



Я. А. Белозорова

Національний авіаційний університет, м. Київ, Україна

МЕТОД ПІДВИЩЕННЯ ІНФОРМАТИВНОСТІ ЧАСТОТИ ОСНОВНОГО ТОНУ В ЗАДАЧАХ ІДЕНТИФІКАЦІЇ МОВЦЯ

Розглянуто питання підвищення точності ідентифікації мовця завдяки аналізу фізичної природи формування особливостей мовного сигналу та математичного опису структури сигналу. Здійснено огляд процесу ідентифікації особи, зроблено висновок, що в сигналі мають бути постійні самоподібні структури, які формуються під час мовлення кожної конкретної особи. Подібність структур самим собі можлива за рахунок їхнього масштабування в мовних фрагментах. На підставі проведеного аналізу визначено основні параметри опису індивідуальних особливостей голосу мовця у вигляді частоти основного тону та спектральних характеристик мовного сигналу. Проведений огляд методів визначення частоти основного тону дав змогу виділити напрями поліпшення точності ідентифікації мовця внаслідок точнішого математичного опису унікальних ознак мовного сигналу. Під час аналізу зроблено висновок, що найвдалішим інструментом виявлення самоподібних структур можна вважати фрактальний та вейвлет-аналіз. Обґрунтовано використання комплексного вейвлету Морле для опису мовного сигналу. У вигляді оцінки частоти основного тону мовного сигналу в роботі розглянуто відстані між локальними частотними максимумами скейлограм. Важливим фактором стійкості і достовірності оцінок частоти основного тону для цього методу є можливість оцінки частоти основного тону не тільки по локальних максимумах, але і по кореляції між фрагментами областей максимумів. На підставі вейвлет-перетворення та мультифрактального спектра запропоновано алгоритм виділення характеристик самоподібних структур, притаманних мовцю, та розроблені методи оброблення мовного сигналу дають змогу використовувати їх для побудови систем ідентифікації мовного сигналу та для створення інтелектуальних систем взаємодії користувача й комп'ютера. На підставі алгоритму запропоновано метод підвищення інформативності частоти основного тону для мовної ідентифікації особи, в якому, на відміну від наявних, за ознаку для розпізнавання використано значення коефіцієнтів вейвлет-перетворення на відрізках мовного сигналу, де спостерігаються екстремуми кореляційних функцій частоти основного тону. Проведений аналіз точності запропонованого методу показав достатній для використання рівень його ефективності.

Ключові слова: мовна ідентифікація особи; вейвлет Морле; сегментація мовного сигналу; форманти; атомарні структури; мультифрактальний спектр; скейлограма.

Вступ / Introduction

Мовлення є універсальною формою спілкування. Розпізнавання мовця – це процес ідентифікації спікера за вокальними особливостями даної промови. Це відрізняється від розпізнавання мовлення, де процес ідентифікації обмежується вмістом, а не мовцем. Процес розпізнавання мовця заснований на виявленні та виділенні унікальних характеристик мовлення оратора. Характеристики голосу людини також відомі як біометрія голосу.

Системи розпізнавання мовця використовуються для ідентифікації та розрізнення спікерів і виділення унікальних характеристик, які можуть бути використані для верифікації або автентифікації користувача. Розпізнавання мовця є відомим як процес ідентифікації мовця з даного мовного фрагменту способом порівняння голосових біометричних даних даного зразка мовця.

Об'єкт дослідження – процес ідентифікації мовця.

Предмет дослідження – моделі та методи цифрового оброблення й ідентифікації мовного сигналу (МС).

Мета роботи – розробити метод підвищення інформативності частоти основного тону (ОТ) в задачах ідентифікації мовця на підставі апарату фрактального та вейвлет-аналізу.

Для досягнення зазначеної мети визначено такі основні завдання дослідження:

- 1) Проаналізувати математичні моделі формування та сприйняття МС для задач їх цифрового оброблення.
- 2) Використовуючи фрактальний аналіз МС, розробити метод його сегментації.
- 3) На базі вейвлет-аналізу розробити алгоритм виділення характеристик самоподібних структур у МС, що можуть використовуватись як унікальні ознаки під час ідентифікації особи.
- 4) На підставі 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.

Аналіз останніх досліджень та публікацій. Осно-

Інформація про автора:

Белозорова Яна Андріївна, канд. техн. наук, доцент, кафедра інженерії програмного забезпечення.

Email: bryukhanova.ya@gmail.com; <https://orcid.org/0000-0002-0688-3436>

Цитування за ДСТУ: Белозорова Я. А. Метод підвищення інформативності частоти основного тону в задачах ідентифікації мовця. Науковий вісник НЛТУ України. 2023, т. 33, № 1. С. 82–88.

