

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

БЄЛОЗЬОРОВА 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 і систем

АВТОРЕФЕРАТ

дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ – 2021

Дисертацією є рукопис.

Роботу виконано на кафедрі інженерії програмного забезпечення Факультету кібербезпеки, комп'ютерної та програмної інженерії Національного авіаційного університету.

Науковий керівник доктор технічних наук, доцент
ЗИБІН Сергій Вікторович,
професор кафедри інженерії
програмного забезпечення
Національного авіаційного університету

Офіційні опоненти: доктор технічних наук, професор
ЯКОВИНА Віталій Степанович,
професор кафедри систем штучного інтелекту
Національного університету
"Львівська політехніка"

доктор технічних наук, професор
АЛЬ-АММОРИ Алі Нурддинович,
професор, завідувач кафедри
інформаційно-аналітичної діяльності
та інформаційної безпеки
Національного транспортного університету

Захист відбудеться «26» серпня 2021 року о 12:00 на засіданні спеціалізованої вченої ради Д 26.062.19 Національного авіаційного університету за адресою: 03058, м. Київ, просп. Любомира Гузара, 1, ауд. 6.205.

З дисертацією можна ознайомитися в бібліотеці Національного авіаційного університету за адресою: 03058, м. Київ, просп. Любомира Гузара, 1.

Автореферат розіслано «23» липня 2021 р.

Вчений секретар
спеціалізованої вченої ради,
доктор технічних наук, доцент



Р.С. Одарченко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність. 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 ідентифікації мовних сигналів.

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 мовного сигналу (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 ідентифікації 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 ідентифікації 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 значно розширює алгоритмічну і методичну базу для створення інформаційних технологій обробки і аналізу МС.

Дана дисертаційна робота присвячена розробці методу, спрямованого на підвищення точності ідентифікації МС в умовах шуму, а також пошуку параметрів, які описують акустичні характеристики мови особи на основі вейвлет аналізу та інваріантних характеристик до інтенсивності сигналу.

Зв'язок роботи з науковими програмами, планами, темами. Основні дослідження за темою дисертації було виконано на кафедрі інженерії програмного забезпечення Факультету кібербезпеки, комп'ютерної та програмної інженерії Національного авіаційного університету в рамках фундаментальних тем НДР № 29/09.01.02 «Онтології у інженерії програмного забезпечення» (2016–2018 р.); № 58/09.01.02 «Методологія підвищення ефективності процесів життєвого циклу розробки програмного забезпечення у гнучких підходах його розробки» (2019–2022 р.).

Мета і 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 наступні методи: 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 отриманих результатів. У ході виконання дисертаційного дослідження були отримані наступні основні результати, що відображають наукову новизну роботи:

1. 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 у 96,5 %, при побудові системи ідентифікації МС.

2. Уперше 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.

3. В рамках фрактального аналізу 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.

4. 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.

На основі вище викладеної новизни розроблено новий підхід до обробки МС, що базується на запропонованих методах та реалізує функції попередньої обробки, сегментації МС, класифікації самоподібних структур і ідентифікації особи.

Обґрунтованість і вірогідність отриманих у роботі наукових результатів, висновків і рекомендацій підтверджується результатами дисертаційного дослідження: коректним використанням методів фрактального та вейвлет аналізу, теорії цифрової обробки сигналів, обробки експериментальних даних та аналізу результатів досліджень, а також залученням широкої наукової громадськості до апробації результатів роботи на наукових конференціях і їх публікації у наукових фахових виданнях.

Практичне значення отриманих результатів:

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.

Реалізація та впровадження результатів роботи. Отримані результати впроваджено на кафедрі інженерії програмного забезпечення Факультету кібербезпеки, комп'ютерної та програмної інженерії Національного авіаційного університету у навчальний процес при проведенні лабораторних робіт згідно з програмою навчальних дисциплін «Інтелектуальні системи», «Людино-машинна взаємодія» та «Мультимедійні, ігрові та розважальні системи» із спеціальності 121 «Інженерія програмного забезпечення» та у виробничий процес ТОВ «ІСТ ТАУЕР» і ТОВ «КіберБіонік Систематікс».

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 результатів дисертації. Основні положення та результати дисертаційної роботи доповідалися й обговорювалися 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).

Публікації результатів. Основні положення та результати дисертації 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, 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 й обсяг дисертації. Дисертаційна робота складається із вступу, чотирьох розділів, висновків, списку використаних джерел, доповнень. Загальний обсяг – 153 сторінки, включаючи 57 рисунків, 7 таблиць, 4 доповнення, список використаних джерел – 138 одиниць.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність дисертаційної роботи, сформульовано мету, задачі, а також методи дослідження, подано відомості про зв'язки обраного напрямку дослідження із планами організації, у якій виконана робота. Викладено основні наукові результати роботи і відзначено її практичне значення.

У **першому** розділі проведено аналіз вітчизняної та зарубіжної літератури з питань, пов'язаних з темою дисертації. Створювані системи ідентифікації особи пройшли етапи від простого порівняння МС на слух до багатокомпонентних моделей, що використовують сучасні математичні методи представлення МС, ймовірного розкладання, передбачення та інше. Проведено аналіз методів параметризації МС, методів попередньої обробки МС і його сегментації, основні

підходи до ідентифікації МС, які застосовуються в існуючих системах розпізнавання як складових частинах інтелектуальних систем.

