ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ РОЗПІЗНАВАННЯ ДИКТОРА ЗА РАХУНОК СУМІСНОГО ВИКОРИСТАННЯ ЧАСТОТИ ОСНОВНОГО ТОНУ ТА ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ

Биков М.М., Ковтун В.В.
Вінницький національний технічний університет
Україна, м. Вінниця, вул. Хмельницьке шосе, 95,
e-mail: nmbdean@ksu.vstu.vinnica.ua

Abstract
The method of increasing of the speaker identification efficiensy by the way of joint using of the fundamental frequency and wavelet-transformation coefficients of a speech signal is offered in this article. The efficiency of different types of wavelet-transformation and classifier for speaker identification are investigated.

Вступ

Задача підвищення якості систем розпізнавання диктора безпосередньо пов'язана з питанням вибору ефективних методів цифрової обробки мовних сигналів. Процедура цифрової обробки сигналів повинна відповідати наступним основним вимогам: точне перетворення сигналу, можливість ефективного стиснення та можливість точного відтворення оригінального сигналу. Виходячи з особливостей задачі розпізнавання диктора, обробка мовного матеріалу повинна здійснюватись таким чином, щоб були збережені і точно локалізовані флуктуації частотно-часових характеристик мовного сигналу, притаманні конкретному диктору [1]. В роботі наводяться результати досліджень ефективності вейвлет-перетворення для розв'язання задачі розпізнавання диктора, де під ефективністю в даному контексті розуміється підвищення точності розпізнавання диктора порівняно з методом, запропонованим в [1]. Проведено аналіз впливу типу, виду та ступеня вейвлет-перетворення на результат розпізнавання.

Постановка проблеми, аналіз останніх досліджень і публікацій

Найчастіше для частотного аналізу складних сигналів, в тому числі мовних, використовують класичне перетворення Фур'є (неперервне та дискретне) яке, однак, в окремих випадках виявляється недостатньо ефективним [2, 3]. Наприклад, перетворення Фур'є не відрізняє сигнали із двох синусоїд з різними частотами, один з яких являє собою суму синусоїд, другий – дві синусоїди, що слідують одна за одною. В обох випадках їх спектр буде зображений у вигляді двох піків на фіксованих частотах [2, 3]. Отже, перетворення Фур'є у своєму традиційному вигляді не пристосовано для аналізу нестаціонарних сигналів, у тому числі локалізованих на деякому інтервалі часу, оскільки воно часто нівелює інформативні часові характеристики сигналу.

Вказаного недоліку аналізу мовних сигналів можна позбавитися за рахунок використання вейвлет-перетворення. Вейвлет-перетворення сигналів являє собою розклад сигналу за базисом солітоноподібних функцій (вейвлетів) двох аргументів – масштабу та зсуву в часі [3]. На відміну від традиційного перетворення Фур'є, вейвлет-перетворення забезпечує двовимірне представлення досліджуваного сигналу в частотній області та в площині частота-положення. Аналогом частоти при цьому є масштаб аргументу базисної функції (найчастіше – часу), а положення характеризується її зсувом. Це дозволяє розрізняти як завгодно малі де-

талі сигналу, одночасно локалізуючи їх на шкалі часу. Тому в даній роботі ставиться і досліджується проблема підвищення ефективності ідентифікації диктора за рахунок ознак, отриманих в результаті вейвлет-перетворення мовного сигналу.

Дана робота продовжує цикл статей, які присвячені підвищенню точності розпізнавання диктора шляхом використання інформативних ознак. Зокрема, в [1] автори обґрунтували можливість підвищення точності розпізнавання диктора за рахунок використання в якості ознак кореляційних функцій частоти основного тону. В даній статті автори пропонують вдосконалити згаданий метод шляхом використання в якості додаткових ознак для розпізнавання диктора значення коефіцієнтів неперервного та дискретного вейвлет-перетворення на ділянках сигналу, де спостерігаються екстремуми кореляційних функцій частоти основного тону диктора в часі.