Citation APA: Bielozorova, Ya. A. (2023). Method of increasing awareness of main tone frequencies in speaker identification problems. *Scientific Bulletin of UNFU*, 33(1), 82–88. <https://doi.org/10.36930/40330111>

вою будь-якої системи пошуку голосів є параметри індивідуальних характеристик голосу мовця. У сучасних системах автоматичної ідентифікації мовця параметри індивідуальних характеристик голосу зазвичай визначаються на підставі двох головних факторів – частоти основного тону ОТ і спектральних характеристик [3, 30]. Ефективність таких систем значно залежить від методів визначення цих факторів і стійкості спектральних характеристик і частоти ОТ.

На сьогодні відомо, що вокалізовані фрагменти МС особи зберігають періодичність коливань голосових зв'язок. Мовний сигнал потрібно розглядати у вигляді пульсувальних коливань повітряного потоку, де період повторення імпульсів МС має назву частоти ОТ. Досліджено, що ОТ визначає структуру МС, а також є головним параметром мовного сигналу. 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 містить великий обсяг інформації, що міститься в МС особи. З погляду представлення цієї інформації, сам процес виділення частоти ОТ має незалежний інтерес [1]. Під час обчислення 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 таким моментам [1]:

- 1) можливість роботи за різних рівнів 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 призводять 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.

Найпопулярніші методи виділення частоти ОТ розглянемо нижче.

Піковий метод, що вперше запропонував Доланський [7], 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 [7] дослідження різних варіантів пікових методів показали їх погану роботу в умовах шуму та складних змін форми акустичного сигналу, окрім цього, ці методи дуже чутливі до пригнічення низькочастотної компоненти МС.

Кепстральний метод, що вперше запропонував Нолл [20], розраховує частоту ОТ так. 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 це 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 за різних змін МС на переходах між вокалізованими і невокалізованими фрагментами та не визначає моменти збудження мовного тракту [21].

Автокореляційний метод, що запропонував Пірогов, обчислює частоту ОТ за рахунок автокореляційної функції, що шукає максимуми у МС. Цей метод дуже чутливий 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 [17].

Метод Pitch, що запропонував Рабінер-Голд [9], полягає у тому, що спочатку МС обмежують фільтром низьких частот в діапазоні 600-800 Гц. Далі, вже у відфільтрованому МС, визначають відстані між усіма максимумами сигналу, знаходять відстані між усіма мінімумами сигналу та відстані між мінімумами і максимумами. І вже після всіх цих кроків визначають період за найбільшою кількістю збігів виділених відстаней. На жаль, цей метод також не визначає моменти збудження мовного тракту та має збільшену кількість помилок за рахунок швидких змін МС та не дає змоги відстежувати миттєві значення частоти ОТ [10].

Фільтровий метод, що вперше запропонував Сапжков [24], полягає у тому, що на мовний сигнал накладаються обмеження фільтром нижніх частот так, що після цієї своєрідної фільтрації, МС містить тільки 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 відстежити швидкі зміни частоти ОТ, саме на переходах між вокалізованими та невокалізованими фрагментами, за рахунок того, що вимагає детальнішого встановлення частоти зрізу фільтра низьких частот.

Метод ЛЛК (Лобанова – Левіна – Ковалю) обчислює частоту ОТ так. Визначається автокореляційна функція МС, знаходять найімовірніше значення періоду основного тону, далі піковим методом обчислюють значення максимумів МС. Під час підрахунків частоти ОТ цим методом маємо велику кількість помилок, особливо на ділянках між вокалізованими та невокалізованими фрагментами.

Методи лінійного передбачення. Принципова різниця цієї групи методів полягає у тому, що всі вони використовують зворотній лінійний фільтр. Весь механізм зводиться до обчислення параметрів МС для отримання певних характеристик його спектральних властивостей. З наближенням частоти повторення у зворотному фільтрі до частоти ОТ спостерігається більше і сильніше вирівнювання спектра. Складність цього методу полягає в постійному розрахунку спектра МС та підлаштування частоти повторення у зворотному фільтрі. Загалом робота методу має задовільні оцінки, але тільки доки спектр вирівнюється добре. Також достовірно

відомо, що за частот ОТ вище 200 Гц методи цієї групи показують погані результати [12].

Отже, точність методів визначення частоти ОТ можна вважати недостатньою через наявність недоліків кожного з розглянутих методів. Тому розроблення методів визначення частоти ОТ, що дають змогу точніше стежити за її траєкторією та використовують такі унікальні ознаки МС, що характеризують його особливості, є актуальною науковою задачею.