Виділено, що основними причинами, які перешкоджають використанню мовних технологій у практиці, є низька якість ідентифікації особи та відсутність надійних акустико-фонетичних ознак. Розглянуто сучасні методи фоноскопічної ідентифікації людини та ряд завдань, які потребують визначення. В якості основного інформаційного критерію ідентифікації особи визначено частоту ОТ та формантні частоти, як найбільш рекомендовані експертами в сфері мовних технологій. Проведений аналіз дозволив сформулювати задачі досліджень.

У **другому** розділі дисертації виконано основні теоретичні розробки, а саме, модифіковану модель слухової системи людини, яка дозволила сформулювати принцип оброблення інформативних ознак для ідентифікації МС особи; узагальнену послідовність математичних перетворень інформаційного каналу мовного спілкування, на основі якої виділено ознаки для сегментації МС на вокалізовані фрагменти та паузи, та принципи визначення характерних ознак МС, що можуть бути використані при ідентифікації.

Враховуючи притаманну слуховій системі людини властивість ідентифікувати особу за мовою, в якості моделі потенційної системи для ідентифікації МС особи розроблено модель слухової системи пересічної людини.

Структурну схему узагальненої моделі слухової системи людини зображено на рисунку 1.

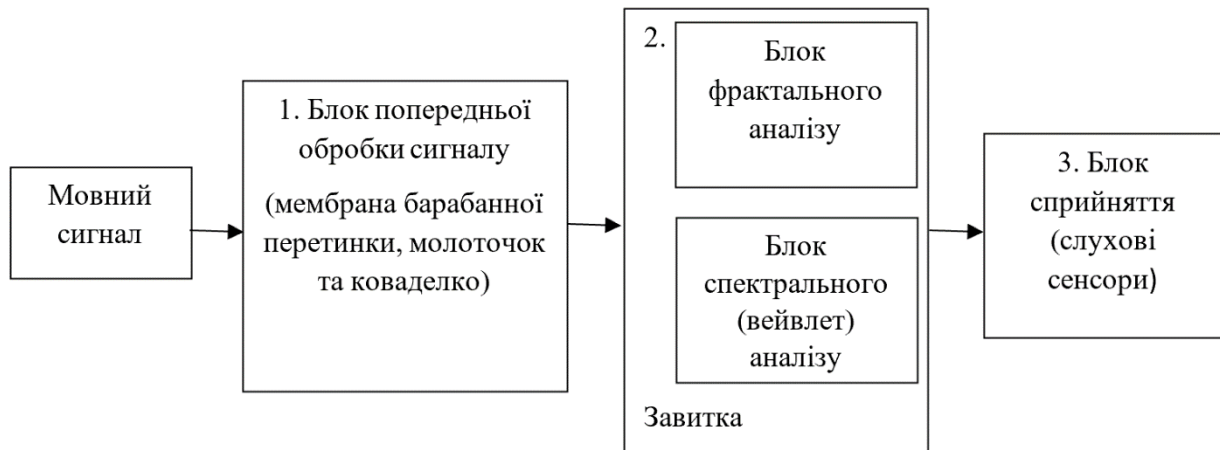


Рис. 1. Структурна схема узагальненої моделі слухової системи

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 представляє собою фрактальну самоподібну структуру, яка виконує сегментацію сигналу, на вокалізовані фрагменти та паузи. Блок спектрального аналізу реалізує пошук самоподібних структур з метою ущільнення інформації в сигналі та визначення основних мовних особливостей особи.

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

Розглянемо математичне представлення основних блоків, що виконують обробку МС. Визначимо принципи необхідності використання фрактального аналізу при дослідженні МС.

В роботі було показано, що турбулентні потоки повітря мовної системи людини формують МС, створюючи вкладення в його структуру. Згідно з теоремою вкладення вектор

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

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 $X(n) \rightarrow X(n + 1)$, можна 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 є досить великою.

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

$$I(T) = \sum_{n=1}^{N-T} P(s(n), s(n + T)) \cdot \log_2 \left[\frac{P(s(n), s(n + T))}{P(s(n)) \cdot P(s(n + T))} \right] \quad (2)$$

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

$$T_D = \min \left\{ \arg \min_{T \geq s_0} I(T) \right\} \quad (3)$$

На основі виконаного дослідження та представлення процесів, що відбуваються при формування звуків мовлення встановлено, що одним з важливих характеристик МС 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 характерні ознаки, притаманні 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 спектральний (вейвлет) аналіз. Зважаючи на це, в створюваній моделі слухової системи людини присутній блок саме такого аналізу.

Визначимо принципи використання вейвлет аналізу в задачі ідентифікації МС особи. Зважаючи на розглянуту в розділі 1 побудову мовної системи людини запропоновано подання сигналу у вигляді складових з різною структурою:

$$f(t) = x_1 f_1(t) + x_2 f_2(t) + \dots + x_s f_s(t) \quad (4)$$

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:

$$\sum_i g_i f_j(t_i) f_k(t_i) = \delta_{jk} \quad (5)$$

де $g_i = 1/\sigma_i^2$.

Зважаючи на те, що функції f_i в (4) 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:

$$f_i(t) = \sum_n c_{in} \varphi_{in}(t) \quad (6)$$

де $f_i \in L^2(R)$, φ_{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:

$$f_M = \sum_{m \in I_M} \langle f, \varphi_m \rangle \varphi_m \quad (7)$$

де I_M - множина індексів, яка визначається властивостями функції f .

Математична конструкція (4) з урахуванням введених властивостей (5) - (7) 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 .

