівнях декомпозиції використовується мультифрактальний спектр.

3. Запропоновано метод підвищення інформативності частоти ОТ та алгоритм визначення формантих частот, що моделюють природний процес виділення інформаційних ознак особи з мовного фрагменту за рахунок фрактального та вейвлет аналізу.

4. Запропоновано міру подоби порівняння обчислених статистичних характеристик мовних сигналів для прийняття рішення ідентифікації МС. Показано, що використання в ролі статистичних характеристик щільностей вірогідностей характерних критеріїв ідентифікації МС дозволяють в достатній мірі виконати порівняння мовних сигналів.

## РОЗДІЛ 4

### ПІДБІР ПАРАМЕТРІВ ДЛЯ ПОБУДОВИ МЕТОДУ ІДЕНТИФІКАЦІЇ МОВНОЇ ІНФОРМАЦІЇ

#### 4.1. Ідентифікація мовного сигналу як клас задач загальної ідентифікації

В рамках задачі ідентифікації особи можна виділити два взаємопов'язані завдання ідентифікації (рис. 4.1) та верифікації (рис. 4.2) особи. У першій задачі мета полягає в ідентифікації фрагмента МС як виголошеного однією з осіб з розглянутої множини, у другому - встановлення належності мовного сигналу конкретній еталонній особі.

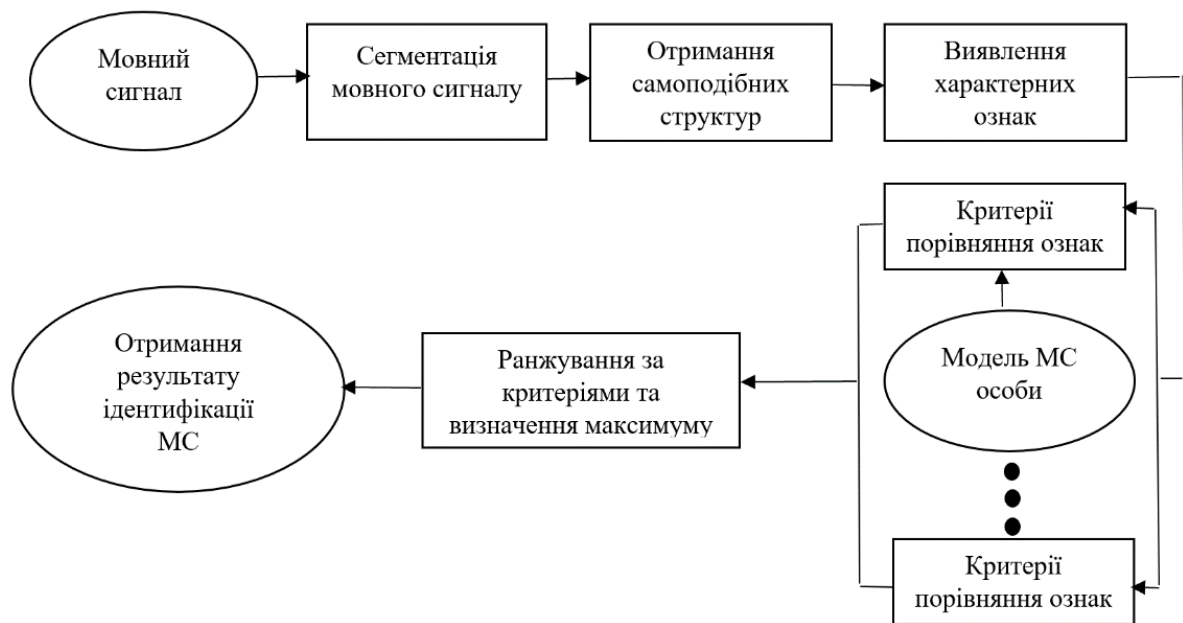


Рис. 4.1. Схема задачі ідентифікації МС

На підставі поданих схем система розбивається на три частини:

1. визначення індивідуальних ознак МС;
2. представлення характерного еталона особи;
3. прийняття рішення про індивідуальність мовця.

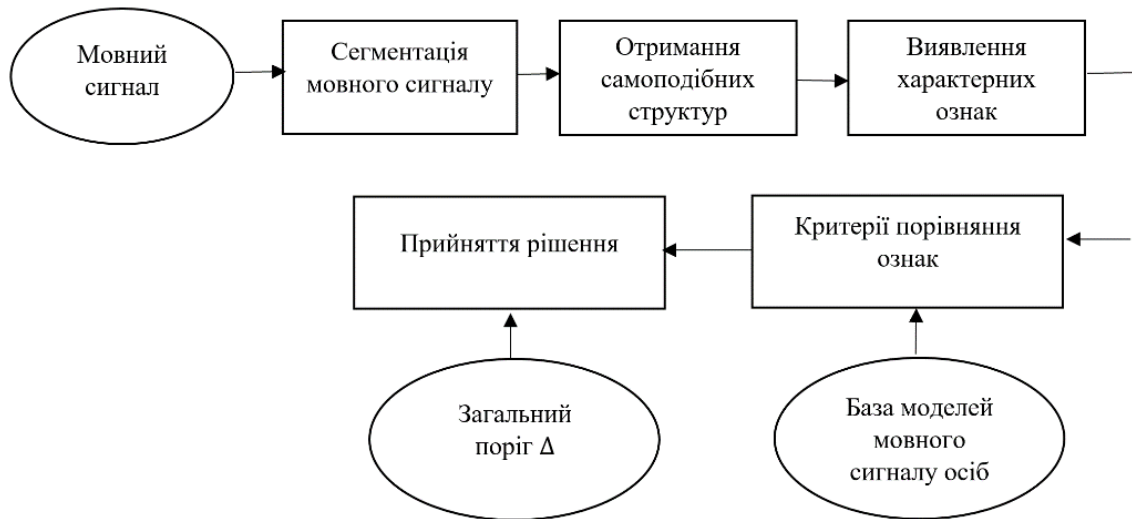


Рис. 4.2. Схема задачі верифікації

На підставі вищевикладеного можна виділити наступні основні етапи реалізації системи розпізнавання особи:

*Вимірювання фрактальної розмірності компонентів сигналу.* Простий у реалізації етап, але досить ефективний в наборі всіх заходів розрізнюваності. Реалізація його можлива як з постійним вікном, так і з адаптивним типом вікна.

*Визначення границь фрази.* Для вирішення даного завдання найбільш раціонально використовувати запропоновану в 3 розділі метод сегментації мови на основі мультифрактального підходу. На основі цього підходу в тих елементах сигналу, де зміна фрактальної розмірності перевищує деякий встановлений поріг, де передбачається, починається вокалізований фрагмент.

*Виділення основного тону та основних формант.* Для вирішення завдання виділення основного тону та формант існує необхідність використання методів слідкування за траєкторією ОТ та формант. В якості базового методу виділення може бути взятий підхід, заснований на використанні апроксимації сигналу на основі вейвлет-перетворення з подальшим статистичним аналізом розподілу вейвлет-максимумів, що фізично пояснюється наявністю самоподібних структур характерних для сигналів, пов'язаних з резонаторами. Для вирішення даного завдання найбільш

раціонально використовувати запропонований в 3 розділі метод підвищення інформативності частоти ОТ та виділення частот формант.

*Виділення характеристик основного тону та частот формант.* Для вирішення цього завдання можна скористатися знаходженням залежностей показника Гельдера від сингулярності мультифрактального спектру та гістограм розподілу частоти ОТ і частот формант.

*Порівняння параметрів сигналу з еталонними параметрами.* Після здійснення процесу порівняння параметрів мови з еталонними потрібно вибрати з бази найбільш «близьку» особу. Для цього необхідно порівняти виділені параметри основного тону з бази на основі ймовірнісного підходу.

#### **4.2. Вибір параметрів та дослідження ефективності методу сегментації**

Запропонований метод обчислення фрактальної розмірності може бути використаний у багатьох галузях аналізу мовних сигналів. Як приклад розглянемо використання пропонованого методу в завданні сегментації мови. З огляду на структурну складність завдання, була запропонована наступна методологія проведення досліджень (рис. 4.3).

При проведенні дослідження ефективності сегментації в якості ознак прийняття рішення «вокалізований/не вокалізований фрагмент» були обрані:

1. середньоквадратичне відхилення фрактальної розмірності;
2. діапазон розмірів часового вікна.

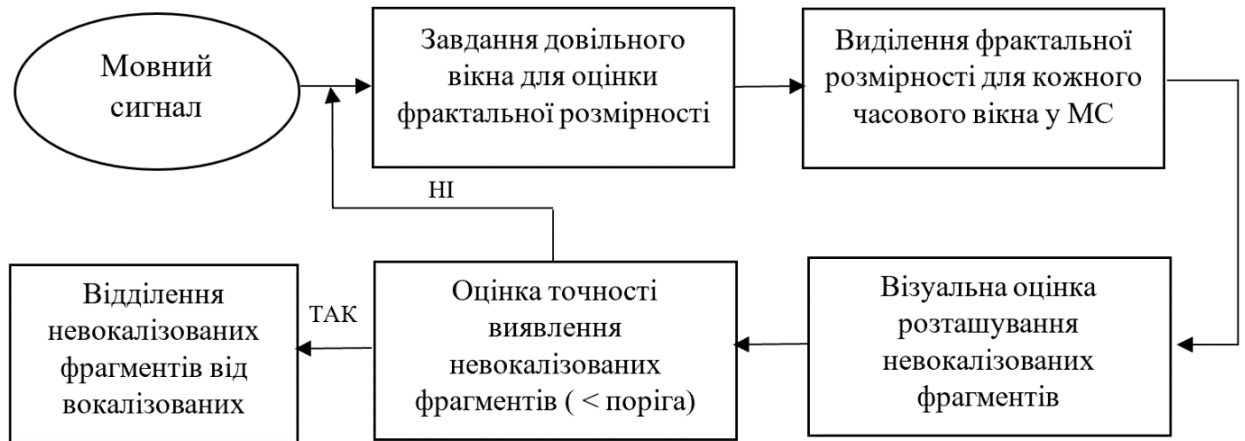


Рис. 4.3. Методологія проведення досліджень сегментації МС

З огляду на залежність фрактальної розмірності від розмірів вікна на наступному етапі задаємося довільним значенням вікна для оцінки фрактальної розмірності. На наступному етапі обчислюємо фрактальну розмірність для кожного вікна в мовному сигналі. Далі відбувається поділ мовного сигналу на паузи і фрагменти мови (візуально). Суб'єктивність методу поділу пауз і фрагментів мови на даному етапі досліджень не повинна надавати істотного впливу на наступні висновки і результати, зважаючи на істотне відмінності фрактальної розмірності пауз і фрагментів мови. Зокрема використання подібного підходу саме по собі наочно відображає зміну фрактальної розмірності при переході з мови на паузи і навпаки (рис. 4.4).

На Рис. 4.4 – величина  $D$  - фрактальна розмірність для фрагмента мовного сигналу для всього часового вікна.  $SDI$  - середньоквадратичне відхилення фрактальної розмірності для часових фрагментів по 20 мс від середнього значення. На даному етапі дослідження вибір оптимального порогу вікна сканування при оцінки фрактальної розмірності здійснюємо на основі достатньої точності ідентифікації пауз для оператора. При необхідності послідовність операцій завдання вікна фрактальної розмірності і обчислення його параметрів може бути повторена. Таким чином, шляхом оцінки точності виявлення пауз на підставі порога здійснюється відділення пауз і мови.

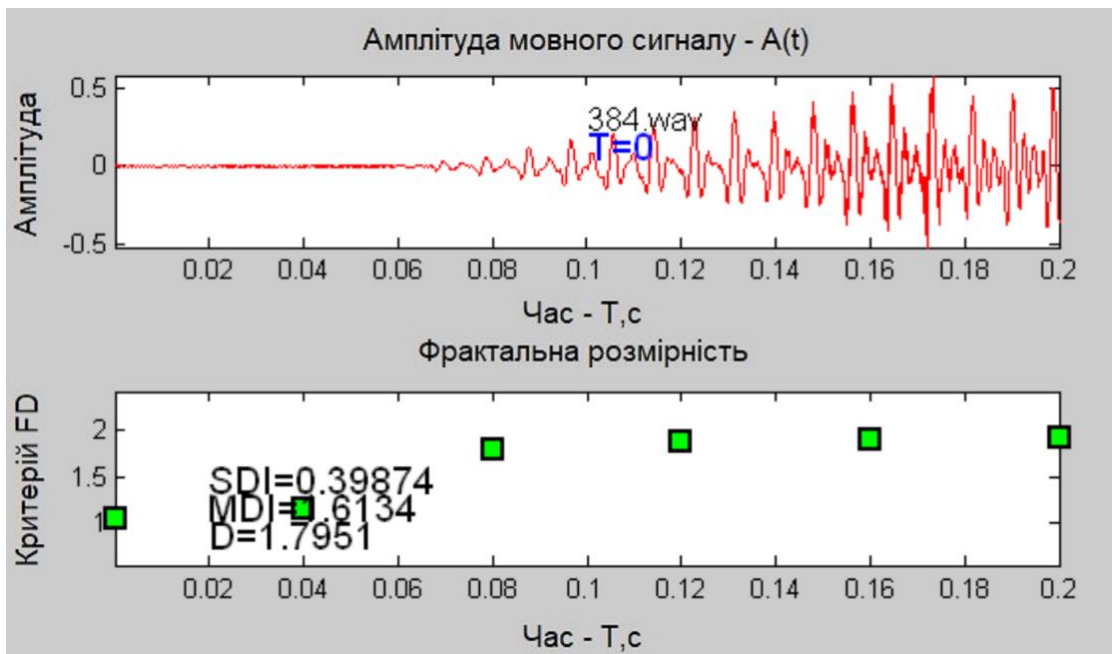


Рис. 4.4. Зміна фрактальної розмірності при переході з паузи на мову

Проаналізувавши 300 різних мовних сигналів, можна помітити значну різницю фрактальної розмірності пауз від фрагментів мови.

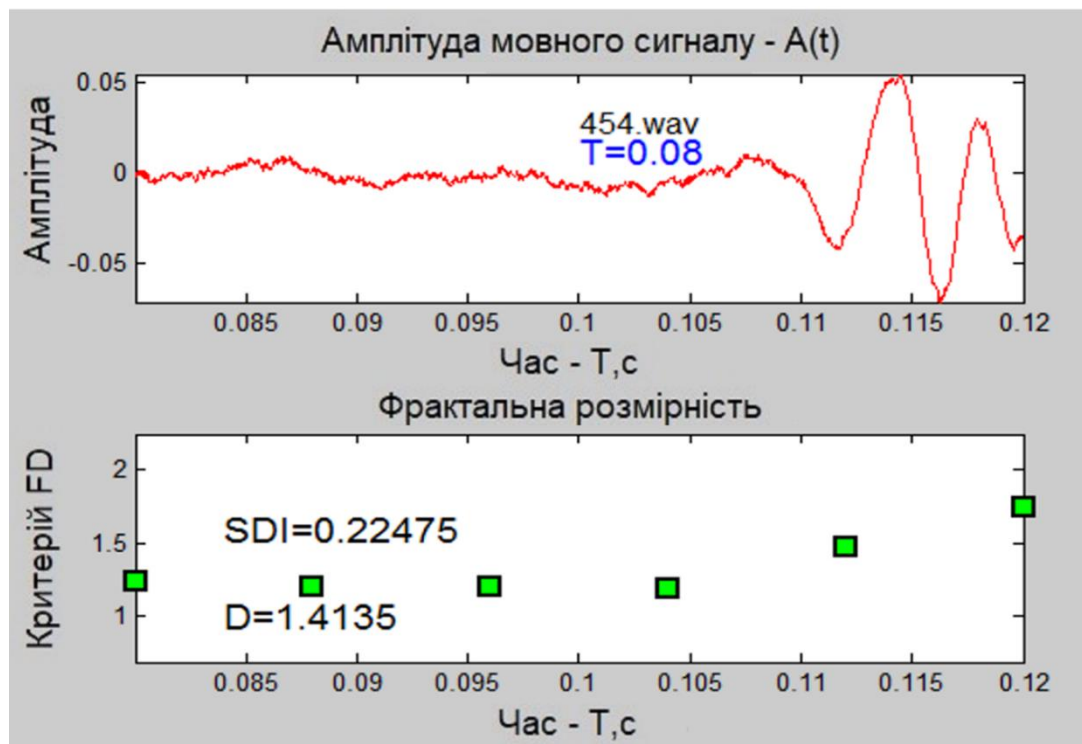


Рис. 4.5. Типовий розподіл фрактальної розмірності для фрагмента мовного сигналу з початковою паузою і починається промовою з 0,11 с.