Результати дослідження та їх обговорення / Research results and their discussion

Припустимо, що можливість впізнавати особу за мовою є здібністю виділяти у фрагменті мови характерних ознак, притаманних особі, що упізнається. Причому ідентифікація особи є дією з виділення ознак, що повторюються в сигналі (є постійними у сигналі), а отже, можна зробити висновок, що в сигналі мають бути постійні самоподібні структури, які формуються під час мовлення кожної конкретної особи.

Подібність структур самим собі можлива завдяки їхньому масштабуванню в мовних фрагментах. Тому найвдалішим інструментом виявлення самоподібних структур можна вважати вейвлет-аналіз.

Найпопулярнішим методом представлення результатів вейвлет-перетворення є скейлограма (рис. 1), що дає змогу візуально подати та оцінити розташування екстремумів поверхні, що побудовані на підставі вейвлет-коефіцієнтів $W(a,b)$.

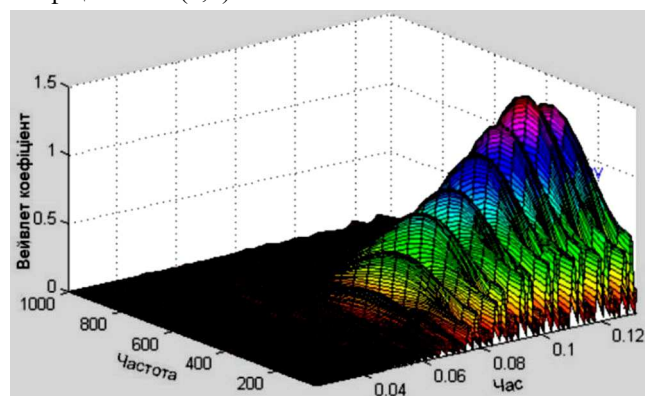


Рис. 1. Скейлограма фрагмента промови / Scale diagram of a speech fragment

Особливий інтерес становлять локальні екстремуми поверхні коефіцієнтів. Теоретично, аналіз самоподібних структур можливо виконувати на підставі параметрів скейлограми, але існує ціла низка статистичних функцій, що дають змогу оцінювати характеристики спектра ефективніше.

Загальний стан подібних статистичних функцій міри можна подати у вигляді

$$M(q,a) = \sum_{l \in L(a)} |W(a,t_l(a))|^q, \quad (1)$$

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

Згідно з даними [19], виконується залежність

$$M(q,a) \sim a^{\tau(q)}, \quad (2)$$

де $\tau(q)$ визначається для значення q обчисленням нахилу $\ln(M(q,a))$ від $\ln a$, що називають скейлінговою експонентною. Задаючи значення q отримуємо залежність

$\tau(q)$. Залежність $\tau(q)$ дає змогу отримати мультифрактальний спектр МС на підставі вейвлет-перетворення [22], який може описати основні характеристики самоподібних структур.

Стійкість такого методу отримання характеристик самоподібних структур полягає у використанні частотно-часового вікна, що автоматично виконує операції усереднення, а також в отриманні модулів вейвлет-коефіцієнтів під час розрахунку. Треба відзначити, що з погляду енергетичного підходу з питань аналізу самоподібних структур найбільш значущими є саме максимумами коефіцієнтів вейвлет-перетворення на різних рівнях декомпозиції.

Будемо розглядати фрагменти мови в сигналі як дискретні часові ряди амплітуди звукової хвилі. Поставимо задачу визначення характеристик самоподібних структур у 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 зі збереженням форми. Одним з варіантів представлення цих структур є використання вейвлет-перетворення МС. У роботі обґрунтовано використання комплексного вейвлету Морле для опису МС:

$$C_{mor}(t_i, T_k, F_b, F_c) = (\pi F_b)^{0.5} \text{Exp}(2j\pi F_c t_i) \text{Exp}\left(\frac{-(t_i - T_k)^2}{F_b}\right), \quad (3)$$

де: F_b – параметр ширини вейвлету; F_c – центральна частота вейвлету; t_i – дискретні часові відліки; T_k – часовий відлік відповідний центральній частині часового вікна; j – комплексна одиниця.

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

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

Відстані між локальними частотними максимумами скейлограм є в цьому підході оцінкою частоти основного тону мови. Важливим фактором стійкості і достовірності оцінок частоти ОТ для цього методу є можливість оцінити частоту ОТ не тільки за локальними максимумами, але і за кореляцією між фрагментами областей максимумів. На малих інтервалах ці області є приблизно самоподібними структурами. Під час аналізу такої самоподібності можливо доволно "вирізати" однакові за розмірами часового вікна структури не орієнтуючись на максимуми скейлограм [5].