Запропонований підхід до побудови моделі часового ряду, яким представляється МС зі складною структурою, і використання вейвлет-перетворення, в якості методу ідентифікації його структурних складових надає широкі можливості в вирішенні задачі аналізу мовних сигналів з метою виділення самоподібних структур.

Проведені дослідження дозволили запропонувати модель слухової системи людини, описати елементи моделі, визначити в якості ідентифікаційних ознак особи самоподібні структури в МС та математичні аспекти алгоритму виділення характеристик самоподібних структур в МС, що побудовані на принципах фрактального та вейвлет аналізу.

У **третьому розділі** розроблено принципи використання фрактального та вейвлет аналізу в задачі ідентифікації особи, розроблено метод сегментації МС на вокалізовані фрагменти та паузи, а також алгоритм виділення характеристик самоподібних структур.

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.

Проведений в роботі аналіз дозволив запропонувати наступний метод обчислення фрактальної розмірності, що дозволив виконувати сегментацію МС на вокалізовані фрагменти та паузи. Кожне часове вікно МС представимо у вигляді набору прямокутників розміру $a \times b$, що покривають графічне представлення МС. Нехай масштаб представлення p :

$$p = k \cdot b \quad (8)$$

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

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

$$D = 2 - \lim[\ln(N(p))/\ln(p)] \quad (9)$$

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.

Огляд залежності (9) показує, що велике значення має масштаб представлення. Оцінимо зміну масштабу для конкретної реалізації часового вікна МС. Для цього обчислимо залежність логарифму числа прямокутників та логарифму масштабу представлення $\ln(N(p)) = f(\ln(p))$. Для отримання нелінійного характеру залежностей апроксимуємо 10 точок графіку:

$$F(p) = c \ln(p) + c_0 \quad (10)$$

де c і c_0 – коефіцієнти апроксимації.

На основі виконаної апроксимації фрактальна розмірність D обчислюється наступним чином:

$$D = 2 - abs(c) \quad (11)$$

На рис. 2 представлена величина D – фрактальної розмірності для фрагмента мовного сигналу для всього часового вікна. Представлена залежність фрактальної розмірності МС відображає достатні можливості для реалізації методу сегментації МС на вокалізовані фрагменти та паузи.

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

$$C_{mor}(t_i, T_k, F_b, F_c) = (\pi F_b)^{0.5} \exp(2j\pi F_c t_i) \exp(-(t_i - T_k)^2 / F_b) \quad (12)$$

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.

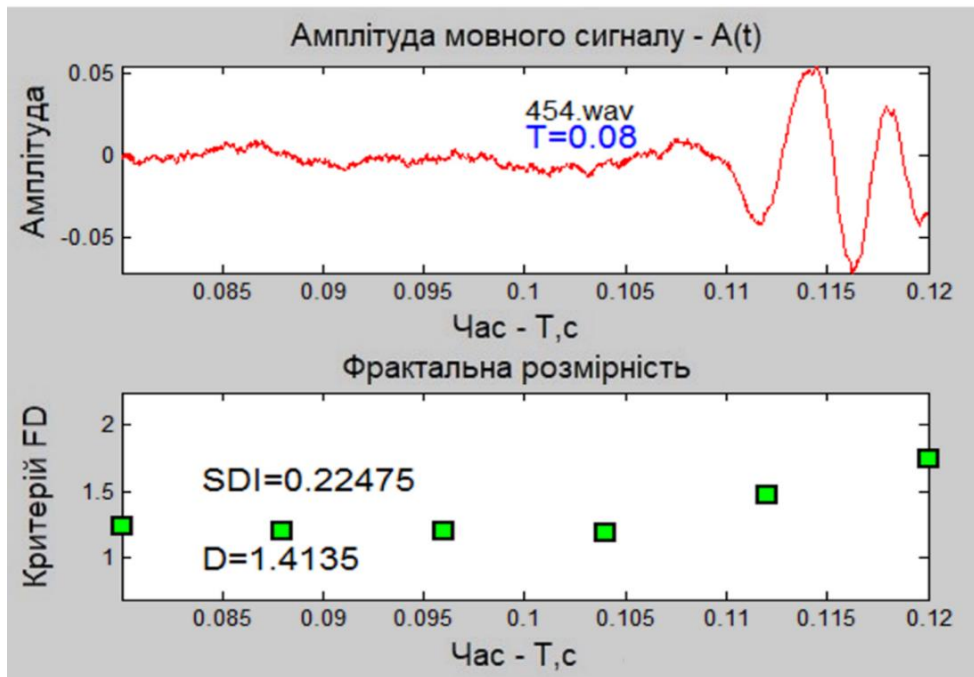


Рис. 2. Динаміка зміни фрактальної розмірності при переході з паузи на мову

Нехай $A(t_i)$ – величина амплітуди фрагменту промови мовного сигналу в момент часу t_i . Розглянемо часове вікно фрагменту промови з інтервалом δT . Цей інтервал, на основі проведеного в роботі аналізу, 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 мовного сигналу у вигляді

$$C(T_k, F_b, F_c) = (1/N) \text{abs} \left(\sum_{t_{ij}=0}^{N_m} C_{mor}(t_i, T_k, F_b, F_c) A(t_i) \right) \quad (13)$$

де $C(T_k, F_b, F_c)$ – значення модуля коефіцієнта вейвлет-перетворення, N – кількість дискретних відліків на інтервалі δT 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.