Фрактальна розмірність для пауз в 99 % перебувала в межах  $1,04 \leq D \leq 1,45$ , а фрактальна розмірність фрагментів мови не спостерігалася менш  $D = 1,55$  (рис. 4.5) для часового вікна розміром 20 мс. Зменшення часового вікна дозволяє виділити у фрагменті мови окремі фонемічні складові (рис. 4.6), однак використання даного факту вимагає додаткового дослідження, що не входить до задач дисертаційної роботи.

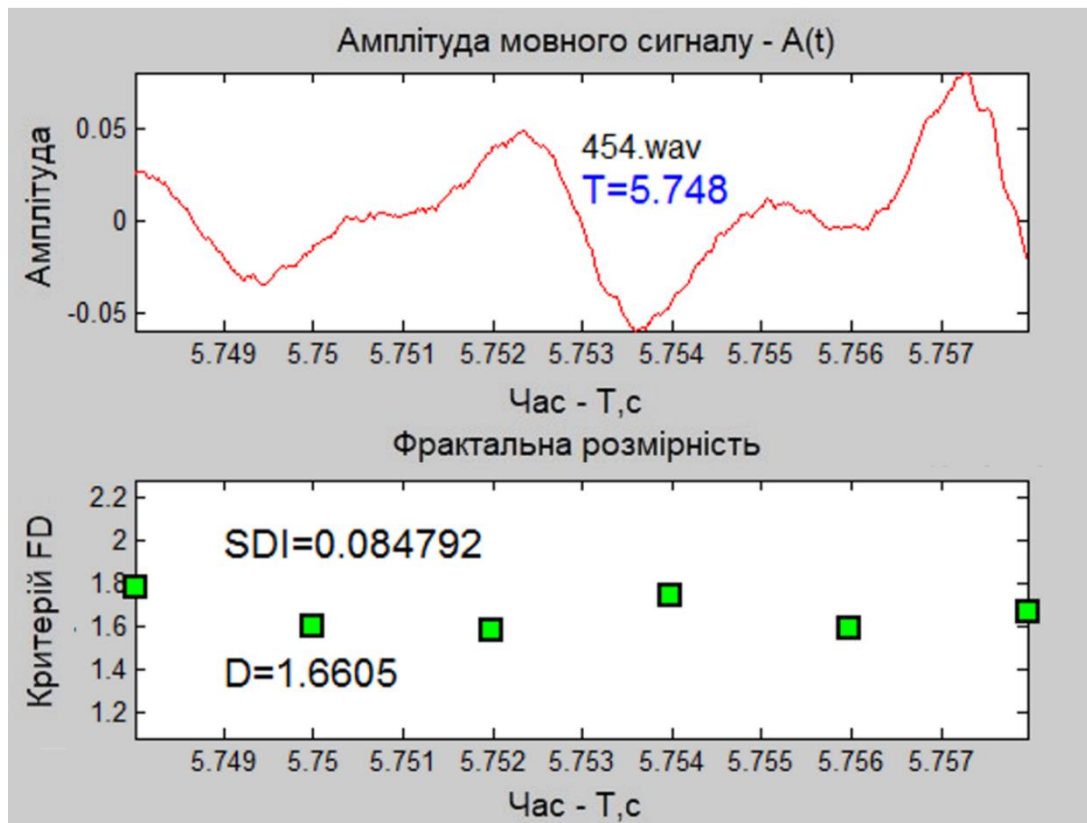


Рис. 4.6. Зміна фрактальної розмірності при переході між фонемічними складовими мови

Важливим фактором є стабільність фрактальної розмірності як фрагментів пауз, так і фрагментів мови на ділянках близько 150-200 мс.

На цьому етапі дослідження вибір оптимального порогу вікна сканування при оцінюванні фрактальної розмірності здійснюється на основі достатньої точності ідентифікації пауз особи. При необхідності послідовність операцій завдання вікна фрактальної розмірності і обчислення його параметрів може бути повторена. Таким чином, шляхом оцінювання точності виявлення пауз на підставі порога здійснюється відділення пауз й мови.

В дослідженні по визначенню раціональних параметрів ознак прийняття рішення використовувались 300 мовних фрагментів, спеціально відібраних для підбора параметрів ознак. Проведене дослідження дозволило визначити раціональні параметри діапазону розмірів мовного вікна в 20-40 мс, а величина середньоквадратичного відхилення обчисленої фрактальної розмірності  $\partial D = 0,01 - 0,08$ .

Для оцінювання фрактальної розмірності діапазон розмірів вікна для визначення фрактальної розмірності становив від 20 до 40 мс. Після численних досліджень і проведених експериментів початкова величина 20 мс була вибрана не випадково. Саме до цієї певної кількості точок графіки носять нелінійний характер. На представлених графіках часове вікно для визначення фрактальної розмірності становило 20 мс, як показали дослідження це часове вікно є оптимальним для сегментації мовного файлу на вокалізовані та невокалізовані фрагменти.

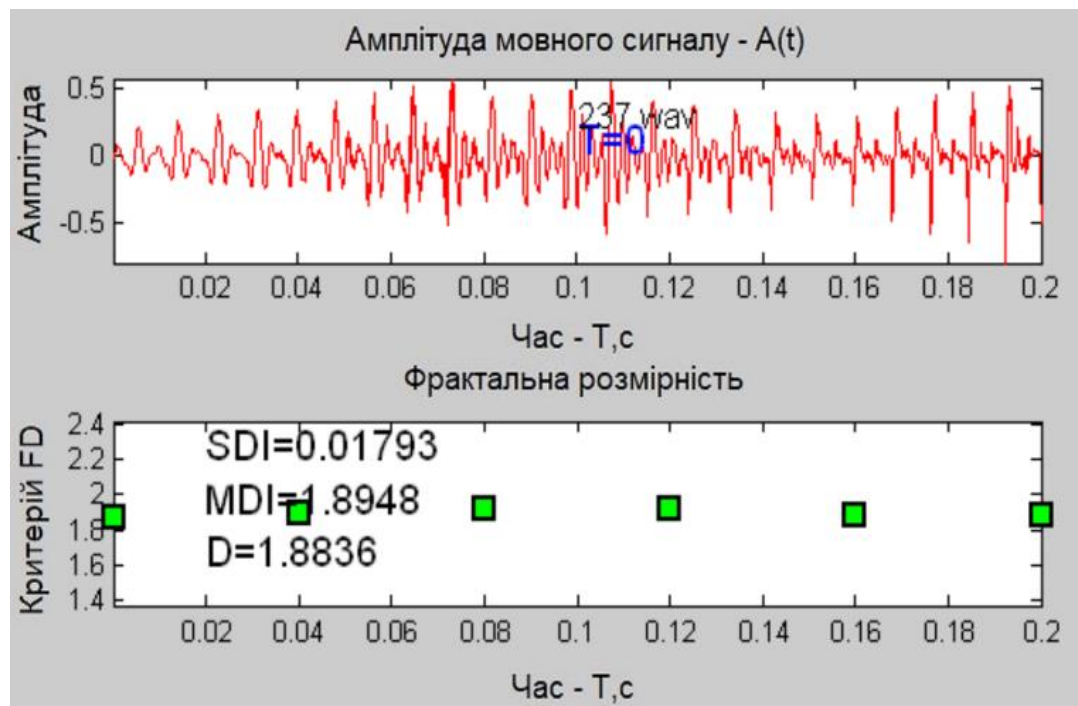


Рис. 4.7. Типовий розподіл фрактальної розмірності для фрагмента мовного файлу без пауз

Середньоквадратичне відхилення фрактальної розмірності на таких фрагментах не перевищує в середньому 0,02. Досить істотні зміни фрактальної

розмірності для фрагментів промови відбуваються на ділянках мовних сигналів, які можливо ідентифікувати з фонемічними складовими мови. На фрагментах пауз також можливі суттєві зміни величини фрактальної розмірності, наприклад, при візуально істотних змінах характеристик пауз.

Для проведення порівняльного аналізу методів сегментації МС були реалізовані наступні методи: покадровий; сингулярний; відхилю від середнього; блокування чорної області; аналізу спектральної форми та запропонований в розділі 3 власний метод. У дослідженні брали участь 50 осіб (чоловіків і жінок з різними мовними даними), які вимовляли однаковий набір слів. В створеному тестовому наборі була виконана автоматична розмітка на вокалізовані фрагменти та паузи. При виконанні порівняння всі методи використовували свої стандартні налаштування, рекомендовані розробниками. Кожен метод виконував сегментацію розміченого МС, при співпадінні розмітки та рішення метода сегментації для кожного поміченого фрагменту цьому методу встановлювалося значення 0 або 1. При визначенні похибок 1 та 2 роду виконувався аналіз та підсумок відповідності значень попередньої розмітки 0 та 1, встановлених за кожним методом сегментації відносно загальної кількості пауз та вокалізованих фрагментів. Результати дослідження представлені в таблиці 4.1.

Таблиця 4.1

## Результати дослідження ефективності методів сегментації МС

Метод	Похибка 1-го роду	Похибка 2-го роду
Покадровий	0,177	0,1
Сингулярний	0,133	0,1
Відхилю від середнього	0,180	0,17
Блокування чорної області	0,156	0,11
Аналізу спектральної форми	0,141	0,1
Запропонований метод	0,108	0,1

Таким чином, за ефективністю розроблений метод сегментації МС краще виконує свої функції, та може бути рекомендований до використання в створюваній системі ідентифікації особи.

### 4.3. Вибір параметрів та дослідження ефективності методу підвищення інформативності частоти основного тону

Як вказувалось у розділі 2, в якості базового вейвлету для виділення самоподібних структур будемо використовувати комплексний вейвлет Морле. Важливою особливістю даного вейвлету є можливість визначення миттєвої частоти при використанні аналітичної форми вейвлету [132]. Вважається, що вейвлет Морле є вейвлетом малих коливань, тобто забезпечує центральну частоту близько 1 Гц, має гарну локалізацію у часі та частотну роздільну здатність. Саме цей набір властивостей робить вейвлет Морле одним з найкращих для аналізу мовного сигналу. Аналітичне представлення вейвлету Морле може бути записано у вигляді

$$\psi(t) = \left[ \exp\left(-\frac{t^2}{2\sigma_t^2}\right) - \sqrt{2} \exp\left(-\left(\frac{\omega_c^2 \sigma_t^2}{4} + \frac{t^2}{\sigma_t^2}\right)\right) \right] \exp(j\omega_c t) \quad (4.1)$$

де  $\omega_c = 2\pi F_c$ ,  $F_c$  – центральна частота вейвлету.

$$\sigma_t = \frac{1}{2\pi\sigma_f} \quad (4.2)$$

де  $\sigma_t$  – стандартне відхилення Гаусіана ( $4\sigma_f$  – ширина вейвлету).

Добуток  $\omega_c \sigma_t$  зв'язує ширину вейвлет Гаусіана з частотою її коливань. Для вейвлету Морле цей добуток має приймати досить великі значення ( $\omega_c \sigma_t \geq 5$ ). Найчастіше використовується діапазон  $5 \leq \omega_c \sigma_t \leq 10$ , при  $0,8 \leq F_c \leq 1$  [133].

Зважаючи що  $\omega_c \sigma_t$  забезпечує зв'язок ширини вейвлету та частоти його коливань, то для визначення раціональних параметрів вейвлету необхідно провести аналіз поєднання параметрів  $(F_c, \omega_c \sigma_t)$ , для більш точної ідентифікації особливостей та структур, притаманних мовному сигналу.

Відомо [134], що у діяльності фонетиків в якості аналізу використовуються голосні звуки, частотний діапазон яких є достатньо широким, але умовно його можна поділити на наступні піддіапазони в залежності від конкретного звуку або їх набору. Зважаючи на це, загальний частотний діапазон дослідження був розбитий на піддіапазони 100-500 Гц, 500-1500 Гц, 1500-4000 Гц, від 4000 Гц. Ці діапазони будуть використані для фіксації шкал налаштувань вейвлет перетворення.

З метою підбору параметрів вейвлету було синтезовано МС із заздалегідь заданими характеристиками. Для синтезування використовувалося відкрите програмне забезпечення SoundGen [135]. Це бібліотека на мові програмування R, що дозволяє синтезувати мову з заданими параметрами. Для генерації формантних частот використовувалась функція `estimateVTL`, а для генерації МС – функція `soundgen`. Було створено сигнал з частотами 450 Гц, 1400 Гц, 2800 Гц та 4300 Гц (рис. 4.8).

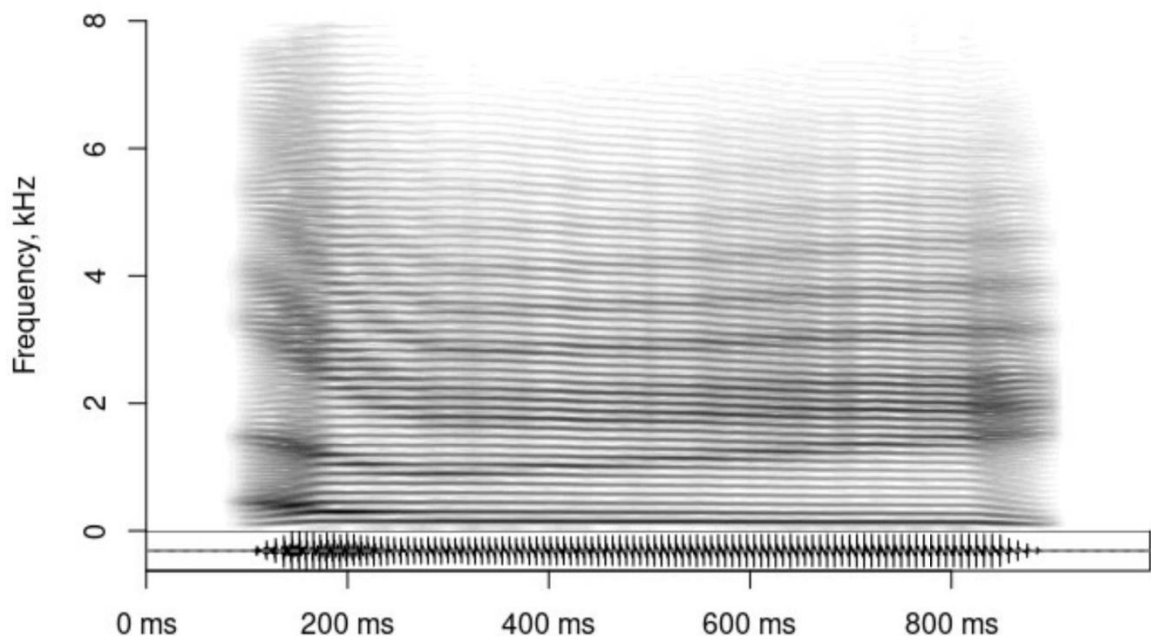


Рис. 4.8. Синтезований МС із заданими частотними параметрами (представлення рисунку подається без змін – як отримано в програмному забезпеченні)

При підборі параметрів змінювалась центральна частота вейвлету з кроком 0,5 Гц в діапазоні 0,8-1 Гц. Для кожного значення центральної частоти

змінювався параметр  $\omega_c \sigma_t$  з кроком 0,5 в діапазоні з 5 до 10 [126]. Для кожної комбінації параметрів обчислювався модуль вейвлет перетворення, знаходилися шкали, які відповідають максимуму модуля коефіцієнтів, та обчислювалися частоти, виводячи фазові коефіцієнти вейвлет перетворення, що відповідають цим шкалам. Отримані параметри налаштувань вейвлет перетворення представлені в таблиці 4.2.