Так, наприклад, розглянемо мовний фрагмент тривалістю 1 с, за частоти дискретизації 44100 Гц. Будемо аналізувати його спектральні характеристики і параметри частоти основного тону з використанням часового вікна в 20 мс на підставі цієї моделі. Прийнемо мінімально можливий крок з дискретності переміщення вікна аналізу рівним величині зворотного частоті дискретизації. Тоді, кількість оцінок віконних спектрів і параметрів частоти основного тону за одну секунду буде порядку 44080 Гц. Ці значення скорельовані 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.

Перший рівень вейвлет-декомпозиції МС дає змогу 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 пропонуємо розглянути також таких рівні декомпозиції МС для визначення характеристик самоподібних структур на підставі аналізу мультифрактального спектра.

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 визначаються як [27]:

$$L_x(j, k) = \sup_{\lambda' \in 3\lambda_{j,k}} |d_x(j, k)|, \quad (4)$$

де: j – ступінь вейвлет-декомпозиції; k – час; $d_x(j, k)$ – вейвлет-коефіцієнти.

Для виділення самоподібних структур $L_x(j, k)$ використовуємо алгоритм відповідно до теорії мультифрактального спектра [13].

Відомо, що мовні експерти у ролі ідентифікації мовних сигналів осіб виконують пошук ударних голосних та порівняння їх спектрограм. Тому основну, найважливішу інформацію, несуть в собі вершини піків спектрограм. Зважаючи на це, запропонуємо такий 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 підставі (3).
4. Для визначення положення кожної окремої вершини піку тривимірного спектра Морле на даному інтервалі виконуємо порівняльний аналіз площини під кривою кожного зрізу тривимірного спектра Морле для кожного інтервалу.
5. На визначеному зрізі (з максимальним значенням площини на інтервалі) тривимірного спектра 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 спектра.
7. На етапі порівняння двох МС визначаються два типи залежностей частоти ОТ:
 - 7.1. Перша залежність – за результатами отриманих значень різниць максимумів будуємо нормовану гістограму розподілу частоти ОТ.
 - 7.2. Друга залежність – за результатами отриманих значень різниць максимумів будуємо нормовану гістограму розподілу частоти ОТ тільки для тих фрагментів, які мають близькі значення сингулярності мультифрактального спектра двох МС.

Проведені попередні дослідження статистичних характеристик особи, отримані на підставі самоподібних структур, показали стійку залежність щільності вірогідності від індивідуальних особливостей МС особи (рис. 2 і 3).

Наведені залежності щільності вірогідності на рис. 2 показують представлення цієї залежності для двох МС різних осіб. Видно, що центральні піки частоти ОТ припадають відповідно на 248 та 398 Гц. Зазвичай наявні 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 використовуються щільності вірогідності.

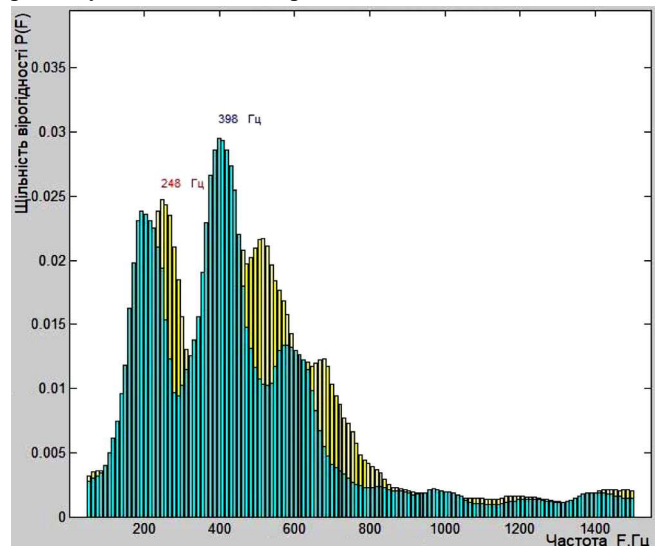


Рис. 2. Розподіл щільності вірогідності максимумів самоподібних структур для частоти ОТ для двох МС різних осіб / Probability density distribution of maxima of self-similar structures for the fundamental tone frequency for two speech signals of different persons

Про потребу використання щільності вірогідності також свідчить рис. 3, де представлені МС однієї і тієї ж особи з двома піками розподілу в 237 та 248 Гц відповідно (такий розподіл не дасть змоги ідентифікувати МС однієї особи, бо частота ОТ на них відрізняється).