Розглянуте подання фрагмента мови у вигляді просторової скейлограми володіє важливими особливостями, що дозволяють істотно підвищити ефективність виявлення самоподібних структур. На рис. 3 наведена ілюстрація, на якій поєднані фрагменти сигналу мови і скейлограма, яка побудована на основі розглянутої моделі. Аналіз показує, що розташування «гребнів» скейлограм по часовому параметру строго відповідає локальним максимумам амплітуди мовного сигналу в часовій області. При цьому локальні максимуми відповідають сплескам амплітуди мовного сигналу, зумовленими частотою OT .

Суттєвою особливістю характеристик «гребнів» є їх форма. 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.

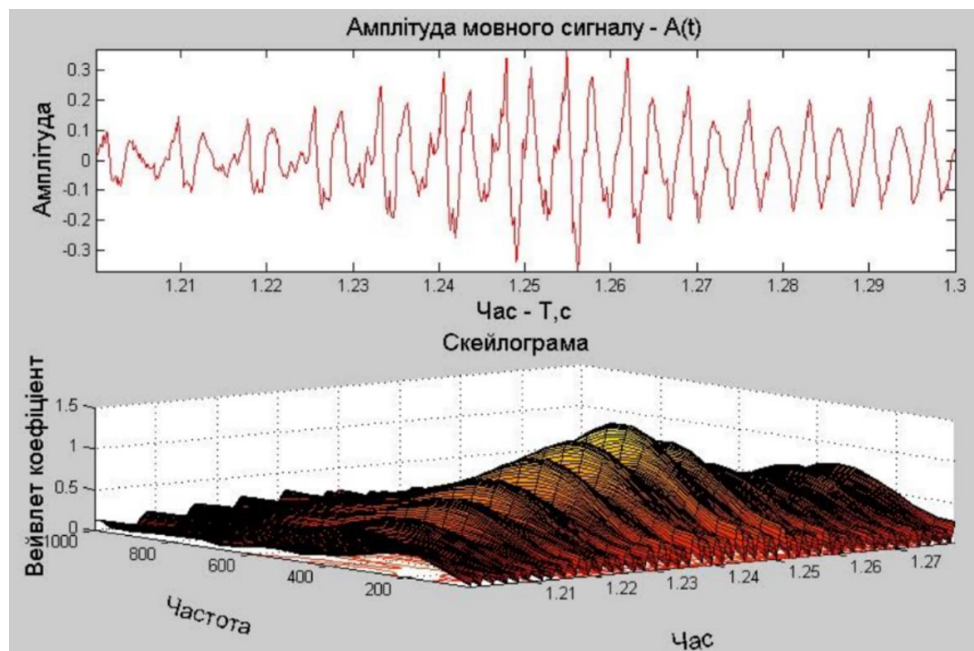


Рис. 3. Ілюстрація виявлення самоподібних структур

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 мовного сигналу в областях локальних максимумів. Вейвлет-перетворення Морле ефективно при даному підході виділяє ці структури. Розглянута подоба має вельми прозоре фізичне трактування. Структури амплітуди мовного сигналу 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.

Наявність самоподібних структур, що характеризують МС конкретної особи, може лежати на різних рівнях розкладання сигналу. Тому, запропоновано використання декомпозиції МС для визначення характеристик самоподібних структур на основі аналізу мультифрактального спектру. Визначено принципи вибору максимумів вейвлет-перетворення, що характеризують самоподібні структури на різних рівнях декомпозиції МС. Описано алгоритм виділення характеристик самоподібних структур, що використовує мультифрактальний спектр МС на основі вейвлет-перетворення.

Відомо, що фонетики в якості ідентифікації мовних сигналів осіб виконують пошук ударних голосних та порівняння їх спектрограм. Зважаючи на це, основну, найбільш важливу інформацію, несуть в собі вершини піків спектрограм. На основі цього запропоновано наступний метод підвищення інформативності частоти OT на основі характеристик самоподібних структур спектрограми МС:

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 (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:

$$S_{i-1} < S_i < S_{i+1} \quad (14)$$

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

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

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

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

– по результатам отриманих значень різниць максимумів будуємо нормовану гістограму розподілу частоти ОТ;

– по результатам отриманих значень різниць максимумів будуємо нормовану гістограму розподілу частоти ОТ тільки для тих фрагментів, де значення параметрів мультифрактального спектру двох МС мають найбільш близькі характеристики.

Запропоновано в якості характеристик самоподібних структур використовувати криві сингулярності мультифрактального спектру. Подоба кривих сингулярності мультифрактального спектра оцінювалася на основі гістограм відносної частоти показника Гельдера і ранжувалися за ступенем близькості (близькість оцінювалася на основі абсолютного значення різниці відповідних елементів гістограми відносної частоти показника Гельдера), 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.

Визначення формантних ознак МС здійснюється за рахунок алгоритму пошуку максимумів вейвлет-перетворення при огинанні спектру сигналу кубічним сплайном. Згідно аналізу існуючих досліджень, найбільш інформативними та достатніми для ідентифікації мовного сигналу є чотири перші форманти, тому в роботі вони приймалися в якості формантних ознак МС, що використовуються при його ідентифікації.

Запропоновано метод сегментації МС, алгоритм виділення самоподібних структур та обчислення на основі них частоти ОТ та максимумів формант, що моделюють природний процес сегментації та виділення інформаційних ознак особи з мовного фрагменту за рахунок фрактального та вейвлет аналізу.