Таблиця 4.2

Параметри налаштування ( $F_c, \omega_c \sigma_t$ ) для обраних частотних діапазонів

Діапазон частот, Гц	$F_c$	$\omega_c \sigma_t$
100-500	0,8	5
500-1500	1	7
1500-4000	0,9	8
>4000	1	9

Згідно методу підвищення інформативності частоти ОТ, що запропоновано в розділі 3, необхідно виконати підбір параметрів та уточнення параметрів методу. Опишемо процес визначення характеристик самоподібних структур на основі мультифрактального спектру:

1. виконуємо розділення МС без пауз на часові фрейми по 20 мс. Після сегментації може статися випадок, що розмір фрейму буде менший за 20 мс, такі часові фрейми відкидаємо з розгляду.

2. обчислюємо вейвлет декомпозицію до 3 рівня.

3. визначаємо положення характерних вейвлет максимумів.

4. представляємо залежність  $M(q, a)$  в координатах  $\log - \log$  (залежність 2.33).

5. розраховуємо параметри мультифрактального спектру  $q$  і  $\tau(q)$  (залежність 2.34).

6. визначаємо сингулярність мультифрактального спектру  $D(h)$  та показник Гельдера  $h$  (залежність 2.35).

Для розрахунку мультифрактального спектру використовувались наступні вхідні параметри:

1. розмір МС не менш 30 с – необхідний для забезпечення достовірності статистичного набору даних.
2. Обраний вейвлет – біортогональний (bior1.5), пов'язано з тим, що цей підхід не може використовувати неортогональні вейвлети, яким є вейвлет Морле, що використовується в інших розрахунках цієї роботи.
3. Використовуються однакові вагові коефіцієнти для кожного рівня вейвлет розкладання.

Розрахунок мультифрактального спектру виконувався з використанням функції `dwtleader` мультифрактального аналізу пакету Matlab v.R2018a, в інших розрахунках цієї роботи використовувався Matlab v.R2012b.

Результати розрахунків характеристик самоподібних структур на основі мультифрактального спектру представлено на рис. 4.9 – 4.11.

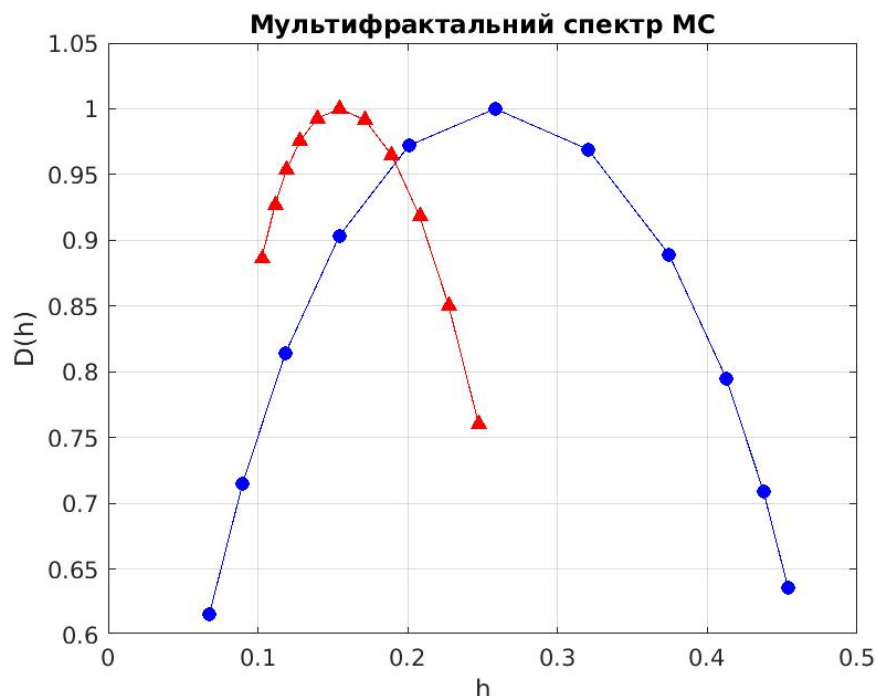


Рис. 4.9. Залежності характеристик самоподібних структур на основі мультифрактального спектру вокалізованого фрагменту (зображено крапкою) та паузи (зображено трикутником)

При перегляді рисунку 4.9. чітко видно значну відмінність МС та паузи, що пов'язано з відомими поняттями монофрактальності та

мультифрактальності [136]. Крім того, відомо, що залежності мультифрактального спектру мають прямий зв'язок з фрактальною розмірністю. Все це підтверджує можливість і ефективність запропонованого в розділі 3 методу сегментації МС на основі фрактальної розмірності.

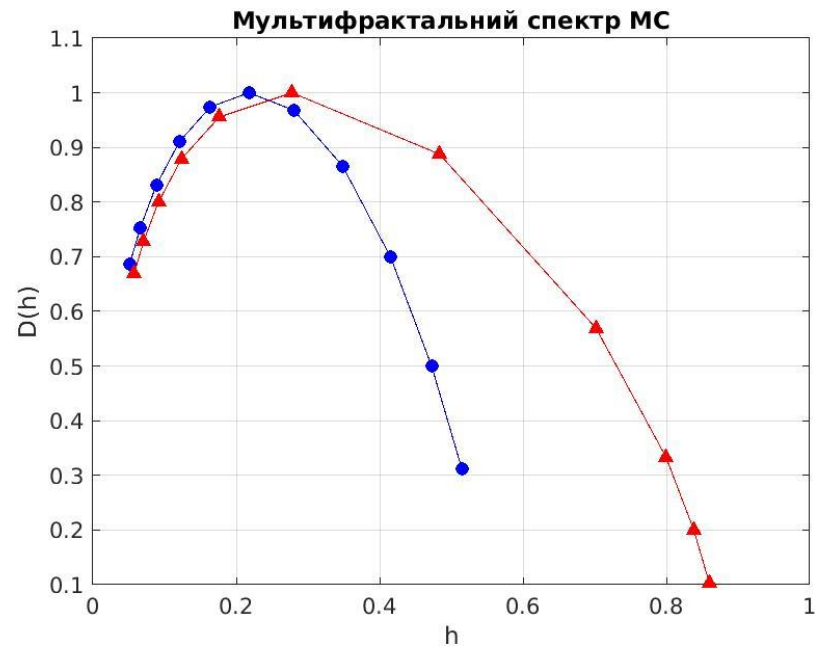


Рис. 4.10. Залежності характеристик самоподібних структур на основі мультифрактального спектру МС різних осіб

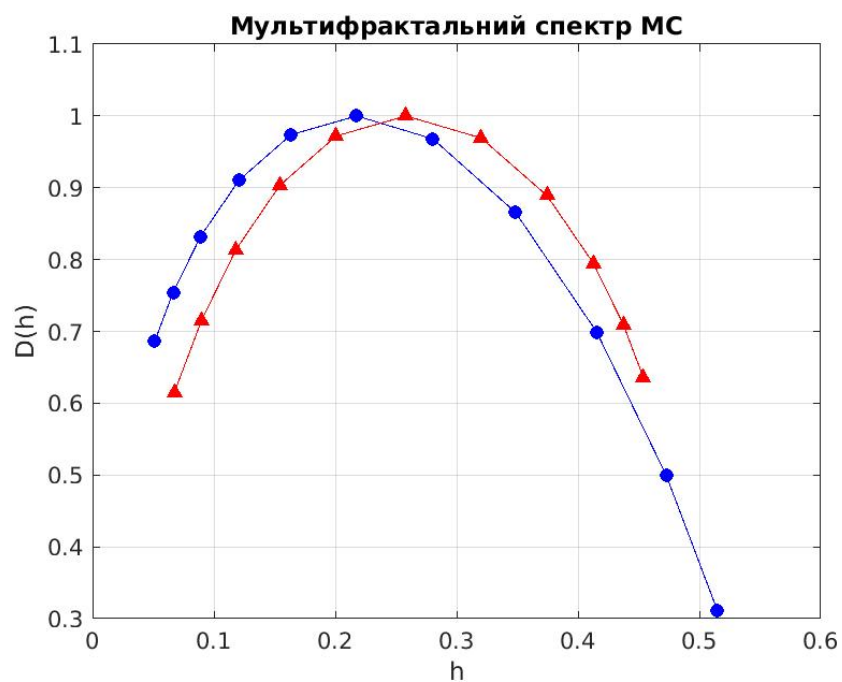


Рис. 4.11. Залежності характеристик самоподібних структур на основі мультифрактального спектру, отримані з різних фрагментів одного МС



Представлені залежності характеристик самоподібних структур на основі мультифрактального спектру для однієї особи і різних осіб відображають достатню варіативність сигналу, що може бути використана для розробки систем мовної ідентифікації особи. В нашому дослідженні вважаємо використання мультифрактального спектру, як додаткову ознаку ідентифікації.

Розглянемо принципи побудови залежностей розподілу частоти ОТ з урахуванням характеристик самоподібних структур при використанні мультифрактального спектру. Як було описано в розділі 3, в процесі визначення частоти ОТ в базі зберігається для кожного фрагменту сигналу два набори даних:

- залежність характеристик самоподібних структур на основі мультифрактального спектру (сингулярність мультифрактального спектру) для фрагменту;

- визначені на основі вейвлету Морле частотні різниці між 7 максимумами спектрограми.

Припустимо, що найбільш ефективні в ідентифікації голосні звуки, і саме по їх патернах найбільш легко порівнюють фонетики, а також беручи до уваги, що в мові близько 30 % голосних в процентному співвідношенні в промові [137], будемо брати 30 % найбільш схожих розподілів мультифрактального спектру. Подоба кривих сингулярності мультифрактального спектра оцінювалася на основі гістограми відносної частоти показника Гельдера (для кожного діапазону по осі  $u$  з кроком 0.1 обчислювалася кількість точок залежності) і ранжувалися за ступенем близькості (близькість оцінювалася на основі абсолютного значення різниці відповідних елементів гістограми відносної частоти показника Гельдера – подібний принцип порівняння використовувався у більшості оцінок близькості двох залежностей в роботі), для кожного фрейму одного МС знаходився фрейм максимально схожий в іншому МС на основі порівняння гістограм. Надалі вибиралося 30 % від загальної кількості фреймів з

найкращими показниками близькості. Передбачається, що відповідність мультифрактальних характеристик буде у подібних одна одній структур. Тільки відібрані таким чином дані наборів різниць частот максимумів для відповідних фреймів беруть участь у формуванні набору даних для побудови гістограми частоти ОТ з урахуванням самоподібних структур для кожного МС, потім об'єднуються в гістограму і нормуються.

Для отримання залежностей розподілу частоти ОТ без урахування характеристик мультифрактального спектру виконується побудова гістограми зі збережених в базі частотних різниць, що на одному рівні з мультифрактальним варіантом відбору фрагментів для визначення частоти ОТ використовується в процесі ідентифікації МС особи.

Введемо для мовної ідентифікації особи за параметрами частоти ОТ наступну міру подоби:

$$\Delta OT(i, j) = \sum_{t=1}^N |p_i^{FOMF}(F_t) - p_j^{FOMF}(F_t)| + \sum_{t=1}^N |p_i^{F0}(F_t) - p_j^{F0}(F_t)| \quad (4.3)$$

де  $p_i^{FOMF}(F)$ ,  $p_i^{F0}(F)$  – значення щільності вірогідності для частоти ОТ з урахуванням та без урахування характеристик мультифрактального спектру, що відповідає особі  $i$ ,  $p_j^{FOMF}(F)$ ,  $p_j^{F0}(F)$  – значення щільності вірогідності для частоти ОТ з урахуванням та без урахування характеристик мультифрактального спектру, що відповідає особі  $j$ .

Тоді порівняння осіб  $i$  і  $j$  для метода визначення частоти ОТ може здійснюватися за допомогою визначального правила при заданому пороговому значенні  $\Delta_{OTlim}$

$$OT(i, j) = \begin{cases} 1, & \text{якщо } \Delta OT(i, j) < \Delta_{OTlim} \\ 0, & \text{інакше} \end{cases} \quad (4.4)$$

При визначенні порогового значення  $\Delta_{OTlim}$  використовувався мовний датасет VTR-TIMIT з вибраними записами 20 жінок і 20 чоловіків, оцінювались параметри компонентів розробленого методу підвищення інформативності частоти ОТ та було обране порогове значення  $\Delta_{OTlim} = 0.516$ .

Наступним етапом перевірки розроблених методів було дослідження ефективності визначення частоти ОТ. Проведені порівняльні випробування запропонованого в роботі методу підвищення інформативності частоти ОТ з методами виділення частоти ОТ, що базуються на піковому методі Доланського, кепстральному методі Нолла, автокореляційному методі Пірогова, методі Pitch Рабінера-Гоулда, фільтровому методі Сапожкова і методі Лобанова Левіна Ковалю (Фірма «Центр мовних технологій»), для чистого сигналу, сигналу з доданим шумом, сигналу, обмеженого смугою телефонного каналу і для гранично обмеженого сигналу з різним рівнем обмеження. Налаштування параметрів методів виконувалося шляхом підбору таким чином, щоб узагальнена похибка виділення частоти ОТ для тестового мовного матеріалу була б мінімальною.

В якості тестового мовного матеріалу біло використано мовні сигнали 20-ти дикторів (10 чоловіків, 10 жінок). Для всіх сигналів виконувалося ручне визначення параметрів частоти ОТ, саме за цією частотою виконувалися операції порівняння методів. Для отримання сигналу з шумом виконувалось додавання білого шуму з необхідним співвідношенням. Співвідношення шуму обчислювалось наступним чином

$$C/Ш = 10 \log \frac{E}{\sigma^2} \quad (4.5)$$

де  $\sigma^2$  – дисперсія шуму,  $E$  – енергія вокалізованого фрагменту МС, яка визначалась як сума квадратів максимумів вейвлет-перетворення. Для отримання сигналу, що моделює роботу телефонного каналу, використовувався фільтр Баттерворта з пропускнуою смугою 300-3500 Гц.