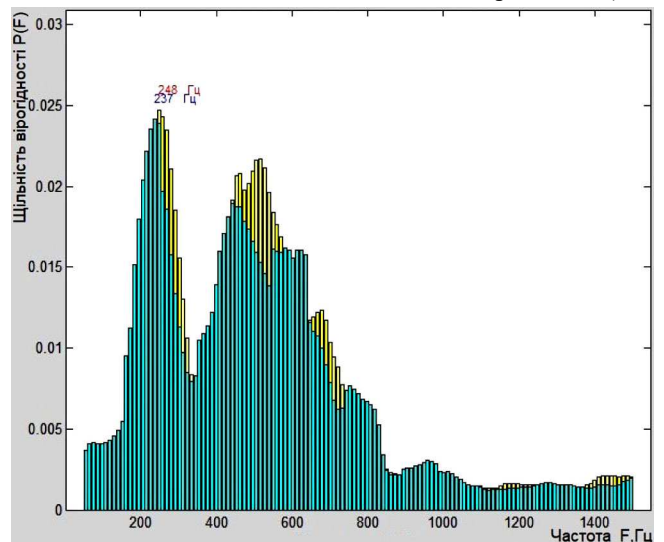


Рис. 3. Розподіл щільності вірогідності максимумів самоподібних структур для частоти ОТ для двох МС однієї особи / Probability density distribution of the maxima of self-similar structures for the fundamental tone frequency for two speech signals of the same person

Отже, наведена закономірність розподілу щільності вірогідності максимумів самоподібних структур для

частоти ОТ та чотирьох формант може бути використана в методі ідентифікації мовця.

Для отримання залежностей розподілу частоти ОТ виконується побудова гістограм зі збережених у 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 таку міру подоби:

$$\Delta Ot(i, j) = \sum_{t=1}^N |p_{i^{F_oMF}}^{F_o}(F_t) - p_{j^{F_oMF}}^{F_o}(F_t)| + \sum_{t=1}^N |p_{i^{F_o}}^{F_o}(F_t) - p_{j^{F_o}}^{F_o}(F_t)|, \quad (5)$$

де: $p_{i^{F_oMF}}^{F_o}(F_t), p_{j^{F_oMF}}^{F_o}(F_t)$ – значення щільності вірогідності для частоти ОТ з урахуванням та без урахування характеристик мультифрактального спектра, що відповідає особі i ; $p_{i^{F_o}}^{F_o}(F_t), p_{j^{F_o}}^{F_o}(F_t)$ – значення щільності вірогідності для частоти ОТ з урахуванням та без урахування характеристик мультифрактального спектра, що відповідає особі j .

Тоді порівняння осіб i і j для метода визначення частоти ОТ може здійснюватися за допомогою визначального правила за заданого порогового значення ΔOt_{lim} :

$$Ot(i, j) = \begin{cases} 1, \text{ якщо } \Delta Ot(i, j) < \Delta Ot_{lim}; \\ 0 - \text{ інакше.} \end{cases} \quad (6)$$

У ролі тестового мовного матеріалу було 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 із шумом виконувалося додавання білого шуму з необхідним співвідношенням. Співвідношення шуму обчислювалось так:

$$C/Ш = 10 \log \frac{E}{\sigma^2}, \quad (7)$$

де: σ^2 – дисперсія шуму; E – енергія вокалізованого фрагменту МС, яка визначалась як сума квадратів максимумів вейвлет-перетворення.

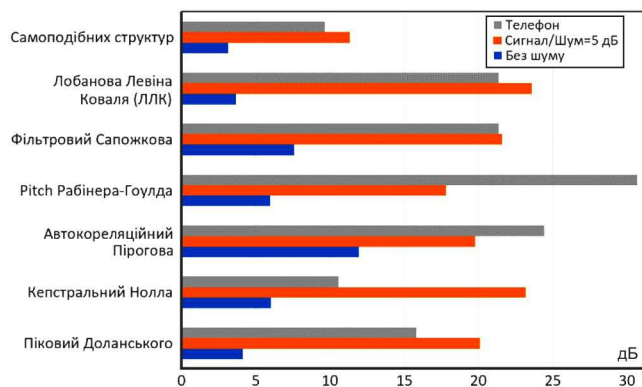


Рис. 4. Результати випробувань методів виділення ОТ (%) / Results of tests of methods of the main tone selection (%)

Сумарні результати випробувань за оцінкою узагальненої помилки подано на рис. 4, для десяти чоловіків, десяти жінок, 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 у визна-

ченні частоти ОТ та забезпечив правильне стеження за траєкторією частоти ОТ упродовж усього проголошення, навіть за співвідношення $C/Ш = 5$ dB, тоді як інші методи виявилися практично непрацездатні. Тому можна вважати, що метод є конкурентоспроможним порівняно з іншими розглянутими методами виділення ОТ, які вимагають ручного підлаштування параметрів.