У **четвертому розділі** на базі запропонованих методу та алгоритму розроблено рекомендації щодо проектування системи ідентифікації особи за МС. Досліджено інформативність алгоритмів оброблення МС, інформативних ознак ідентифікації особи та методів прийняття рішень. Проведено оцінювання достовірності розпізнавання особи на основі запропонованого методу системи ідентифікації особи за МС.

В рамках задачі ідентифікації особи можна виділити два взаємопов'язані завдання ідентифікації та верифікації особи. У першій задачі мета полягає в ідентифікації фрагмента МС як виголошеного одною з осіб з розглянутої множини, в другій – у встановленні належності мовного сигналу конкретній еталонній особі.

На підставі аналізу задач ідентифікації побудова системи повинна складатися:

1. Визначення індивідуальних ознак МС.
2. Представлення характерного еталона особи.
3. Прийняття рішення про індивідуальність мовця.

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

Вимірювання фрактальної розмірності компонентів сигналу. Простий у реалізації етап, але досить ефективний в наборі всіх засобів пошуку відмінностей. Реалізація його можлива як з постійним вікном, так і з адаптивним.

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

Виділення основного тону та основних формант. Для вирішення завдання виділення основного тону та формант існує необхідність використання методів слідування 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 виконується порівняння виділення характерних мовних ознак з бази на основі ймовірнісного підходу.

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

1. Середньоквадратичне відхилення фрактальної розмірності.
2. Діапазон розмірів часового вікна.

На цьому етапі дослідження вибір оптимального порогу вікна сканування при оцінюванні фрактальної розмірності здійснюється на основі оцінки достатньої точності ідентифікації пауз. 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 фрактальної розмірності $\partial D = 0,01 - 0,08$.

В результаті виконаного дослідження, були помічені достатні до використання в задачі сегментації відмінності фрактальної розмірності пауз від фрагментів мови. Фрактальна розмірність D для невокалізованих фрагментів в 99 % перебувала в межах $1,04 \leq D \leq 1,45$, а фрактальна розмірність вокалізованих фрагментів не спостерігалася менш $D = 1,55$ (Рис. 4) для часового вікна розміром 20 мс.

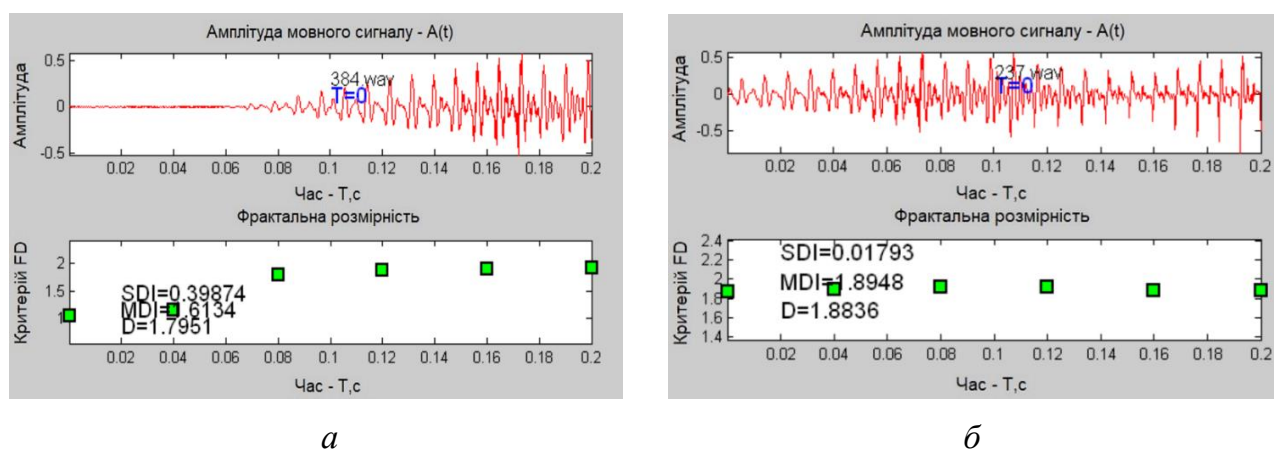


Рис. 4. Типовий розподіл фрактальної розмірності для фрагмента мовного файлу
a – з початковою паузою і початком промови з 0,08 с; *б* – без пауз

Важливим фактором є стабільність фрактальної розмірності як для вокалізованих так і невокалізованих фрагментів на ділянках 20-40 мс. Середньоквадратичне відхилення фрактальної розмірності на таких фрагментах не перевищує в середньому 0,02. Досить істотні зміни фрактальної розмірності для фрагментів промови відбуваються на ділянках мовного сигналу, 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. 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. Результати дослідження представлені в таблиці 1. Таким чином, за ефективністю розроблений метод сегментації МС краще виконує свої функції, та може бути рекомендований до використання в створюваній системі ідентифікації особи.

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

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

Сумарні результати випробувань за оцінкою узагальненої помилки представлені на рис. 5, 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.