Основні теоретичні передумови застосування вейвлет-аналізу для розпізнавання диктора

Наведемо теоретичні положення, що є основою використання вейвлет-аналізу цифрових сигналів для ідентифікації диктора [2-5]. Вейвлет-аналіз представляє собою одну з форм частотно-часового аналізу, який призначений для виявлення локальних флуктуацій сигналу. Враховуючи, що флуктуації мовного сигналу спостерігаються на досить малому проміжку часу, сам сигнал можна розглядатися як такий, що задано на числовій осі $\Re(-\infty,\infty)$ з нормою

$$\|f(t)\|^2 < 0.$$

Таким чином, базисні функції повинні швидко затухати при $|t| \to \infty$. Тому для того, щоб перекрити обраними базисними функціями весь часовий інтервал, на протязі якого спостерігається досліджуваний сигнал, необхідно, щоб базисні функції являли собою набір відповідних функцій, зміщених в часі. Звичайно, відповідний набір утворюється з функції-прототипа $\psi(t)$, зсунутої по осі t, тобто $\{\psi(t-b)\}$. Для того, щоб забезпечити також і частотний аналіз, базисна функція повинна мати масштабний коефіцієнт, який є аналогом частоти у перетворенні Фур'є, тоді базисні функції для частотно-часового аналізу матимуть наступний вигляд:

$$\psi\left(\frac{t}{a}-\frac{b}{a}\right)=\psi\left(\frac{t-b}{a}\right),\ a,b\in\Re.$$

Базисні функції повинні відповідати умовам обмеженості, локальності та нульового середнього.

Розрізняють неперервне та дискретне вейвлет-перетворення. В неперервному перетворенні (CWT) вводиться базис, який відповідає зазначеним умовам:

$$\psi_{a,b}(t)=\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \qquad (1)$$

де множник $\frac{1}{\sqrt{|a|}}$ потрібен для нормування функції $\|\psi_{f,b}(t)\| = \|\psi(t)\|$.

Нехай $a,b\in\Re$, тобто приймають довільні дійсні значення, тоді неперервне вейвлет-перетворення CWT буде мати такий вигляд:

$$CWT_f(a,b)=\langle f(t),\psi_{a,b}(t)\rangle=\frac{1}{\sqrt{|a|}}\int_{-\infty}^{\infty}f(t)\psi\left(\frac{t-b}{a}\right)dt, \qquad (2)$$

$$f(t)=C_{\psi}^{-1}\int_{-\infty}^{\infty}\frac{da}{a^2}\int_{-\infty}^{\infty}CWT_f(a,b)\psi\left(\frac{t-b}{a}\right)db, \quad (3)$$

де коефіцієнт нормування $C_{\psi}^{-1}=\int_{-\infty}^{\infty}\frac{|\Psi(\omega)|^2}{|\omega|}d\omega<\infty$.

Одержані в роботі результати розраховувались при b=1.

В дискретному вейвлет-перетворенні (DWT) введемо масштабний коефіцієнт $a=a_m=a_0^m\ (a_0>1)$, що еквівалентно розбиттю частотної осі на піддіапазони (частотні смуги). Покладемо $\omega_0=(a_0+1)\Delta_\omega$, тоді частотне вікно буде мати наступні границі $\left(\frac{\omega_0}{a_m}-\frac{\Delta_\omega}{a_m},\frac{\omega_0}{a_m}-\frac{\Delta_\omega}{a_m}\right)=\left(a_0^{-m+1}\Delta_\omega,a_0^{-m+2}\Delta_\omega\right)$, а центральна частота m-того вейвлету:

$$\frac{\omega_0}{a_m}=(a_0+1)a_0^{-m}\Delta_m.$$

Базисом для DWT буде функція, отримана з (1) при $a=a_0^m$:

$$\psi_{m,b}(t)=a_0^{-\frac{m}{2}}\psi\left(a_0^{-m}(t-b)\right).$$

Якщо $\psi_{m,b}(t)$ відповідає властивостям, характерним для базисних функцій, то будь-яка функція, що відповідає вищенаведеним умовам, може бути представлена в дискретному вигляді по $m\in Z$ послідовності:

$$DWT_f(m,b)=\langle f,\psi_{m,b}\rangle=a_0^{-\frac{m}{2}}\int_{-\infty}^{\infty}f(t)\psi\left(a_0^{-m}(t-b)\right)dt, \quad (4)$$

$$f(t)=\frac{2}{A+B}\sum_{j,k}DWT_f(j,k)\psi_{jk}(t)+R, \quad (5)$$

де похибку R відтворення вихідної функції $f(t)$ з коефіцієнтів $DWT_j(j,k)$ можна визначити з нерівності $\|R\|\le O\left(\frac{B}{A}-1\right)\|f\|$, де A, B - константи, такі, що $0<A\le B<\infty$.