На базі власного методу розроблено інформаційну підсистему з автоматичного виділення контуру ОТ. Розроблена підсистема стала одним з компонентів розробленої системи ідентифікації особи. Попередні експерименти з використання підсистеми автоматичного виділення частоти ОТ, що працює сумісно з розробленим методом сегментації МС [25], показали ефективність системи загалом.

Обговорення результатів дослідження. Визначення частоти основного тону є важливою задачею в області ідентифікації мовця. Існує значна кількість підходів до цієї задачі, що ґрунтуються на недавніх дослідженнях у галузі нейроанатомії, нейрофізіології, поведінкових досліджень та мовотворення [15, 18, 26, 28, 29]. Особливістю задач визначення частоти ОТ є неможливість порівняння методів визначення за рахунок особливостей отримання значень визначального параметра, тому задача точності визначення частоти основного є частиною задачі ідентифікації мовця і дуже часто не може розглядатися окремо [23]. Останнім часом було проведено низку досліджень для уточнення задачі визначення частоти основного тону [2, 4, 14]. Але, як зазначали раніше, кожен з викладених підходів не дає змоги отримати необхідний рівень точності для ідентифікації мовця.

Отже, за результатами виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження:

1. Уперше, для мовної ідентифікації особи, в ролі набору ознак запропоновано самоподібні структури, що дало змогу сформулювати новий спосіб представлення індивідуальних ознак для розпізнавання цієї особи.
2. Уперше запропоновано метод підвищення інформативності частоти основного тону (ОТ) для мовної ідентифікації особи, в якому, на відміну від наявних, за ознаку для розпізнавання використано значення коефіцієнтів вейвлет-перетворення на відрізках МС, де спостерігаються екстремуми кореляційних функцій частоти ОТ.

Практична значущість результатів дослідження.

Запропонований алгоритм виділення характеристик самоподібних структур та розроблені методи оброблення МС дають змогу використовувати їх:

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

Висновки / Conclusions

Розроблено метод підвищення інформативності частоти основного тону (ОТ) в задачах ідентифікації мовця на підставі апарату фрактального та вейвлет-аналізу. За результатами виконаної роботи можна зробити такі основні висновки:

1. Проведений огляд методів визначення частоти ОТ показав недостатню точність цих методів під час їх реалізації, тому потрібно розробити метод підвищення інформативності частоти ОТ, що дає змогу точніше

стежити за її траєкторією та використовує унікальні ознаки мовного сигналу (МС), що характеризує особу.

2. Обґрунтовано використання теорії фракталів для оцінювання ступеня мінливості мовного сигналу на підставі аналізу динаміки його поведінки. Показано, що подібність структур у мовному сигналі можливе завдяки їх масштабуванню в МС. Зважаючи на це, обґрунтовано, що основними підходами в задачі ідентифікації МС можуть бути фрактальний та вейвлет-аналіз.
3. Розроблено алгоритм виділення характеристик самоподібних структур МС на підставі максимумів вейвлет-перетворення. При виділенні використовується комплексний вейвлет Морле, що найефективніше описує самоподібні структури МС на першому рівні декомпозиції. Для отримання характеристик самоподібних структур на інших рівнях декомпозиції використовується мультифрактальний спектр.
4. Запропоновано метод підвищення інформативності частоти ОТ та алгоритм визначення формантних частот [31], що моделюють природний процес виділення інформаційних ознак особи з мовного фрагменту за рахунок фрактального та вейвлет-аналізу.
5. Досліджено ефективність визначення частоти ОТ з використанням методу підвищення інформативності частоти ОТ, що показав достатньо високі характеристики порівняно з іншими сучасними методами.