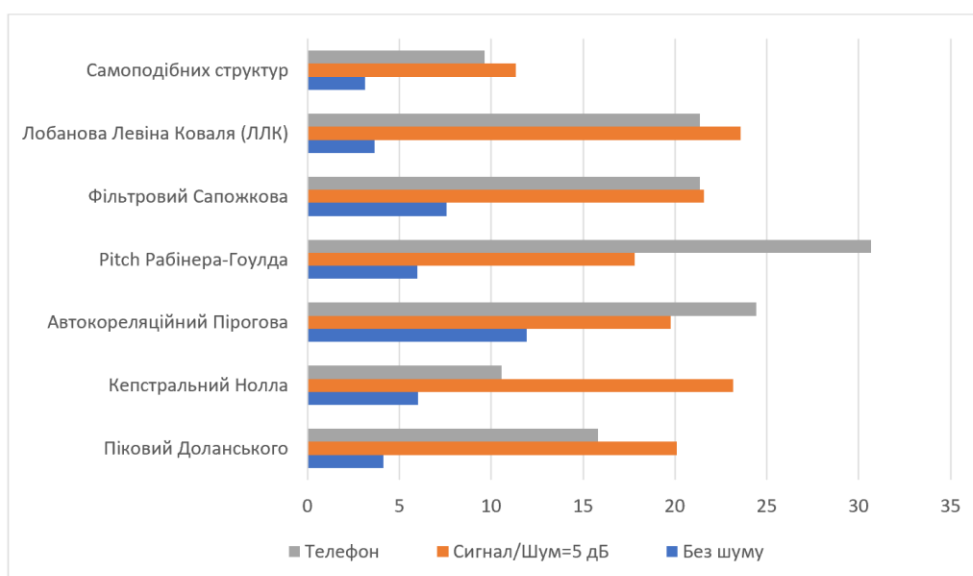


Рис. 5. Результати випробувань методів виділення ОТ (%)

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 $C/\text{Ш} = 5 \text{ dB}$, 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.

На базі власного методу створена інформаційна підсистема з автоматичного виділення контуру ОТ. Підсистема стала одним з компонентів розроблюваної системи ідентифікації особи.

Попередні експерименти з використання запропонованої підсистеми автоматичного виділення частоти ОТ та частот формант, що працює сумісно с 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.

Виходячи з запропонованого вище 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.

3. Програмна система має виконувати статистичну обробку характеристик самоподібних структур та підготовку їх до порівняння.

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

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

Згідно виконаного дослідження побудована програмна система (рис. 6). Програмна система мовної ідентифікації 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 та частот формант, що виділяються з промови, зафіксованої в МС.

– Обчислення близькості кривих функцій щільності вірогідності для кожного з цих ознак окремо.

– Обчислення міри близькості абсолютних максимумів спектрів формант, що виділяються з промови, зафіксованої в МС.

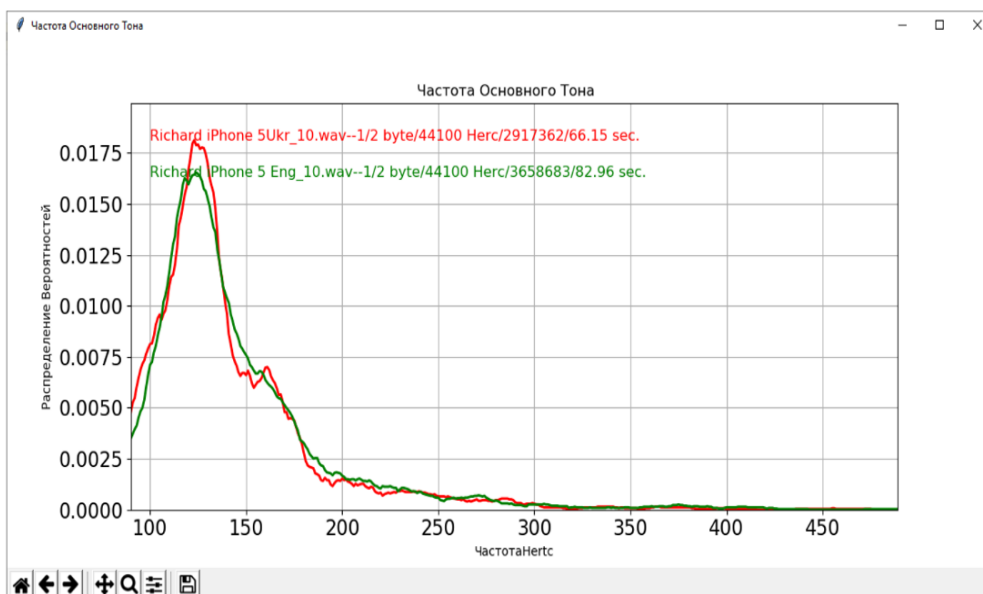


Рис. 6. Порівняння функцій розподілу частоти ОТ

Особливістю методу, який було застосовано для розрахунку спектра є те, що розрахунок спектральних характеристик проводиться з роздільною здатністю 1 Гц. В спектрі кожного часового вікна виділяються максимуми для кожного частотного діапазону. Згідно розподілу максимумів визначаються двовимірна щільність вірогідності для частоти ОТ та частот максимумів формант, що є нелінійними емпіричними функціями амплітуди і частот максимумів формант (Рис. 7). Ступінь близькості таких мовних характеристик двох осіб визначається абсолютними різницями між двовимірними щільностями вірогідності.