Сумарні результати випробувань за оцінкою узагальненої помилки представлені на рис. 4.12, по десяти чоловікам, десяти жінкам, для 300 мовних фрагментів для кожної особи. Узагальнену помилку обчислювали за нормалізованим коефіцієнтом кореляції з одиничною затримкою для розрахунків за кожним методом з наступним підсумовуванням.

Розроблений метод підвищення інформативності частоти ОТ показав кращі середні результати при визначенні частоти ОТ та забезпечив правильне стеження за траєкторією частоти ОТ протягом усього проголошення, навіть при співвідношенні С/Ш = 5 дБ, в той час як інші методи, виявилися практично непрацездатні. Тому можна вважати, що метод є конкурентоспроможним в порівнянні з іншими розглянутими методами виділення ОТ, які вимагають ручного підстроювання параметрів.

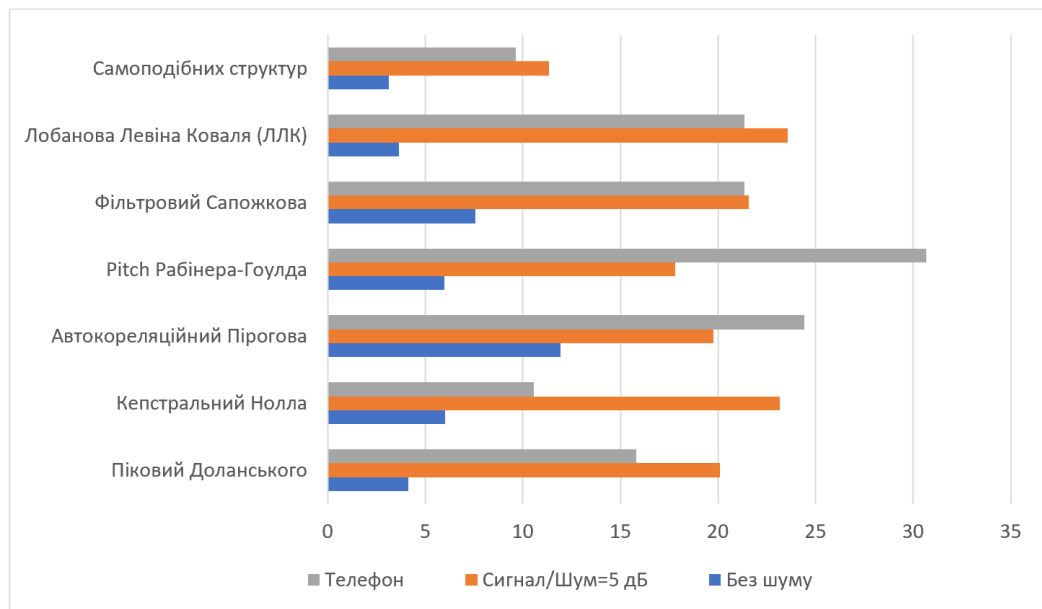


Рис. 4.12. Результати випробувань методів виділення ОТ (%)

На базі власного методу розроблена інформаційна підсистема з автоматичного виділення контуру ОТ. Розроблена підсистема стала одним з компонентів розроблюваної системи ідентифікації особи. Попередні експерименти з використання підсистеми автоматичного виділення частоти ОТ, що працює сумісно з розробленим методом сегментації МС показали ефективність системи в цілому. Суттєво те, що метод сегментації МС та метод підвищення інформативності частоти ОТ використовують ознаки МС, характерні для мови в цілому і не залежні від конкретного типу мови. Випробування показали, що підсистема успішно працювала з проголошенням українською, англійською, німецькою та російською мовами.

#### 4.4. Оцінка ефективності алгоритму визначення формантних частот

Згідно розглянутого в розділі 3 підходу до визначення формантних частот необхідно виконати оцінку та порівняння запропонованого метода з методами, що використовуються у відомих формант-трекерах.

Для дослідження розглядалися формант-трекери PRAAT, SNACK, ASSP та DEEP (опис алгоритмів, які вони використовують, представлено в 3 розділі). Налаштування кожного з них здійснювалось на основі набору параметрів за замовчуванням, що було закладено розробниками цих трекерів. Набір налаштувань для кожного з трекерів представлено у таблиці 4.3.

Основні параметри налаштувань відомих трекерів визначено на основі [100]. Необхідно також відзначити, що кожен з цих відомих трекерів має особливості використання відповідно до статі особи, так PRAAT оптимізовано для мовних сигналів жінок, SNACK та ASSP оптимізовано під мовні сигнали чоловіків. DEEP, як описано у розробника, оптимізувався на наборі даних 67 жінок на 95 чоловіків, тому, можливо, він буде більш точно працювати з мовними сигналами чоловіків.

Таблиця 4.3

Налаштування формант-трекерів, використаних у дослідженні

Параметр	PRAAT	SNACK	ASSP	DEEP	Сплайн-алгоритм
Форманти	5	4	4	4	4
Кількість коефіцієнтів LPC	10	12	18	-	-
Тип вікна	gauss	cos4	blackman	невідомий	gauss
Довжина вікна, мс	25	25	25	невідомий	25
Крок вікна, мс	10	10	10	10	10

Налаштування часових інтервалів власного алгоритму має відповідати розміру подібних інтервалів для інших формант-трекерів для коректного порівняння, тому використовується розмір вікна 25 мс з кроком 10 мс.

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

В якості параметрів порівняння використовувались середньоквадратичне відхилення формант та середня різниця між еталонною розміткою мовних сигналів та результатами формант-трекерів (табл.4.4).

Таблиця 4.4

## Середньоквадратичне відхилення визначення формант (Гц)

Форманта	Стать	PRAAT	SNACK	ASSP	DEEP	Сплайн-алгоритм
F1	жінка	97	104	82	61	78
	чоловік	161	92	71	53	70
F2	жінка	185	197	172	99	89
	чоловік	215	209	109	84	91
F3	жінка	194	183	215	156	167
	чоловік	247	261	113	171	144

Розгляд саме трьох формант, замість чотирьох, пов'язаний з тим, що в датасеті VTR-TIMIT розмічені експертами є тільки три форманти. Крім того, проведений аналіз цього датасету [138] показав середнє відхилення частоти максимуму для 1-3 форманти відповідно 78, 100, 111 Гц, тому отримані значення є достатньо відповідними раніше проведеним дослідженням.

Середня різниця між еталонною розміткою мовних сигналів та результатами формант-трекерів згідно таблиці 4.5, показує стабільні характеристики різниці для сплайн-алгоритму, але з невеликим підвищенням від порядку форманти.

Проведений порівняльний аналіз показує достатньо високу точність визначення формантих частот в порівнянні з існуючими формант-трекерами. Виключення стосується DEEP, але вважаючи не те, що він натренований на саме цьому датасеті, то можна вважати, що метод достатньо добре показує себе в порівнянні з іншими. Поряд з цим, необхідно відзначити простоту реалізації, низьку обчислювальну складність, швидкість та відповідність методу існуючим фізичним процесам.

Таблиця 4.5

Середня різниця між еталонною розміткою мовних сигналів та результатами формант-трекерів (Гц)

Форманта	Стать	PRAAT	SNACK	ASSP	DEEP	Сплайн-алгоритм
F1	жінка	25	19	10	72	13
	чоловік	97	8	12	47	17
F2	жінка	9	2	154	4	11
	чоловік	95	7	9	56	21
F3	жінка	11	37	179	96	32
	чоловік	156	85	20	21	39

#### 4.5. Створення програмної системи мовної ідентифікації особи

Виходячи з запропонованих в розділі 3 методу ідентифікації МС сформульовані вимоги до програмної системи (ПС), що виконує ідентифікацію особи за МС:

1. Програмна система має автоматично виконувати сегментацію мовного сигналу на вокалізовані та невокалізовані фрагменти;
2. Програмна система має виділяти самоподібні структури з вокалізованих фрагментів мовних сигналів;
4. Програмна система має виконувати статистичну обробку параметрів самоподібних структур та підготовку їх до порівняння;

5. Програмна система має визначати міру близькості характерних ознак осіб на основі зіставлення статистичних параметрів самоподібних структур двох МС;

6. Результатом роботи програмної системи є характеристики зіставлення графіків щільності вірогідності статистичних параметрів отриманих самоподібних структур, з відображенням ідентифікаційної помилки.

Розглянемо архітектуру реалізованої системи ідентифікації мови особи на мові UML у вигляді діаграм класів і послідовностей. Діаграма класів відображає статичну структуру системи. Вона складається з опису класів і взаємозв'язків між ними. Діаграма послідовностей відображає динамічні зв'язки в системі, наприклад, послідовність викликів.

На рис. 4.13. представлена діаграма викликів при попередній підготовці до виділення характеристик мови особи. У режимі попереднього розпізнавання мови система завантажується з підготовленим конфігураційним файлом і вхідним сигналом. Розпізнавання буде здійснюватися через диспетчер конфігурації.

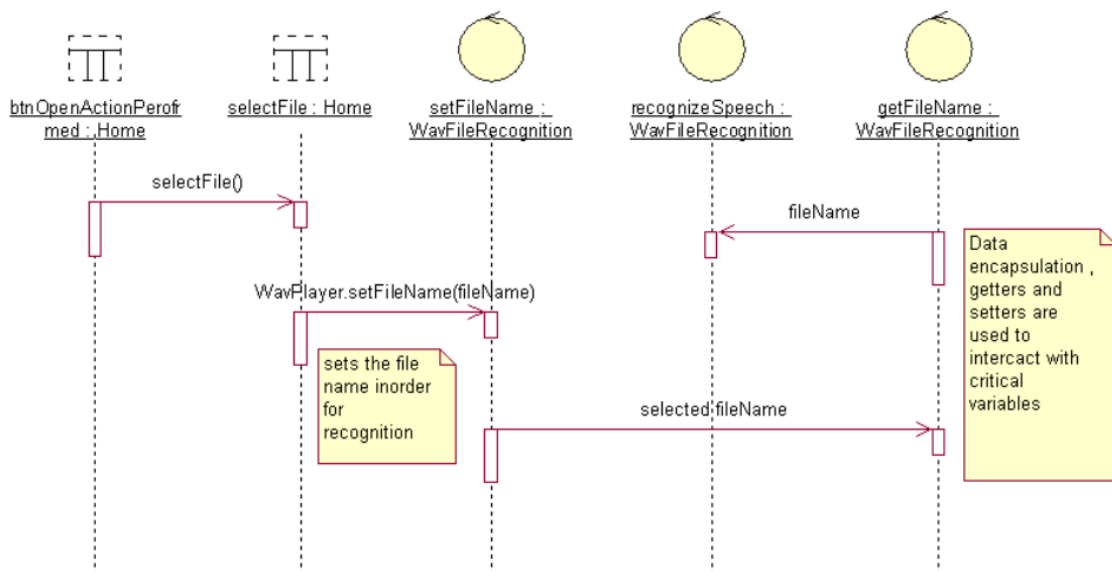


Рис. 4.13. Послідовність викликів попередньої обробки

На рис. 4.14. представлена діаграма виклику при постобробці при розпізнаванні мови особи. На цьому етапі вхідний цифровий сигнал буде



проходити через процес поділу на вокалізовані і невокалізовані фрагменти, розкладання вейвлетом Морле з подальшим статистичним аналізом розподілу вейвлет-максимумів і визначенням частоти основного тону та формантних частот для кожного фрагменту мовного сигналу. Класи `AudioFileDataSource` і `Recognizer` реалізують функції для виконання цих завдань. Результатом послідовності викликів є мітки класу особи і мови, до яких класифікатор відніс вхідний мовний сигнал.

На рис. 4.15. представлена діаграма класів сутностей, які є об'єкними уявленнями даних, якими керує система ідентифікації.

None виконує роль графічного інтерфейсу програмної системи, який безпосередньо взаємодіє з `DBSpeaker` і `WavFileRecognizer`. `DBSpeaker` виконує функції уявлення і опису збережених мовних сигналів осіб. `WavFileRecognizer` призначений для реалізації процесу читання мовного сигналу (зі стріму або з файлу) та ідентифікації диктора. `AudioFileDataSource` реалізує функцію читання мовного сигналу, а `Recognizer` - ідентифікацію МС.

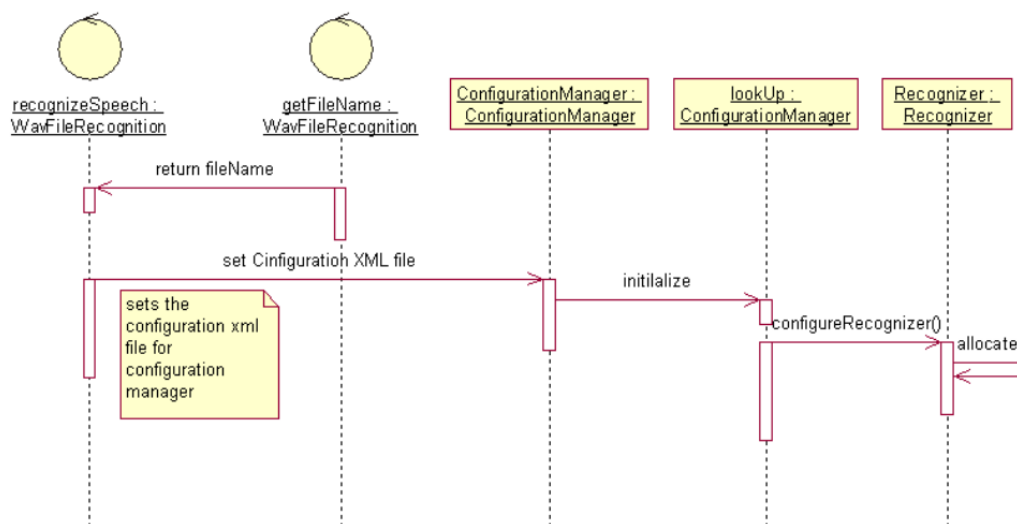


Рис. 4.14. Послідовність викликів процесу розпізнавання мови особи

Абстрактний клас `VoiceFeatures` призначений для зберігання і обчислення ознак вхідного мовного сигналу. Клас складається з масиву об'єктів `VoiceFeatureValue` і методу отримання `ExtractFeatures`, що виконує витяг ознак з отриманого на вхід мовного сигналу. Спадкоємцями класу є

класи, що виконують пофрагментний аналіз: розрахунок мультифрактального спектру; вейвлет-перетворення, з визначенням максимумів; побудова огибальної; визначення попередніх характеристик для обчислення частоти ОТ.

Абстрактний клас `PersonClassifier` призначений для реалізації класифікуючого алгоритму. Клас складається з методів `Train` і `Classify`, а також об'єкта `Parameters`, який містить всі необхідні для роботи класифікатора параметри. Метод `Train` приймає на вхід словник, в якому ключем є мітка класу, а значенням - об'єкт типу `Features`, і повертає об'єкт `Parameters`. Метод `PersonClassify` приймає об'єкт `VoiceFeatureValue` і повертає значення вирішальної функції, а також мітку класу - рішення.