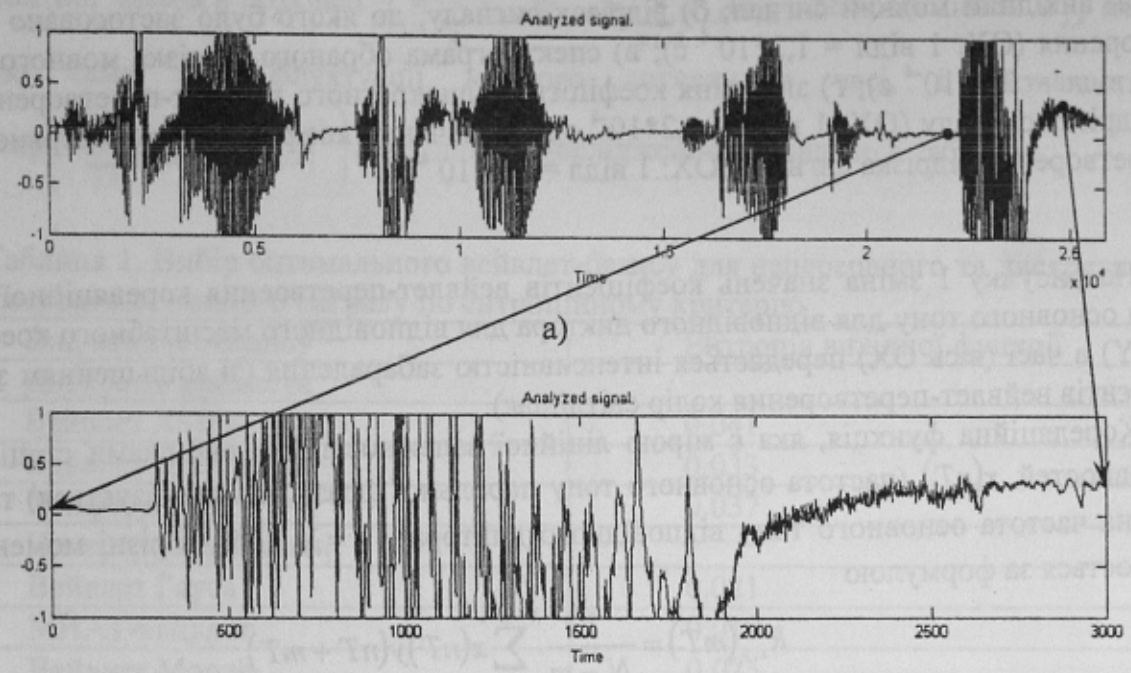
Результати експериментів та їх аналіз

Проведені в даній роботі експерименти мали своєю метою розв'язання таких задач:

1. Дослідження ефективності розпізнавання диктора за рахунок використання в якості ознак значенн масштабних коефіцієнтів вейвлет-перетворення на проміжках часу, де спостерігаються екстремуми кореляційної функції частоти основного тону диктора;
2. Визначення оптимальних вейвлетів для дискретного та неперервоного вейвлет-перетворення мовної інформації при вирішенні задачі розпізнавання диктора;
3. Дослідження ефективності класифікаторів різного типу при використанні вказаних ознак.