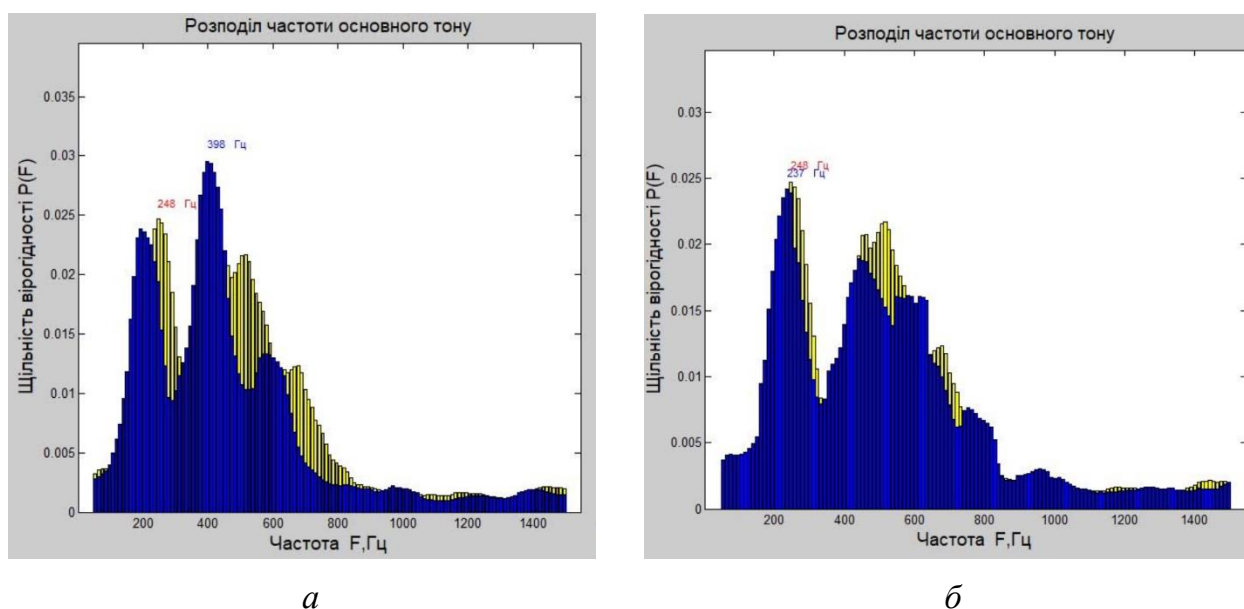


Рис. 7 – Розподіл щільності вірогідностей для частоти ОТ на основі самоподібних структур *a* – двох МС різних осіб; *б* – двох МС для однієї особи

В роботі порівняння МС осіб i і j здійснювалось за допомогою визначального правила при заданому пороговому значенні Δ_{lim}

$$g(i, j) = \begin{cases} 1, & \text{якщо } \Delta(i, j) < \Delta_{lim} \\ 0, & \text{інакше} \end{cases} \quad (15)$$

де $\Delta(i, j)$ міра подоби двох осіб, визначається як

$$\begin{aligned} \Delta(i, j) = \gamma & \left(\sum_{t=1}^N |p_i^{F0MF}(F_t) - p_j^{F0MF}(F_t)| + \sum_{t=1}^N |p_i^{F0}(F_t) - p_j^{F0}(F_t)| \right) \\ & + \sum_{t=1}^N |p_i^{F1}(F_t) - p_j^{F1}(F_t)| + \sum_{t=1}^N |p_i^{F2}(F_t) - p_j^{F2}(F_t)| + \\ & + \sum_{t=1}^N |p_i^{F3}(F_t) - p_j^{F3}(F_t)| + \sum_{t=1}^N |p_i^{F4}(F_t) - p_j^{F4}(F_t)| \end{aligned} \quad (16)$$

де γ це коефіцієнт, що регулює ступень впливу частоти ОТ відносно максимумів частот формант, N – кількість точок розбиття щільності вірогідності, $p_i^{F0MF}(F)$, $p_i^{F0}(F)$, $p_i^{F1}(F)$, $p_i^{F2}(F)$, $p_i^{F3}(F)$, $p_i^{F4}(F)$ – значення щільності вірогідності для частоти ОТ з урахуванням та без урахування характеристик мультифрактального спектру і чотирьох формант, що відповідає особі i , $p_j^{F0MF}(F)$, $p_j^{F0}(F)$, $p_j^{F1}(F)$, $p_j^{F2}(F)$, $p_j^{F3}(F)$, $p_j^{F4}(F)$ – значення щільності вірогідності для частоти ОТ з урахуванням та без урахування характеристик мультифрактального спектру і чотирьох формант, що відповідає особі j .

Для налаштування параметрів γ, Δ використовували датасет VoxCeleb з мовними сигналами 7000 осіб. При проведенні комп'ютерного експерименту обчислювались значення похибки 1 та 2 роду (рис. 8) в залежності від коефіцієнту впливу частоти ОТ (в діапазоні $\gamma = 0 - 4$) та міри подоби. 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 %).