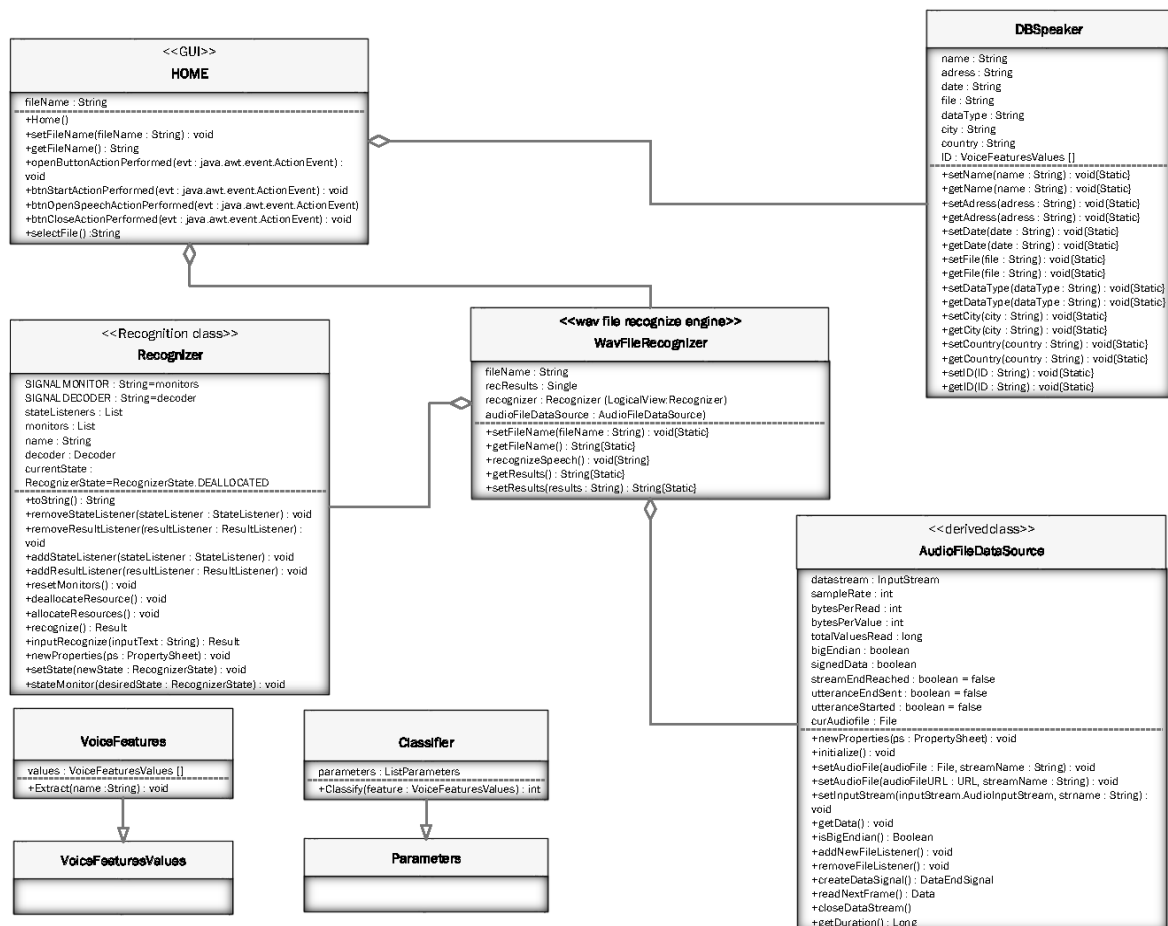


Рис. 4.15. Діаграма класів – сутностей

Клас `SpeechUtils` містить допоміжні методи, необхідні для обчислення ознак і класифікації, такі як, наприклад, обчислення поділу на вокалізовані і невокалізовані фрагменти та очищення від шуму.

Таким чином, запропонована архітектура програмного забезпечення для задач ідентифікації МС особи, що використовує мультифрактальний підхід в описі структури мови. Використання подібної архітектури та використання мультифрактального підходу дозволить в цілому підвищити точність ідентифікації особи.

На основі запропонованої архітектури розроблена програмна система на мові програмування Python з використанням бази даних SQLite.

Програмна система ідентифікації МС особи є пошуковою, оскільки є результатом ранжирування за ступенем близькості окремих параметрів МС.

ПС по цифровим записів інформаційних повідомлень здійснює автоматичний розрахунок параметрів мовних характеристик і подальше ранжування цих характеристик в базі даних осіб.

В ПС використаний метод ранжирування за трьома різними критеріями. До них відносяться:

- обчислення близькості кривих функцій двовимірної щільності вірогідності для частоти ОТ і розташуванню в спектрі трьох частотних діапазонів, що виділяються з промови, зафіксованої на фонограмі;
- обчислення близькості кривих функцій щільності вірогідності для кожного з цих ознак окремо;
- обчислення міри близькості абсолютних максимумів спектрів формант, що виділяються з промови, зафіксованої на фонограмі.

Результатом прийняття рішення розробленої програмі є графічне представлення близькості кривих функцій двовимірної щільності вірогідності для кожної з ознак (рис. 4.16, 4.17).

Перед перевіркою ефективності ідентифікації МС створеною програмною системою необхідно визначитись з параметрами в мірі прийняття рішення (залежності 3.27, 3.28). Визначення параметрів та перевірка ефективності ідентифікації повинна виконуватись на різних наборах даних МС для забезпечення чистоти експерименту. Для налаштування параметрів  $\gamma, \Delta$  використовували датасет VoxCeleb з мовними сигналами 7000 осіб. При

проведенні комп'ютерного експерименту обчислювались значення похибки 1 та 2 роду (рис. 4.18) в залежності від коефіцієнту впливу частоти ОТ (в діапазоні  $\gamma = 0 - 4$ ) та міри подоби.

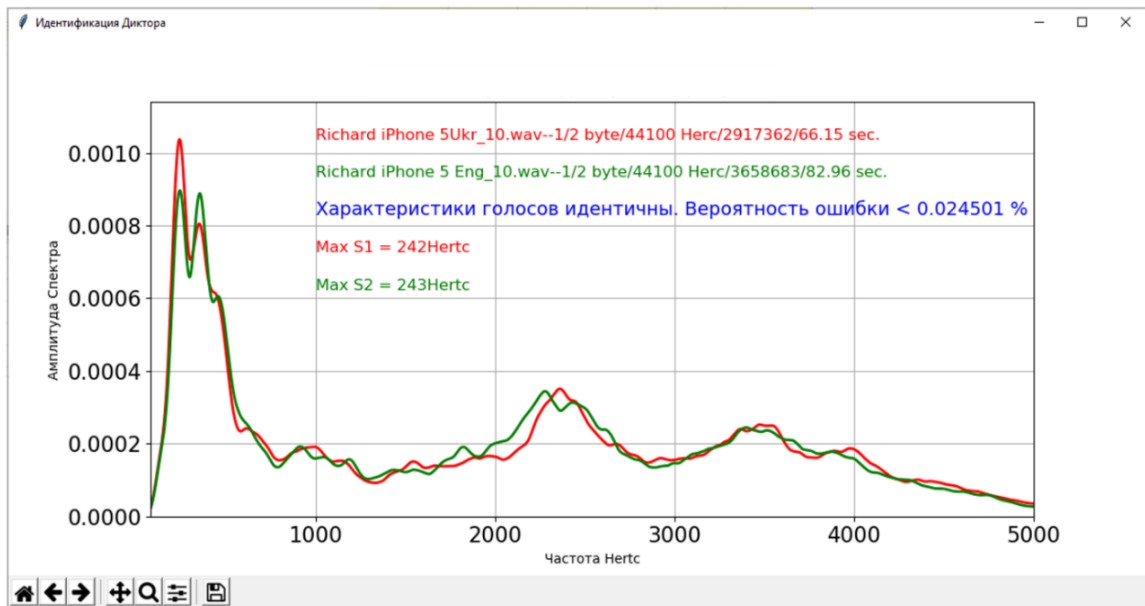


Рис. 4.16. Результуючий графік порівняння мовних сигналів осіб по одній з формант

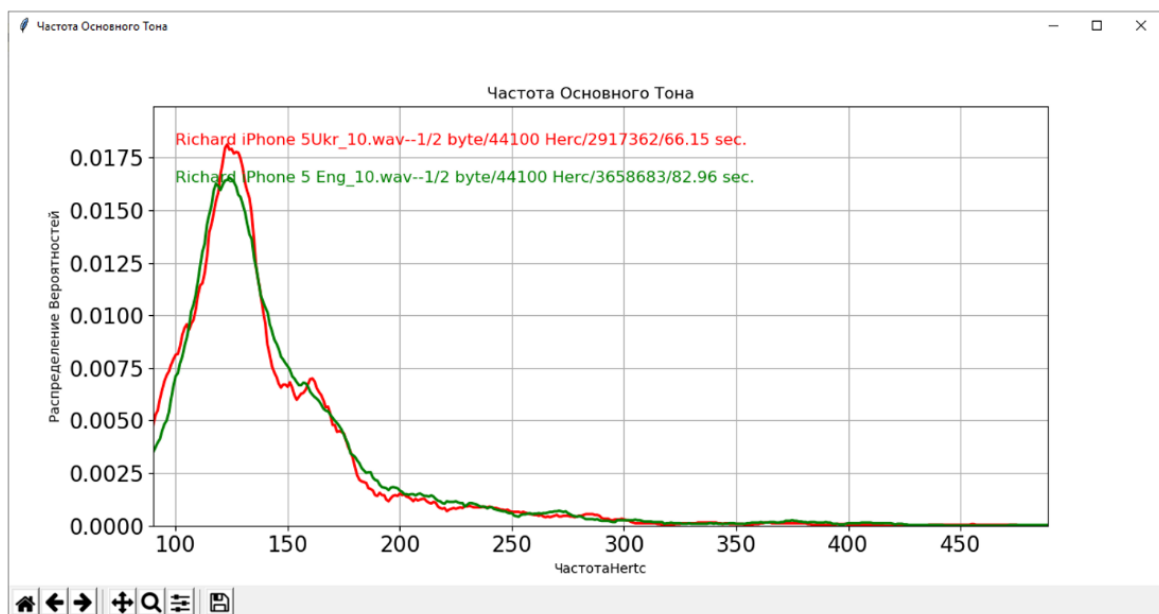


Рис. 4.17. Результуючий графік порівняння мовних сигналів осіб за частотою ОТ

В результаті проведеного дослідження визначено раціональні параметри  $\gamma = 1,34$  та  $\Delta_{lim} = 2,47$  (на перетині кривих похибок 1 та 2 роду з похибкою

3,18 % (рис. 4.19)). Зважаючи на це, для визначення відповідності двох МС одній особі необхідно отримати міру прийняття рішення для цих сигналів, та на основі графіку визначитись з  $\Delta_{lim} < 2,47$ , яке буде визначати похибки 1 та 2 роду для порівняння сигналів.

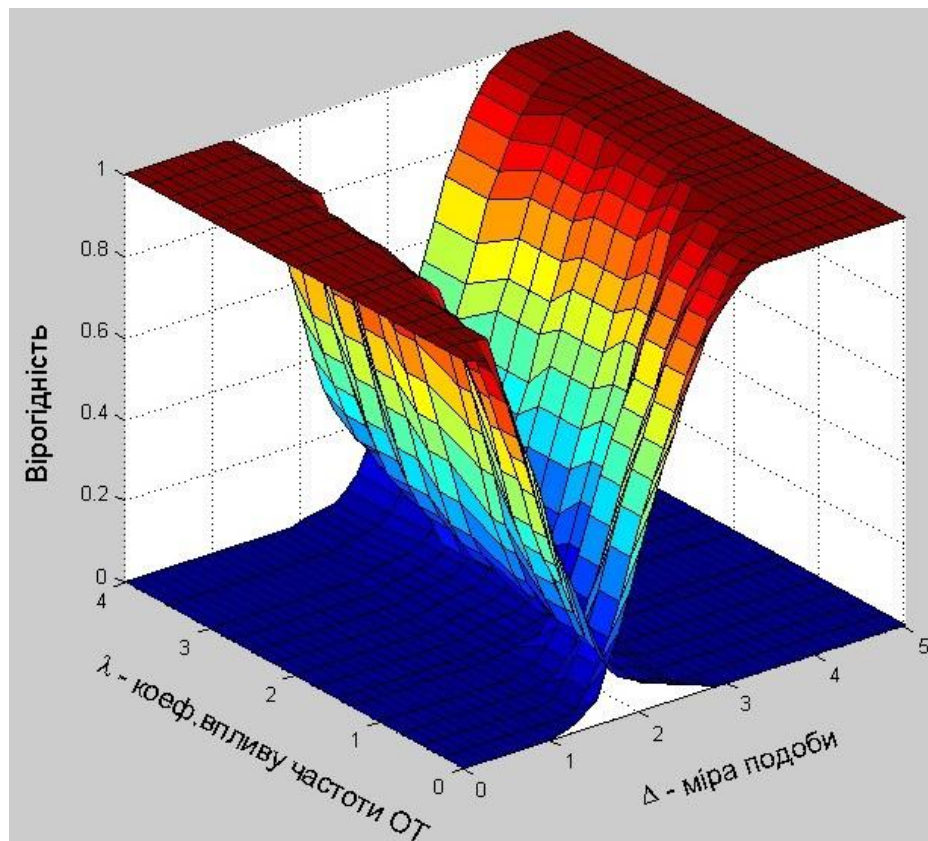
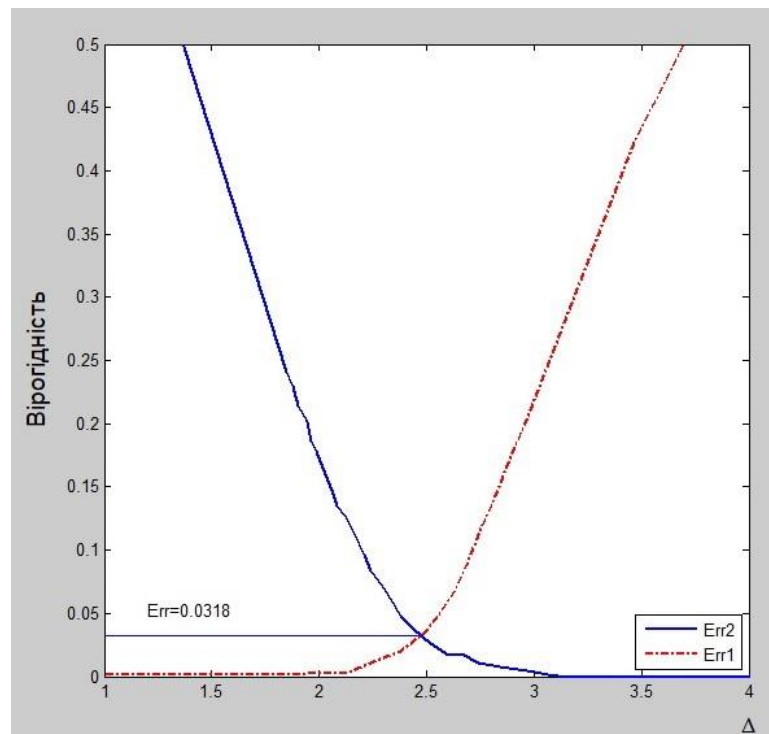


Рис. 4.18. Залежності похибок 1 та 2 роду від коефіцієнта впливу частоти ОТ та міри подоби

Похибка 2 роду є більш важливою для систем ідентифікації, то раціональним є вибір такого значення міри подоби  $\Delta$ , щоб забезпечити менше значення похибки 2 роду відносно похибки 1 роду. Таким чином, якщо вибрати  $\Delta_{lim} = 2,5$ , ідентифікація буде виконуватись з похибкою 1 роду = 4,59 %, 2 роду = 2,12 %.