References

1. Agranovsky, A. V., & Lednov, D. A. (2004). Theoretical aspects of algorithms for processing and classifying speech signals. *Radio and communications*, 164. [In Russian].
2. Aidan, O. T. Hogg, Christine, Evers, & Patrick, A. Naylor. (2019). Speaker Change Detection Using Fundamental Frequency with Application to Multi-talker Segmentation. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 20(1). <https://doi.org/10.1109/ICASSP35589.2019>
3. Beet, S. W. (1990). Automatic speech recognition using a reduced auditory representation and position-tolerant discrimination. *Computer Speech & Language*, 4(1), 17–33. [https://doi.org/10.1016/0885-2308\(90\)90021-W](https://doi.org/10.1016/0885-2308(90)90021-W)
4. Candy, Olivia Mawalim, Kasorn, Galajit, Jessada, Karnjana, Shunsuke, Kidani, & Masashi Unoki. (2022). Show more Speaker anonymization by modifying fundamental frequency and x-vector singular value. *Computer Speech & Language*, 73(1), 101–126. <https://doi.org/10.1016/j.csl.2021.101326>
5. Cheveigne, A., & Kawahara, H. (2002). A fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America*, 111(4), 200–218. <https://doi.org/10.1121/1.1458024>
6. Diak, T. P., & Hrytsiuk, Yu. I. (2020). Design of the system of automated generation of poetry works. *Ukrainian Journal of Information Technology*, 3(2), 01–14. <https://doi.org/10.23939/ujit.2021.02.001>
7. Dolansky, L. O. (1955). An Instantaneous Pitch-Period Indicator. *The Journal of the Acoustical Society of America*, 27(1), 38–67. <https://doi.org/10.1121/1.1907499>
8. Fenii, N. S., & Hrytsiuk, Y. I. (2020). Automation of the process of classification of text news from internet sites by neural network methods. *Scientific Bulletin of UNFU*, 30(4), 123–133. <https://doi.org/10.36930/40300421>
9. Gold, V., & Rabiner, L. (1969). Parallel processing techniques for estimating pitch period of speech in the time domain. *The Journal of the Acoustical Society of America*, 46(2), 442–448. <https://doi.org/10.1121/1.1911709>
10. Gold, V., & Rabiner, L. (1976). Theory and Application of Digital Signal Processing. *Englewood Cliffs*, 7(4), 126–187. <https://doi.org/10.1002/piuz.19760070413>
11. Heller Murray, E. S., Lien, Y. S., Van Stan, J. H., Mehta, D. D., Hillman, R. E., Pieter Noordzij, J., and Stepp, C. E. (2017). Relative fundamental frequency distinguishes between phonotraumatic and non-phonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research*. 60(6), 1507–1515. <https://doi.org/10.1044/2016.JSLHR-S-16-0262>
12. Hsieh, C.-T. (2017). Segmentation of continuous speech into phonemic units. *International Journal of Computer Science and Mobile Computing*, 6(4), 420–424.
13. Jaffard, S., Lashermes, B., & Abry, P. (2006). Wavelet Leaders in Multifractal Analysis. *Wavelet Analysis and Applications. Signal Processing*, 6(89), 219–264. <https://doi.org/10.1007/978-3-7643-7778-6>
14. Jennifer, M. Vojtech, & Roxanne, K. Segina. (2019). Refining algorithmic estimation of relative fundamental frequency: Accounting for sample characteristics and fundamental frequency estimation method. *The Journal of the Acoustical Society of America*, 146(5), 31–84. <https://doi.org/10.1121/1.5131025>
15. Kimberly, L. Dahl, & Cara, E. Stepp. (2021). Changes in Relative Fundamental Frequency Under Increased Cognitive Load in Individuals With Healthy Voices. *Journal of Speech, Language, and Hearing Research*, 64(4), 1189–1196. <https://doi.org/10.1044/2021.JSLHR-20-00134>
16. Kornak, O. I., & Hrytsiuk, Y. I. (2021). Designing a fund management system for a charity fund. *Scientific Bulletin of UNFU*, 31(2), 121–132. <https://doi.org/10.36930/40310220>
17. Markel, J. D., & Gray, A. H. (1977). *Linear Prediction of Speech*. *Linguistic Society of America*, 53(3), 723–752. <https://doi.org/10.2307/413194>
18. Mohamed, A. Serry, Cara, E. Stepp, & Sean, D. Peterson. (2021). Physics of phonation offset: Towards understanding relative fundamental frequency observations. *The Journal of the Acoustical Society of America*, 149(5), 36–54. <https://doi.org/10.1121/10.0005006>
19. Muzy, J. F., Bacry, E., & Arneodo, A. (1991). Wavelets and multifractal formalism for singular signals: application to turbulence data. *American Physical Society*, 67(25), 3515–3518. <https://doi.org/10.1103/PhysRevLett.67.3515>
20. Noll, A. M. (1964). Short-time spectrum and "cepstrum" techniques for vocal-pitch detection. *The Journal of the Acoustical Society of America*, 36(5), 296–302.
21. Noll, A. M. (1969). Pitch determination of human speech by harmonic product spectrum, the harmonic sum spectrum and a maximum likelihood estimation. *Proc. of a symposium on Computer culture*, 5(3), 779–797.
22. Pavlov, A. N., & Anishchenko, V. S. (2007). Multifractal signal analysis based on wavelet transform. *Series: Physics*, 7(1), 3–25. <https://doi.org/10.18500/1817-3020-2007-7-1-3-25>
23. Rashid, Jahangir, Ying, Wah The, Henry, Friday Nweke, & Ghulam, Mujtaba. (2021). Speaker identification through artificial intelligence techniques: A comprehensive review and research challenges. *Expert Systems with Applications*, 171(1), 114–136. <https://doi.org/10.1016/j.eswa.2021.114591>
24. Sapozhkov, M. A. (1963). Speech signal in cybernetics and communications. *Radio and communications*, 452. [In Russian].
25. Solovyov, V. I., & Belozero, Y. A. (2013). Using the fractal dimension of audio files in the problem of audio file segmentation. *Scientific journal. Bulletin of the Eastern Ukrainian National University named after Volodymyr Dahl*, 5(194), 165–168. [In Russian]
26. Teichert, T., Gnanateja, G. Nike, & Sadagopan, S. (2022). A Linear Superposition Model of Envelope and Frequency Following Responses May Help Identify Generators Based on Latency. *Neurobiology of Language*, 3(3), 441–468. https://doi.org/10.1162/nol_a_00072
27. Wendt, H., & Abry, P. (2007). Multifractality Tests Using Bootstrapped Wavelet Leaders. *IEEE Transactions on Signal Processing*, 55(10), 4811–4820. <https://doi.org/10.1109/TSP.2007.896269>
28. Yeonggwang, Park, Feng, Wang, Manuel, Díaz-Cádiz, & Jennifer, M. Vojtech. (2021). Vocal fold kinematics and relative fundamental frequency as a function of obstruent type and speaker age. *The Journal of the Acoustical Society of America*, 149(4), 21–89. <https://doi.org/10.1121/10.0003961>
29. Yunan, Wu, Vibha, Viswanathan, & Taylor, Abel. (2022). Auditory cortical responses to speech are shaped by statistical learning