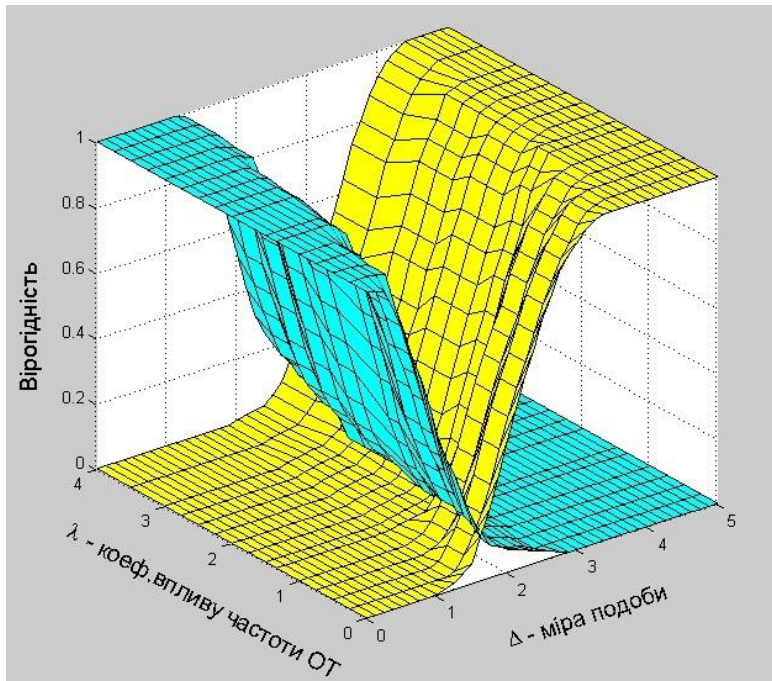


Рис. 8. Залежності похибок 1 та 2 роду від коефіцієнта впливу частоти ОТ та міри подоби

Наступним етапом була оцінка ефективності розробленої програмної системи ідентифікації особи, що використовує створені метод сегментації МС та метод підвищення інформативності частоти ОТ та алгоритм виділення максимумів формант. Дослідження виконувалось з набором з 300 МС від 5 осіб для однакових умов запису та шуму МС. Результати представлені в таблиці 2. Згідно за таблицею створена програмна система ідентифікації особи по ефективності перевищує існуючі комерційні аналоги та може бути рекомендованою до використання в мовних базах даних, що потребують мовної ідентифікацію особи.

Таблиця 2

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 %

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 %).

Таким чином, у четвертому розділі було виконано перевірку ефективності розроблених методів та розроблена ПС, що реалізує обробку МС і ідентифікацію особи; проведено порівняльний аналіз ефективності розпізнавання особи в порівнянні з існуючими подібними програмними системами.

ВИСНОВКИ

У дисертаційній роботі представлено вирішення актуальної наукової задачі попередньої обробки МС, сегментації, виділення ідентифікаційних ознак особи у вигляді параметрів частоти ОТ та частот формант і ідентифікації 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 характеристик 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 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.

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.

3. Соловьев В.И., Белозерова Я.А. Использование фрактальной размерности аудиофайлов в задаче сегментации звукового файла / В.И. Соловьев., Я.А. Белозерова // Научный журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2013. – № 5(194) ч.2. – С. 165 – 168.

4. Соловьев В.И., Белозерова Я.А. Анализ алгоритмов построение системы идентификации диктора / В.И. Соловьев., Я.А. Белозерова // Научный журнал. Вісник Східноукраїнського національного університету імені Володимира Даля. – 2013. – № 6(195) ч.1. – С. 62 – 66.

5. Белозерова Я.А. Выделение идентификационных характеристик диктора на основе вейвлет-разложения голосового сигнала / Я.А. Белозерова // Научный 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 // Научный 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 // Научный журнал. Інженерія програмного забезпечення. – 2017. – № 2(30). – С. 24 – 31.

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 // Научный журнал. Інженерія програмного забезпечення. – 2017. – № 3(31). – С. 34 – 38.

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.

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

12. Zybin Serhii, Bielorova Yana. Practical approach to speech identification / Serhii Zybin, Yana Bielorova // International Journal “Information models & analyses”. – 2020. – Vol.9. – № 3. – pp. 224 – 231.

13. Zybin Serhii, Bielorova Yana. Risk-based decision-making system for information processing systems / Serhii Zybin, Yana Bielorova // International

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 б 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.

АНОТАЦІЯ

Белозьорова Я.А. Метод застосування вейвлет аналізу в задачах ідентифікації мовної інформації. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.03 – математичне та програмне забезпечення обчислювальних машин і систем. – Національний авіаційний університет, Київ, 2021.

Дисертаційна робота присвячена розробці на основі апарату фрактального та вейвлет аналізу моделі та методу, що здійснюють обробку, визначення індивідуальних характеристик особи та ідентифікацію мовного сигналу. У роботі обґрунтовано необхідність використання фрактального та вейвлет аналізу для ідентифікації мовної інформації. Для сегментації мовного сигналу розроблено метод, що використовує фрактальні характеристики сигналу. Застосування методу дозволяє виділяти вокалізовані та невокалізовані фрагменти сигналу, незалежно від мовних даних особи. Запропоновано алгоритм виділення характеристик самоподібних структур в мовному сигналі на основі аналізу максимумів вейвлет-перетворення на різних рівнях декомпозиції мовного сигналу, що дозволяє визначити параметри частоти основного тону та частот формант у вигляді кривих щільності розподілу вірогідності. Розроблено метод ідентифікації особи, що використовує метод сегментації мовного сигналу та алгоритм виділення характеристик самоподібних структур. На основі виконаного дослідження створено програмну систему, яка по мовним записам здійснює автоматичний розрахунок мовних характеристик, виконує ранжування цих характеристик в базі даних, за визначеними в роботі критеріями, та ідентифікацію особи в мовному сигналі.

Ключові слова: мовний сигнал, вейвлет перетворення, фрактальна розмірність, частота основного тону, сегментація сигналу, вокалізований фрагмент.

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.

Підписано до друку 21.07.2021р. Формат 60x90 1/16.
Папір офсетний. Умовн. др. арк. 0,9.
Друк різнограф. Тираж 100 прим. Зам. № - 2107/01.

Надруковано ФОП Гузік О.М.
Податковий номер № 2705814113
м. Київ, вул. Богдана Гаврилишина, 16
Тел.: 338-16-61.