Наступним етапом була оцінка ефективності розробленої програмної системи ідентифікації особи, що використовує створені методи сегментації МС та підвищення інформативності визначення частоти ОТ. Дослідження виконувалось з набором з 300 МС від 5 осіб для однакових умов запису та шуму МС. Результати представлені в таблиці 4.6.

Рис. 4.19. Зріз поверхонь похибок 1 та 2 роду для  $\gamma = 1,34$ 

Таблиця 4.6

## Характеристики мовної ідентифікації відомих та розробленої системи

Система мовної ідентифікації	Точність мовної ідентифікації		
	при тривалості сигналу більше 96с	при порівнянні сигналів тривалістю 16 та 96 с відповідно	при порівнянні сигналів тривалістю 16с обидва
ТРАЛ-М («Центр Речевых технологий»)	до 91 %	85 %	82 %
ГОЛОС («СТЭЛ-Компьютерные системы»)	до 90,8 %	89 %	89 %
Розроблена система	96,5 %	93,5 %	91,2 %

Згідно за таблицею створена програмна система ідентифікації особи по ефективності перевищує існуючі комерційні аналоги та може бути рекомендованою до використання в експертних установах для ідентифікації особи.

Проведено тестування розробленої програмної системи мовної ідентифікації особи за набором з 300 мовних записів 5 осіб. Система показала

достовірність ідентифікації 96,5 % (ймовірність виникнення похибки другого роду – 2,12 %, ймовірність виникнення похибки першого роду – 4,59 %).

### **Висновки за 4 розділом**

В розділі 4 на базі запропонованих методів та алгоритму розроблено рекомендації щодо проектування системи ідентифікації особи за МС. Досліджено інформативність методів оброблення МС, інформативних ознак ідентифікації особи та методів прийняття рішень.

1. Розглянуто завдання ідентифікації особи та структурна побудова системи ідентифікації особи. В рамках задачі мовної ідентифікації особи описано підхід, що включає ідентифікацію фрагмента МС як виголошеного одною з осіб з розглянутої множини та прийняття рішення відповідності МС особі.

2. В результаті проведеного дослідження на підставі запропонованого методу сегментації МС, а також модифікованої оцінки фрактальної розмірності фрагментів мовних сигналів, встановлені стійкі характеристики підвищення значення модифікованої фрактальної розмірності для фрагментів мовних сигналів, що містять мову. Фрактальна розмірність для пауз в 99 % перебувала в межах  $1,04 \leq D \leq 1,45$ , а фрактальна розмірність фрагментів мови не спостерігалася менш  $D = 1,55$  для часового вікна розміром 20 мс.

3. Проведено вибір параметрів вейвлет-перетворення та визначені параметри мультифрактального спектру, розглянута побудова залежностей розподілу частоти ОТ з урахуванням характеристик самоподібних структур. Введена міра подоби двох МС за критерієм частоти ОТ. Шляхом численного експерименту визначено порогове значення міри подоби  $\Delta_{OTlim} = 0.516$ .

4. Досліджено ефективність визначення частоти ОТ з використанням метода підвищення інформативності частоти ОТ, що показав достатньо високі характеристики в порівнянні з іншими існуючими методами.

5. Виконана оцінка ефективності алгоритму визначення формантих частот в порівнянні з існуючими формант-трекерами. Проведений порівняльний аналіз показав достатню високу точність, а також низьку обчислювальну складність та простоту реалізації алгоритму, запропонованого в роботі.

6. Розроблено рекомендації щодо проектування системи для ідентифікації особи за МС, та побудовано програмну систему, що реалізує функції ідентифікації МС за запропонованими методами та алгоритмом.

7. Виконано налаштування параметрів системи мовної ідентифікації особи шляхом комп'ютерного експерименту на основі обчислення значень похибок 1 та 2 роду. В результаті проведеного дослідження визначено раціональні параметри системи ідентифікації  $\gamma = 1,34$  та  $\Delta_{lim} = 2,47$  (на перетині кривих похибок 1 та 2 роду з похибкою 3,18 %).

8. Проведено порівняльний аналіз ефективності ідентифікації особи в порівнянні з існуючими подібними комерційними програмними продуктами. Система показала достовірність ідентифікації 96,5 % (ймовірність виникнення похибки другого роду – 2,12 %, ймовірність виникнення похибки першого роду – 4,59 %).



## ВИСНОВКИ

У дисертаційній роботі теоретично обґрунтовано й вирішено важливе науково-технічне завдання розробки на основі апарату фрактального та вейвлет аналізу моделі і методу, що здійснюють обробку, визначення індивідуальних характеристик особи та ідентифікацію МС особи. Основні результати дисертаційної роботи:

У дисертаційній роботі представлено вирішення актуальної наукової задачі попередньої обробки МС, сегментації, виділення ідентифікаційних ознак особи у вигляді параметрів частоти ОТ і розпізнавання особи на основі вейвлет та фрактального аналізу. Аналіз отриманих результатів дозволяє зробити наступні висновки.

1. Проведено аналіз математичних моделей формування та сприйняття МС, алгоритмів параметризації й ідентифікації МС, що показав необхідність використання в якості інформаційного критерію ідентифікації – частоти ОТ та частот формант, застосування методів обробки сигналу, що забезпечують вейвлет та фрактальне розкладання МС, а також дозволив сформулювати постановку задач дослідження. Модифіковано модель слухової системи людини, яка на відміну від існуючих, описує механізм оброблення слуховою системою індивідуальних ознак МС, що дозволяє підвищити інформативність характерних ознак для мовної ідентифікації особи.

2. Одержали подальший розвиток методи сегментації МС. Створений метод, що використовує фрактальний аналіз, дозволяє виділяти вокалізовані і невокалізовані фрагменти МС незалежно від мовних даних особи та інтенсивності МС в результаті аналізу динаміки характеристик його фрактальної розмірності.

3. Розроблено алгоритм виділення самоподібних структур в МС та уперше запропоновано інформативну ознаку для мовної ідентифікації особи, яка на відміну від існуючих, використовує значення коефіцієнтів вейвлет-перетворення МС на відрізках, де спостерігаються екстремуми кореляції

частоти ОТ, що дозволяє комплексно враховувати наявність самоподібних структур, які відповідають за індивідуальність МС, та досягти високу точність ідентифікації.

4. Розроблено метод мовної ідентифікації особи, що використовує метод сегментації МС та алгоритм виділення характеристик самоподібних структур. В якості принципу прийняття рішення методу мовної ідентифікації особи є ранжирування за ступенем близькості окремих параметрів МС, що представляють характерні ознаки особи, виділені при реалізації методу.

5. Проведено експериментальні дослідження методів та алгоритму, для мовної ідентифікації особи. Отримані результати порівнювалися з відомими методами, алгоритмами та програмними системами. Дослідження показало, що кращими є методи, запропоновані в дисертаційній роботі та засновані на фрактальному та вейвлет аналізі. Помилки ідентифікації не перевищують 5 %. Запропоновані методи та алгоритм можуть бути використані при розробці систем мовної людино-машинної взаємодії. Практична значимість підтверджена актами впровадження в навчальний процес Національного авіаційного університету та в процесі розробки програмних продуктів у ТОВ «ІСТ ТАУЕР» і «КіберБіонік Систематікс».

6. Створено програмну систему мовної ідентифікації особи. Проведено тестування створеної програмної системи мовної ідентифікації особи. Система показала точність розпізнавання 96,5 %.

**СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ**

1. Сорокин В.Н. Теория речеобразования. Москва : Радио и связь, 1985. 312 с.
2. Маркел Дж., Грей А.Х. Линейное предсказание речи: Пер. с англ. / Под ред. Ю.Н. Прохорова, В.С. Звездина. Москва : Связь, 1980. 308 с.
3. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: Пер. с англ. / Под ред. М.В. Назарова, Ю.Н. Прохорова. Москва : Радио и связь, 1981. 495 с.
4. Сапожков М.А. Речевой сигнал в кибернетике и связи. Москва : Радио и связь. 1963. 452 с.
5. Фант Г. Акустическая теория речеобразования. Москва : Наука. 1964. 283 с.
6. Чистович Л.А., Венцов А.В., Гранстем М.П. и др. Физиология речи. Восприятие речи человеком. В серии «Руководство по физиологии». Ленинград : Наука, 1976. 388 с.
7. Цвикер Э., Фельдкеллер Р. Ухо как приемник информации: Пер. с нем. / Под ред. Б.Г. Белкина. Москва : Связь. 1971. 256 с.
8. Фланаган Дж., Анализ, синтез и восприятие речи: Пер. с англ. / под ред. А.А. Пирогова. Москва : Связь. 1968. 396 с.
9. Лабутин В.К., Молчанов А.П. Модели механизмов слуха. Москва : Энергия. 1973. 200 с.
10. Сиберт У. Преобразование стимула в периферической слуховой системе. – В кн. Распознавание образов: Пер. с англ. / Под ред. Л.И. Титомира. Москва : Мир. 1970. 321 с.
11. Загоруйко Н.Г. Методы распознавания и их применение. Москва : Советское радио. 1972. 206 с.
12. Дворякин С.В. Речевая подпись / Под ред. Заслуженного деятеля науки РФ, д.т.н., проф. А.В. Петракова. Москва : РИО МТУСИ, 2003. 184 с.

13. Seneff S. Pitch and spectral estimation of speech based on an auditory synchrony model. In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP-84, 1984, pages 3621-3624.

14. Seneff S. A joint synchrony/mean-rate model of auditory speech processing. Journal of Phonetics, 16(1): 1988. pages 57-76.

15. Zue V.W., Glass J., Philips M. and Seneff S. Acoustic Segmentation and Phonetic Classification in the SUMMIT System. Proc. IEEE ICASSP. 1989. pages 389-392.

16. Lyon R.F.: A Computational Model of Filtering, Detection and Compression in the Cochlea. Proc. IEEE-ICASSP. 1982. pages 1282-1285.

17. Аграновский А.В., Леднов Д.А. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов. Москва: Радио и связь. 2004. 164 с.

18. Cheveigne A., Kawahara H. Yin. A fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America, 111(4), 2002.

19. Голубинский А.Н. Расчет частоты основного тона речевого сигнала на основе полигармонической математической модели // Вестник Воронежского института МВД России, 2009. № 1. С. 81-90.

20. Makarova V. RUSLANA: a database of russian emotional utterances / V. Makarova, V.A. Petrushin // ICSLP, 2002. P. 2041–2044.

21. Хейдоров И.Э. Классификация эмоционально окрашенной речи с использованием метода опорных векторов / И.Э. Хейдоров, Я. Цзинбинь, [и др.] // Речевые технологии. Вып. 3. СПб., 2008. С. 63–71.

22. Chen Y.T. A study of emotion recognition on mandarin speech and its performance evaluation: Ph. D. dissertation / Y.T. Chen. Tatung, 2008.

23. Morist M.U. Emotional speech synthesis for a radio dj: corpus design and expression modeling: master thesis MTG-UPF dissertation / M.U. Morist. Barcelona, 2010.

24. Siging W. Recognition of human emotion in speech using modulation spectral features and support vector machines: master of science dissertation / W. Siging. Kingston, 2009.

25. Розалиев, В.Л. Моделирование эмоциональных реакций пользователя при речевом взаимодействии с автоматизированной системой // Диссертация на соискание ученой степени кандидата технических наук / В.Л. Розалиев. Волгоград: ВГТУ, 2009.

26. Калюжный, М.В. Система реабилитации слабовидящих на основе настраиваемой сегментарной модели синтезируемой речи // Диссертация на соискание ученой степени кандидата технических наук / М.В. Калюжный. СПб., 2009.

27. Morist M.U. Emotional speech synthesis for a radio dj: corpus design and expression modeling: master thesis MTG-UPF dissertation / M.U. Morist. Barcelona, 2010.

28. Алдошина И.А. Связь акустических параметров с эмоциональной выразительностью речи и пения / И.А. Алдошина // Звукорежиссер. № 2. СПб., 2003

29. Chen Y.T. A study of emotion recognition on mandarin speech and its performance evaluation: Ph. D. dissertation / Y.T. Chen. Tatung, 2008.

30. Siging W. Recognition of human emotion in speech using modulation spectral features and support vector machines: master of science dissertation / W. Siging. Kingston, 2009.

31. Гитлин В.Б. Разработка концепции выделения основного тона речи в информационно-измерительных системах реального времени // Диссертация на соискание ученой степени доктора технических наук. Ижевск : ИГТУ, 2000. 306 с.

32. Dolansky L.O. Instantaneous pitch period indicator // J.Acoust. Sos. Am. - 27. № 11, 1955. P. 67 – 72

33. Гітлін В.Б. Основной тон речевого сигнала. Деп. В ВИНТИ 1998. № 1206-B98. 739 с.

34. Noll A.M. Short-Time spectrum and "Cepstrum" techniques for vocal-pitch detection // J. Acoustic. Soc. Am. N 2. 1955. P.36.
35. Noll A.M. Pitch determination of human speech by harmonic product spectrum, the harmonic sum spectrum and a maximum likelihood estimation // Proc. Symp. Comp. Proc. Commun. N.Y. 1969. 779 -797 pp
36. Пирогов А.А. Пристрій для автоматичного виділення основного тону. - А. с. N 129 739 СРСР. - Пріор, від 08.06.58. - МКІ 21Е 1/20 42е
37. Маркел Дж.Д., Грэй А.Х. Линейное предсказание речи. Москва. СВЯЗЬ 1980. 308 с.
38. Gold B., Rabiner L. Parallel processing techniques for estimating pitch period of speech in the time domain // J. Acoustic Soc. Am., 1969. 46. N 2 (Pt. 2). P. 442 – 448.
39. Рабинер Л.Р., Голд Б. Теория применения цифровой обработки сигналов. Москва. Мир. 1978. 424 с.
40. Методичні рекомендації щодо практичного використання програми SIS при роботі з мовними сигналами / Центр мовних технологій. С. Петербург. - С. Петербург, 1997. С. 279.
41. Hsieh C.-T. Segmentation of continuous speech into phonemic units. IEICS, 1991. pp. 420–424.
42. Al-Manie M.A., Alkanhal M.I., Al-Ghamdi M.M., Mastorakis N., Croitoru A., Balas V., Son E. and Mladenov V. Automatic speech segmentation using the arabic phonetic database. In WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering, no. 10. World Scientific and Engineering Academy and Society. 2009.
43. Makowski R. and Hossa R. Automatic speech signal segmentation based on the innovation adaptive filter. International Journal of Applied Mathematics and Computer Science, vol. 24, no. 2. 2014. pp. 259–270.
44. Cherif A., Bouafif L. and Dabbabi T. Pitch detection and formant analysis of arabic speech processing. Applied Acoustics, vol. 62, no. 10. 2001. pp. 1129–1140.

45. Sharma M. and Mammone R. Subword-based text-dependent speaker verification system with user-selectable passwords. in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, vol. 1. IEEE, 1996, pp. 93–96.

46. Van Hemert J.P. Automatic segmentation of speech. *IEEE Transactions on Signal Processing*, vol. 39. no. 4. 1991. pp. 1008–1012.

47. Lee Y.-S., Papineni K., Roukos S., Emam O. and Hassan H. Language model based arabic word segmentation. in *Proceedings of the 41<sup>st</sup> Annual Meeting of the Association for Computational Linguistics*. 2003. pp 399-406.