- of short-term speech input regularities. *BioRxiv*, 31(5), 534–560. <https://doi.org/10.1101/2022.12.19.520832>
30. Zue, V. W., Glass, J., Philips, M., & Seneff, S. (1990). Speech database development at MIT: Timit and beyond. *Speech Communication*, 9(4), 351–356. [https://doi.org/10.1016/0167-6393\(90\)90010-7](https://doi.org/10.1016/0167-6393(90)90010-7)
 31. Zybin, S., & Bielozorova, Y. (2022). Method of Extracting Formant Frequencies Based on a Vocal Signal. The International Conference on Artificial Intelligence and Logistics Engineering (ICAILE): Advances in Artificial Systems for Logistics Engineering, 135(2), 448–457. https://doi.org/10.1007/978-3-031-04809-8_40

Ya. A. Bielozorova

National Aviation University, Kyiv, Ukraine

METHOD OF INCREASING AWARENESS OF MAIN TONE FREQUENCIES IN SPEAKER IDENTIFICATION PROBLEMS

The paper deals with the issue of increasing the accuracy of speaker identification due to the analysis of the physical nature of the formation of speech signal features and the mathematical description of the signal structure. In the course of research, the person identification process was reviewed, which enabled concluding that the signal should contain constant self-similar structures that are formed during the speech of each specific person. Similarity of structures to themselves is possible due to their scaling in language fragments. On the basis of the conducted analysis, the main parameters of the description of the individual features of the speaker's voice in the form of the frequency of the main tone and the spectral characteristics of the speech signal were determined. The conducted review of methods for determining the frequency of the fundamental tone allowed highlighting the direction of improving the accuracy of speaker identification due to a more accurate mathematical description of the unique features of the speech signal. Moreover, fractal and wavelet analysis is found to be the most successful tool for detecting self-similar structures. The use of the complex Morlet wavelet for describing the speech signal is substantiated. In the form of an estimate of the frequency of the main tone of the speech signal, the work considers the distances between the local frequency maxima of the scalograms. An important factor in the stability and reliability of estimates of the frequency of the fundamental tone for this method is the possibility of estimating the frequency of the fundamental tone not only by local maxima, but also by correlation between fragments of the regions of maxima. On the basis of wavelet transformation and multifractal spectrum, an algorithm for identifying the characteristics of self-similar structures inherent in the speaker is proposed, and the developed speech signal processing methods allow using them to build speech signal identification systems and creating intelligent user-computer interaction systems. Based on the algorithm, a method of increasing the informativeness of the fundamental tone frequency for person's speech identification is proposed, in which the value of the wavelet transformation coefficients on segments of the speech signal unlike the existing ones, where the extremes of the correlation functions of the fundamental tone frequency are observed, are used as a feature for recognition. The analysis of the accuracy of the proposed method showed a sufficient level of its effectiveness for use.

Keywords: language identification of a person; Morlet wavelet; speech signal segmentation; formants; atomic structures; multifractal spectrum; scalogram.