48. Nagarajan T., Murthy H.A. and Hegde R.M. Segmentation of speech into syllable-like units. *Energy*. vol. 1. no. 1.5. 2003. p. 2.

49. Kaur E.A. and Singh E.T. Segmentation of continuous punjabi speech signal into syllables. in *Proceedings of the World Congress on Engineering and Computer Science*. vol. 1. Citeseer. 2010. pp. 20–22.

50. Thangarajan R. and Natarajan A. Syllable based continuous speech recognition for tamil. *South Asian language review*. vol. 18. no. 1. 2008. pp. 72–85.

51. Scharenborg O., Wan V. and Ernestus M. Unsupervised speech segmentation: An analysis of the hypothesized phone boundaries. *The Journal of the Acoustical Society of America*. vol. 127. no. 2. 2010. pp. 1084–1095.

52. SaiJayram A., Ramasubramanian V. and Sreenivas T. Robust parameters for automatic segmentation of speech. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference*. vol. 1. IEEE. 2002. pp. I–513.

53. Tanqueiro H.M.M. Utilitzacio' didiomes. *International Journal of Computer Science and Mobile Computing*. Vol.6 Issue.4. 2017. pp. 308-315.

54. Toledano D.T., Go'mez L.A.H. and Grande L.V. Automatic phonetic segmentation. *IEEE transactions on speech and audio processing*. vol. 11. no. 6. 2003. pp. 617–625.

55. Mporas I., Ganchev T., and Fakotakis N. A hybrid architecture for automatic segmentation of speech waveforms. in *2008 IEEE International*

Conference on Acoustics, Speech and Signal Processing. IEEE. 2008. pp. 4457–4460.

56. Go´mez J.A. and Calvo M. Improvements on automatic speech segmentation at the phonetic level. in Iberoamerican Congress on Pattern Recognition. Springer. 2011. pp. 557–564.

57. Siniscalchi S.M., Schwarz P., and Lee C.-H. High-accuracy phone recognition by combining high-performance lattice generation and knowledge based rescoring. in 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP'07. vol. 4. IEEE. 2007. pp. 4–869.

58. Knill K. and Young S. Hidden markov models in speech and language processing. in Corpus-based methods in language and speech processing. Springer. 1997. pp. 27–68.

59. Juang B.-H. and Rabiner L.R. Automatic speech recognition – a brief history of the technology development. Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara. vol. 1. 2005. p. 67.

60. Chowdhury N., Sattar M.A. and Bishwas A.K. Separating words from continuous bangla speech. Global Journal of Computer Science and Technology. vol. 4. 2009. pp. 172–175.

61. Rahman M.M., Khan M.F. and Moni M.A. Speech recognition front-end for segmenting and clustering continuous bangla speech. Daffodil International University Journal of Science and Technology. vol. 5. no. 1. 2010. pp. 67–72.

62. Zio´lko B., Manandhar S., Wilson R.C. and Zio´lko M. Wavelet method of speech segmentation. In Signal Processing Conference, 14th European. IEEE. 2006. pp. 1–5.

63. Ratsameewichai S., Theera-Umpon N., Vilasdechanon J., Uatrongjit S. and Likit-Anurucks K. Thai phoneme segmentation using dual-band energy contour. ITC-CSCC. Proceedings. 2002. pp. 111–113.

64. Tolba M., Nazmy T., Abdelhamid A., and Gadallah M. A novel method for arabic consonant/vowel segmentation using wavelet transform. International



Journal on Intelligent Cooperative Information Systems. IJICIS. vol. 5. no. 1. 2005. pp. 353–364.

65. Kamarauskas J. Automatic segmentation of phonemes using artificial neural networks. *Elektronika ir Elektrotechnika*. vol. 72. no. 8. 2015. pp. 39–42.

66. Suh Y. and Lee Y. Phoneme segmentation of continuous speech using multi-layer perceptron. in *Spoken Language. ICSLP 96. Proceedings. Fourth International Conference*. vol. 3. IEEE. 1996. pp. 1297–1300.

67. Rahman M.M., Khatun F. and Bhuiyan M.A.-A. Blocking black area method for speech segmentation. *Editorial Preface*. vol. 4. no. 2. 2015.

68. Hossain A., Nahid N., Khan N.N., Gomes D.C. and Mugab S.M. Automatic silence/unvoiced/voiced classification of bangla velar phonemes: New approach. *8th ICCIT. Dhaka*. 2005.

69. Kalamani M., Valarmathy S. and Anitha S. Hybrid speech segmentation algorithm for continuous speech recognition. *International Journal on Applications of Information and Communication Engineering*. 2015. pp. 39-46.

70. Bansal P., Pradhan A., Goyal A., Sharma A. and Arora M. Speech synthesis-automatic segmentation. *International Journal of Computer Applications*. vol. 98. no. 4. 2014.

71. Dines J., Sridharan S. and Moody M. Automatic speech segmentation with hmm. In *Proceedings of the 9th Australian Conference on Speech Science and Technology*. 2002.

72. Stolcke A., Ryant N., Mitra V., Yuan J., Wang W. and Liberman M. Highly accurate phonetic segmentation using boundary correction models and system fusion. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2014. pp. 5552–5556.

73. Рыбальский О.В., Соловьев В.И., Чернявский С.С., Журавель В.В., Железняк В.К. Вероятностный подход к принятию экспертных решений по анализу сложных объектов// *Известия Национальной академии наук Беларуси. Серия физико-технических наук*. 2019. №5. с. 359-365.

74. Greenberg S., Singer C., and Mason L. NIST 2020 CTS Speaker Recognition Challenge Evaluation Plan, NIST 2020 CTS Speaker Recognition Challenge, [online], <https://www.nist.gov/itl/iad/mig/nist-2020-cts-speaker-recognition-challenge> (Accessed June 25, 2021)

75. Hansen J.H.L. A Study on Universal Background Model Training in Speaker Verification / Hansen J.H.L., Hasan T. // Transactions on Audio, Speech, and Language Processing. vol. 19 (7). 2011. pp. 1890–1899.

76. Wu. W. A tree-based kernel selection approach to efficient Gaussian mixture model–universal background model based speaker identification / Z. Xiong and T. F. Zheng and Z. Song and F. Soong and W. Wu // Speech Communication. vol. 48 (10). 2006. pp. 1273–1282.

77. Morrison G.S. A comparison of procedures for the calculation of forensic likelihood ratios from acoustic–phonetic data: Multivariate kernel density (MVKD) versus Gaussian mixture model–universal background model (GMM–UBM) / G.S. Morrison // Speech Communication. vol. 53 (2). 2011. pp. 242–256.

78. Al-Ammouri, A. Development of a mathematical model of information serial redundancy of management information systems of the aircraft fire alarm / A. Al-Ammouri, P. Dyachenko, A. Degtiarova // Eastern-European Journal of Enterprise Technologies. – 2017. – Vol. 2, Issue 9 (86). – P. 4–10. doi: 10.15587/1729-4061.2017.96296.

79. Татарникова Т.А. Рыбальский О.В., Соловьев В.И., Командина Т. В. Общие подходы к экспертизе оригинальности и подлинности материалов цифровой и аналоговой звукозаписи / Науковий вісник НАВС. – К.2011, №4, с. 183-191

80. Чистович Л.А., Венцов А.В., Гранстрем М.П. Физиология речи. Восприятие речи человеком. – Л.: Наука, 1976.

81. Huang Xuedong. Spoken language processing: a guide to theory, algorithm and system development. – New Jersey: Prentice Hall PTR, 2001. 910 p.

82. A bio-inspired geometric model for sound reconstruction. [Электронный ресурс]: Journal of Mathematical Neuroscience. – Режим доступа:

<https://mathematical-neuroscience.springeropen.com/track/pdf/10.1186/s13408-020-00099-4.pdf>

83. Moorer J.A. Signal processing aspects of computer music: A survey // In Proceedings of the IEEE. Vol. 65. no. 8. 1977. pp 1108–1137.

84. Mandelbrot B. Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis. Annals of Economic Social Measurement 1, 1972.

85. Mandelbrot B. The Fractal Geometry of Nature. New York: W. H. Freeman. 1982.

86. Mandelbrot B. A Multifractal Walk Down Wall Street. Scientific American. 1999.

87. Mandelbrot B. Robustness of the rescaled range R/S in the measurement of non-cycling long-run statistical dependence // Water Resources Research. V. № 5. 1969.P. 967-988.

88. Maragos P. and Potamianos A. Fractal dimension of speech sounds: Computation and application to automatic speech recognition. J.Acoust. Soc. Am. vol. 3. 1999. pp 150.

89. Mandelbrot B. Fractal Geometry of Nature, Freeman, San Francisco. 1982.

90. Cheung Clifford, Mangan James. Scattering amplitudes and the Navier-Stokes equation. arXiv preprint arXiv:2010.15970, 2020

91. Fouz Infaz. Fluid Mechanics, Mechanical Engineering Dept. University of Oxford. 2001. p. 96.

92. Teager S.M., Teager H.M. Evidence for Nonlinear Sound Production Mechanisms in the Vocal Tract. vol. 55 of D, Kluwer Academic, Boston, 1990.

93. Thomas T.J. A finite element model of fluid flow in the vocal tract. Comput. Speech Lang. vol. 1. 1986. pp. 131-151.

94. Kaiser J.F. Some observations on vocal tract operation from a fluid flow point of view. Biomechanics Acoustics and Phonatory Control. 1983. pp. 358-386.

95. Иванов Б.Н. Мир физической гидродинамики: От проблем турбулентности до физики космоса. Изд.2. Moscow : URSS. 2010. 240 с.

96. Бунія С.К., Ліенхард Дж. Еволюція поверхневих збурень і розбризкування турбулентних рідинних струменів. Журнал інженерії рідин. 116 (4). 1994. С. 721–727.

97. Плєнкін А.В. Разрывы газодинамических функций в методах сквозного счета, их алгоритмическая локализация и классификация // Диссертация на соискание ученой степени кандидата физико-математических наук. Москва. 2013. 125 с.

98. Mallat S. A wavelet tour of signal processing. / Пер. с англ. Москва: Мир. 2005. 832 с.

99. Закс Л. Статистическое оценивание. Москва. Статистика. 1976. – 598 с.

100. Muzy J.F., Bacry E., Arneodo A. Wavelets and multifractal formalism for singular signals: application to turbulence data // Phys. Rev. Lett. 1991. V.67. P.3515–3518.

101. Павлов А.Н., Анищенко В.С. Мультифрактальный анализ сигналов на основе вейвлет преобразования // Изв. Сарат. ун-та Нов. сер. Сер. Физика. 2007. №1. 3-25 с.

102. Соловьєв В. И., Рыбальский О. В., Железняк В. К. Спектральный анализ и современные речевые технологии // Вестник Полоцкого государственного университета. Серия С: Фундаментальные науки. 2014. № 4 – С.144-148

103. О. В. Рыбальский, В. И. Соловьєв, В. В. Журавель, Л. Н. Тимошенко. Статистическая обработка самоподобных структур, выделенных из шумов фонограмм, при определении идентичности аппаратуры цифровой звукозаписи // Електротехнічні та комп'ютерні системи. - 2016. - № 22. - С. 406-410.

104. Железняк В.К. Рыбальский О.В., Журавель В.В., Соловьєв В.И. Обобщенная модель выделения фрактальных структур из цифровых сигналов методом максимумов вейвлет-преобразования // Вестник Полоцкого

государственного университета. Серия С, Фундаментальные науки. 2016 vol 6 №4 с.13-16

105. Wendt H., P. Abry. Multifractality Tests Using Bootstrapped Wavelet Leaders. *IEEE Transactions on Signal Processing*. Vol. 55. No. 10. 2007, pp. 4811–4820.

106. Jaffard S., B. Lashermes, P. Abry. Wavelet Leaders in Multifractal Analysis. *Wavelet Analysis and Applications*. T. Qian, M. I. Vai, and X. Yuesheng, Eds. 2006/ pp. 219–264.

107. Wet F.D., Weber K., Boves L., Cranen B., Bengio S. and Bourlard H. Evaluation of Formant-Like Features for Automatic Speech Recognition. *Journal of the Acoustical Society of America*, vol. 116, 2004. pp. 1781-1791.

108. Mallat S. *A Wavelet Tour of Signal Processing*. Academic Press. 1999. 805 p.

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

110. Messaoud Z.B., Gargouri D., Zribi S. and 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.

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

112. Acero A. Formant Analysis and Synthesis using Hidden Markov Models. in *Proc. of the Eurospeech Conference*. Budapest, 1999.

113. 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. pp. 464-467.

114. 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.

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

116. Boersma P., Weenink D. Praat: doing phonetics by computer [Computer program]. Version 6.0.23. [Электронный ресурс]: Praat: doing phonetics by computer. – Режим доступа: <http://www.praat.org/>

117. Kåre Sjölander. The Snack Sound Toolkit [Computer program]. [Электронный ресурс]: The Snack Sound Toolkit. – Режим доступа: <https://www.speech.kth.se/snack/>.

118. Scheffer M. Advanced Speech Signal Processor (libassp), [Электронный ресурс]: Advanced Speech Signal Processor. – Режим доступа: <http://www.sourceforge.net/projects/libassp>.

119. Keshet J. DeepFormant, [Электронный ресурс]: DeepFormant. – Режим доступа: <https://github.com/MLSpeech>.

120. Gray A. and Wong D. The Burg algorithm for LPC speech analysis/Synthesis. In IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 28. no. 6. pp. 609-615. December 1980, doi: 10.1109/TASSP.1980.1163489.

121. Krishna H., Wang Y. (1993). The Split Levinson Algorithm is Weakly Stable. SIAM Journal on Numerical Analysis, 30(5), 1498-1508. [Электронный ресурс]: SIAM Journal on Numerical Analysis. – Режим доступа: <http://www.jstor.org/stable/2158249>

122. So H.C., Chan, K.W. (2004). Reformulation of Pisarenko Harmonic Decomposition Method for Single-Tone Frequency Estimation. Signal Processing, IEEE Transactions on. 52. 1128 - 1135. 10.1109/TSP.2004.823473.

123. VTR Formants Database [Электронный ресурс]: VTR Formants Database. – Режим доступа: [http://www.ee.ucla.edu/~spapl/VTR Formants.rar](http://www.ee.ucla.edu/~spapl/VTR%20Formants.rar)

124. Nearey, Terrance & Assmann, Peter & Hillenbrand, James. (2002). Evaluation of a strategy for automatic formant tracking. *The Journal of the Acoustical Society of America*. 112. 2323. 10.1121/1.4779372.
125. Schiel, Florian and Zitzelsberger, Thomas. Evaluation of Automatic Formant Trackers. Proceedings of the Eleventh International Conference on Language Resources and Evaluation {LREC} 2018, Miyazaki, Japan.
126. Markel J.E., Gray A.H. (1982). *Linear Prediction of Speech*. New York, NY: Springer.
127. Sun Don X. (1995): Robust estimation of spectral center-of-gravity trajectories using mixture spline models. In *EUROSPEECH-1995*, 749-752.
128. Schalk-Schupp Ingo. (2012). Improved Noise Reduction for Hands-Free Communication in Automobile Environments. 10.13140/2.1.4068.6724.
129. Белозьорова Я.А. Ідентифікація диктора на основі кратномасштабного аналізу / Я.А. Белозьорова // Науковий журнал. Інженерія програмного забезпечення. – 2017. – № 1(29). – С. 15 – 25.
130. Pradip Niyogi. *Introduction to Computational Fluid Dynamics*. Pearson Education India. 2006. p. 76.
131. Березин И.С., Жидков Н.П. *Методы вычислений*. Москва: Физматлит, 1960. Т. 2. — 620 с.
132. Lardiès J., Ta M.N., Berthillier M. Modal parameter estimation from output-only data using the wavelet transform, *Archive of Applied Mechanics*, Vol. 73. 2004. pp. 718-733.
133. Chui C. *An introduction to wavelets*. Academic Press. 1992. p. 278
134. Соловьев В.И., Рыбальский О.В., Железняк В.К. Мультифрактальная структура шепота и распознавание речевых структур. *Вестник Полоцкого государственного университета. Серия С: Фундаментальные науки*. 2014. № 12. с. 16-20.
135. Набір інструментів з відкритим кодом для синтезу голосу, маніпулювання та аналізу [Електронний ресурс]: Набір інструментів з

відкритим кодом для синтезу голосу, маніпулювання та аналізу. – Режим доступу: <http://cogsci.se/soundgen.html>

136. Кириченко Л., Радивилова Т. Оценивание мультифрактальных характеристик стохастических процессов. International Journal "Information Content and Processing". Volume 5. Number 2. 2018. pp.106-143.

137. Іщенко О.С. Голосні звуки української мови залежно від темпу мовлення: монографія – К.:Інститут української мови НАН України, 2012 – 220с.

138. Deng L., Cui X., Pruvencok R., Huang J., Momen S., Chen Y.N., Alwan A. (2006). A Database of Vocal Tract Resonance Trajectories for Research in Speech Processing. In: Proc. of the Int. Conf. on Acoustics, Speech, and Signal Processing



*Список опублікованих праць за темою дисертації*

*Статті у наукових фахових виданнях України:*

1. Соловьев В.И., Брюханова Я.А. Идентификация заданных фрагментов в звуковых файлах / В.И.Соловьев, Брюханова Я.А. // Науковий журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2008. – №9(127) ч.2. – С. 30 – 33. **Внесок автора:** запропоновано структуру аналізу мовного файлу та ідентифікацію заданих фрагментів на основі інформативного компонента та його обмежувача.

2. Рыбальский О.В., Белозеров Е.В., Соловьев В.И., Белозерова Я.А. Методология проверки подлинности сигналограмм выделением самоподобных структур / О.В. Рыбальский, Е.В. Белозеров, В.И. Соловьев, Я.А. Белозерова // Научно-технический журнал. Захист інформації. – 2010. – № 2(47). – С. 35 – 43. **Внесок автора:** запропоновано метод представлення мовного сигналу та ідентифікації стійких самоподібних (мультифрактальних) структур на основі використання вейвлет аналізу.

3. Соловьев В.И., Белозерова Я.А. Использование фрактальной размерности аудиофайлов в задаче сегментации звукового файла / В.И. Соловьев., Я.А. Белозерова // Науковий журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2013. – № 5(194) ч.2. – С. 165 – 168. **Внесок автора:** запропоновано метод сегментації мовного сигналу на основі використання фрактальної розмірності.

4. Соловьев В.И., Белозерова Я.А. Анализ алгоритмов построение системы идентификации диктора / В.И. Соловьев., Я.А. Белозерова // Науковий журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2013. – № 6(195) ч.1. – С. 62 – 66. **Внесок автора:** запропоновано структуру системи тектонезалежної ідентифікації мовного

сигналу, критерії та алгоритми визначення основних параметрів мовного сигналу.

5. Белозерова Я.А. Выделение идентификационных характеристик диктора на основе вейвлет-разложения голосового сигнала / Я.А. Белозерова // Науковий журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2015. – № 1(218). – С. 62 – 66.

*Статті у наукових фахових виданнях України, які входять до міжнародних наукометричних баз даних:*

6. Белозьорова Я.А. Ідентифікація диктора на основі кратномасштабного аналізу / Я.А. Белозьорова // Науковий журнал. Інженерія програмного забезпечення. – 2017. – № 1(29). – С. 15 – 25.

7. Зибін С.В., Белозьорова Я.А. Побудова архітектури програмної системи ідентифікації диктора / С.В., Зибін, Я.А. Белозьорова // Науковий журнал. Інженерія програмного забезпечення. – 2017. – № 2(30). – С. 24 – 31.  
**Внесок автора:** запропоновано загальна структура, особливості побудови програмної системи ідентифікації мовного сигналу та її практична реалізація.

8. Зибін С.В., Белозьорова Я.А. Поліпшення точності ідентифікації мовного сигналу за рахунок уточнювального підходу до підбору характеристик ідентифікації / С.В., Зибін, Я.А. Белозьорова // Науковий журнал. Інженерія програмного забезпечення. – 2017. – № 3(31). – С. 34 – 38.  
**Внесок автора:** визначено критерії втрати точності при ідентифікації мовного сигналу, запропоновано метод ідентифікації мовного сигналу.

*Статті у закордонних виданнях, які входять до міжнародних наукометричних баз даних:*

9. Solovyov Victor, Byelozorova Yana. Multifractal approach in pattern recognition of an announcer's voice / Victor Solovyov, Yana Byelozorova // ТЕКА.

Commission of motorization and energetics in agriculture. – 2014. – Vol. 14, № 2. – pp. 164 – 170. **Внесок автора:** запропоновано мовна модель, що використовує самоподібні (мультифрактальні) структури, доведено можливість компонування фонем як атомарних структур на основі афінних перетворень.

10. Белозьорова Я.А. Виділення самоподібних структур мовних сигналів в задачах ідентифікації диктора / Я.А. Белозьорова // Scientific Journal «ScienceRice». – 2017. – Vol.5. – №2 (34). – pp. 22 – 27.

11. Bielorozova Yana. Analyse and develop the software of automatic search for an anonymous person in the voice database / Yana Bielorozova // International Journal “Information Technologies & Knowledge”. – 2019. – Vol.13. – № 2. – pp. 152 – 164.

12. Zybin Serhii, Bielorozova Yana. Practical approach to speech identification / Serhii Zybin, Yana Bielorozova // International Journal “Information models & analyses”. – 2020. – Vol.9. – № 3. – pp. 224-231. **Внесок автора:** розроблено критерії ідентифікації особи в мовному сигналі та визначено практичні рекомендації виділення ознак особи в мовному сигналі.

13. Zybin Serhii, Bielorozova Yana. Risk-based decision-making system for information processing systems / Serhii Zybin, Yana Bielorozova // International Journal of Modern Education and Computer Science (IJMECS). – 2021. – Vol.13. – № 4. – pp. 34 – 48. **Внесок автора:** розроблено підхід до прийняття рішень в складних системах, що може використовуватися при побудові систем ідентифікації мовної інформації.

*Тези наукових конференцій:*

14. Белозеров Е.В., Белозерова Я.А. Оценка аутентичности цифровых медиа-файлов с использованием особенностей устройств их формирования / Е.В. Белозеров, Я.А. Белозерова // III Міжнародна науково-технічна конференція «Комп'ютерні системи та мережні технології», Київ, 15-17 червня 2010 р. – Київ: Національний авіаційний університет, 2010. – С. 17.

15. Соловьев В.И., Белозерова Я.А. Модель слуховой системы человека для задач идентификации диктора / В.И. Соловьев, Я.А. Белозерова // XVI Международная научно-практическая конференция «Современные информационные и электронные технологии», Одесса, 25-29 мая 2015 г. – Одеса: Одеський національний політехнічний університет, 2017. – С. 46.

16. Белозьорова Я.А. Ідентифікація диктора на основі поліпшеного алгоритму виділення частоти основного тону / Я.А. Белозьорова // Міжнародна науково-технічна конференція «Інтелектуальні технології лінгвістичного аналізу», Київ, 25-25 жовтня 2017 р. – Київ: Національний авіаційний університет, 2017. – С. 9.

17. Белозьорова Я.А. Особливості побудови системи ідентифікації диктора на основі мультифрактального підходу / Я.А. Белозьорова // Всеукраїнська науково-практична конференція «Безпека соціально-економічних процесів в кіберпросторі», Київ, 27 березня 2019 р. – Київ: Київський національний торговельно-економічний університет, 2019. – С. 126-127.

18. Bielozorova Yana. Analyse and develop the software of automatic search for an anonymous person in the voice database / Yana Bielozolorova // International Conference on Software Engineering, Kyiv, June 03 – 06, 2019. – Kyiv: National Aviation University, 2019. – pp. 60 – 64.

19. Bielozorova Yana. Practical approach to building a speaker identification system / Yana Bielozolorova // International Conference on Software Engineering «SoftEngine 2020», Kyiv, April 13 – 15, 2020. – Kyiv: National Aviation University, 2020. – pp. 59 – 61.

20. Serhii Zybin, Yana Bielozorova. Automatic search for involvants of information messages in the voice databases / Serhii Zybin, Yana Bielozorova // International Conference on Software Engineering «SoftEngine 2021», Kyiv, April 12 – 14, 2021. – Kyiv: National Aviation University, 2021. – pp. 47 – 50.

21. Белозьорова Я.А., Зибін С.В. Критерії побудови системи мовної ідентифікації особи / Я.А. Белозьорова, С.В. Зибін // 13 Всеукраїнська науково-

практична конференція «Стан та удосконалення безпеки інформаційно-телекомунікаційних систем «SITS 2021»», Миколаїв – Коблево, 23 – 25 червня 2021 р. С. 74 – 76.

*Патент на винахід (корисну модель)*

22. Пат. 76105 Україна, МПК G 11 b 27/00, 27/36. Спосіб виявлення слідів цифрової обробки цифрових графічних сигналів [Текст] / Я.А. Белозьорова, М.О. Сидоров, К.С. Лебеденко., Є.В. Белозьоров; заявник и патентовласник: Національний авіаційний університет. – заявл. 31.05.2012; опубл. 25.12.2012, Бюл. №24.

**Акти та довідки впровадження результатів дисертаційної роботи**

ЗАТВЕРДЖУЮ  
Проректор з наукової роботи  
Національного авіаційного  
університету

**АКТ ВПРОВАДЖЕННЯ**


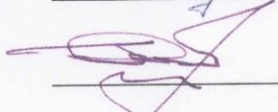
результатів кандидатської дисертаційної роботи Белозьорової Яни Андріївни на тему "Метод застосування вейвлет аналізу в задачах ідентифікації мовної інформації" при виконанні науково-дослідних робіт у Національному авіаційному університеті

Комісія, що є представниками Національного авіаційного університету, в складі декана Факультету кібербезпеки, комп'ютерної та програмної інженерії д.т.н., професора Нестеренко К.С., завідувача кафедри інженерії програмного забезпечення д.т.н., доцента Зибіна С.В., склали акт про те, що результати кандидатської дисертаційної роботи Белозьорової Я.А. впроваджені у НДР кафедри інженерії програмного забезпечення: № 29/09.01.12 «Онтології в інженерії програмного забезпечення», № 58/09.01.02 «Методологія підвищення ефективності процесів життєвого циклу розробки програмного забезпечення у гнучких підходах його розробки» (2019-2022р.).


Назва НД та ДКР, куди впроваджується	Форма впровадження
НДР № 29/09.01.12 «Онтології в інженерії програмного забезпечення»	Метод застосування вейвлет аналізу в задачах ідентифікації мовної інформації та побудова систем ідентифікації мовної інформації
НДР № 58/09.01.02 «Методологія підвищення ефективності процесів життєвого циклу розробки програмного забезпечення у гнучких підходах його розробки»	

Декан ФККП  
д.т.н., професор

Завідувач кафедри ІПЗ  
д.т.н., доцент

 Нестеренко К.С.  
 Зибін С.В.

ПОГОДЖУЮ  
Проректор з навчальної роботи

  
А. Полухін  
" 19 " травня 2021 р.

ЗАТВЕРДЖУЮ  
Проректор з наукової роботи

  
С. Романенко  
" 21 " травня 2021 р.  


**АКТ ВПРОВАДЖЕННЯ  
результатів дисертаційної роботи в  
навчальний процес  
Національного авіаційного університету**

Комісія, що є представниками Національного авіаційного університету, у складі завідувача кафедри інженерії програмного забезпечення д.т.н., доцента Зибіна С.В., д.т.н., професора професора кафедри інженерії програмного забезпечення Трембовецького М.П., д.т.н., доцента професора кафедри інженерії програмного забезпечення Чебанюк О.В. к.т.н., доцента, доцента кафедри інженерії програмного забезпечення Гученко І.В.

склали даний акт у тому, що результати дисертаційної роботи Белозьорової Я.А. на тему «Метод застосування вейвлет аналізу в задачах ідентифікації мовної інформації»,

(назва дисертаційної роботи)

використовуються в навчальному процесі Національного авіаційного університету


Факультету

кібербезпеки, комп'ютерної та програмної інженерії на кафедрі інженерії програмного забезпечення

(ВНЗ, факультет, кафедра)

Що впроваджено (ГЗ, ТТВ, прилад, технологія, програма, алгоритм, модель, методика, рекомендації, регламент, інструкція, винахід, посібник, нормативи та інше)	Форма впровадження (підручник, навчальний посібник, конспект лекцій, методичні розробки, лабораторний практикум, програма курсу, постановка лабораторної роботи, продовження розробки в курсовому та дипломному проекті)	Ефект від впровадження
1. Метод застосування вейвлет аналізу в задачах ідентифікації мовної інформації 2. Методологія сегментації мовного сигналу	1. Використовується в навчальному процесі при проведенні лабораторних робіт згідно з програмою навчальної дисципліни «Мультимедійні, ігрові та розважальні системи» зі спеціальності 121 «Інженерія програмного забезпечення». 2. Використовується в навчальному процесі при проведенні лабораторних робіт згідно з програмою навчальної дисципліни «Інтелектуальні системи» зі спеціальності 121 «Інженерія програмного забезпечення». 3. Використовується в навчальному процесі при проведенні лабораторних робіт згідно з програмою навчальної дисципліни «Людино-машинна взаємодія» зі спеціальності 121 «Інженерія програмного забезпечення».	Підвищення якості навчального процесу шляхом використання запропонованого методу та методології


Голова комісії

  
д.т.н., доцент Зибін С.В.

Члени комісії:

  
д.т.н., проф. Трембовецький М.П.

  
д.т.н., доцент Чебанюк О.В.

  
к.т.н., доцент Гученко І.В.



**ЗАТВЕРДЖУЮ**

**ДИРЕКТОР**

**ТОВ «ІСТ ТАУЕР»**

**О.О. Братко**

**2021 р.**



**АКТ**

**впровадження результатів дисертаційної роботи**

**Белозьорової Яни Андріївни**

Результати дисертаційної роботи Белозьорової Яни Андріївни «Метод застосування вейвлет аналізу в задачах ідентифікації мовної інформації» успішно впроваджені в діяльності ТОВ «ІСТ ТАУЕР». Розроблені автором метод, методологія та алгоритм, застосовані в робочому процесі ТОВ «ІСТ ТАУЕР», показали значну ефективність, дозволили підвищити точність ідентифікації мовної інформації при побудові систем мовної людинно-машинної взаємодії. Запропоновані дисертаційні положення мають практичну цінність і можуть бути використані при побудові програмних проектів, продуктів та окремих модулів систем мовної ідентифікації.