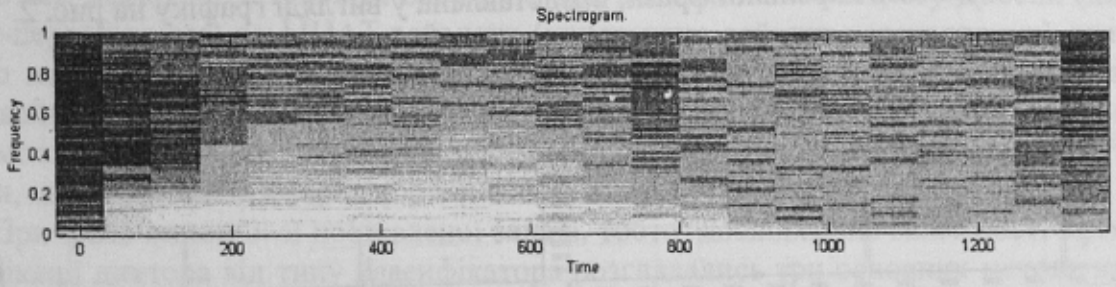
Для вирішення 1-ої поставленої задачі шляхом машинного експерименту було розроблено відповідне програмне забезпечення. Як показали дослідження, проведені в роботі [1], для формування еталонної кореляційної функції частоти основного тону достатньо одержати 5 записів парольної фрази для кожного з дикторів, які приймають участь в експерименті. В якості парольної фрази використано речення «Зима, весна, літо, осінь». В експерименті приймали участь 30 дикторів (15 чоловіків та 15 жінок) віком від 18 до 30 років. На рис. 1 наве-

дено приклад результатів, які дозволяє одержати розроблене програмне забезпечення. Воно здатне проводити вейвлет-аналіз як всього мовного повідомлення, так і довільної його частини, що дозволяє використовувати його при побудові як автоматизованих, так і експертних систем розпізнавання диктора. На рис.1,а) наведено мовний сигнал парольної фрази, промовленої диктором 1, на рис. 1,б) наведено частину цього сигналу, до якого була застосована процедура вейвлет-перетворення, на рис. 1,в), г), д) відповідно зображено спектрограму вибраної ділянки мовного повідомлення, значення коефіцієнтів дискретного та неперервного вейвлет-перетворення. Вейвлет-перетворення мовного повідомлення виконувалось при різних значеннях масштабного коефіцієнта вейвлет-функції: для дискретного вейвлет-перетворення $a = 2,4,8,16,32$, а для неперервного - $a = 1,2,\ldots,32$.

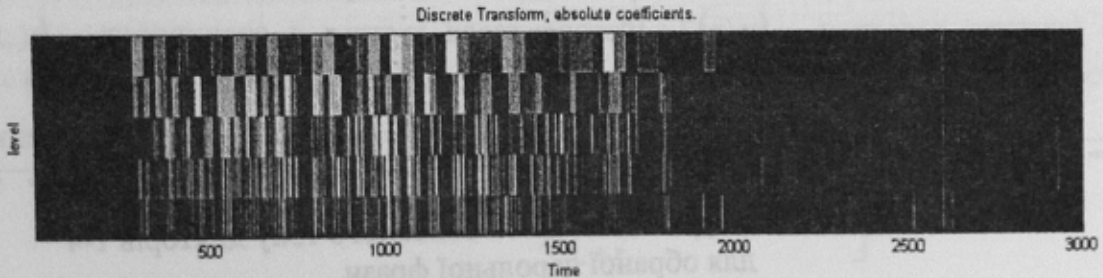


а)

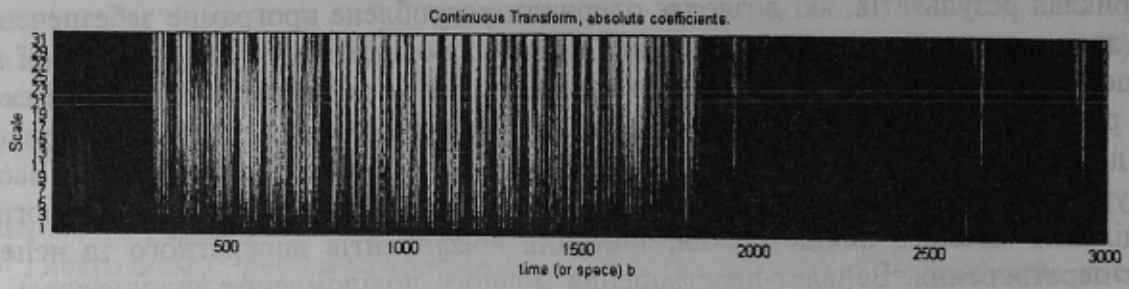
б)



в)



г)



д)

Рис. 1 – Результати виконання вейвлет-перетворення мовного сигналу :
а) вихідний мовний сигнал; б) відрізок сигналу, до якого було застосовано вейвлет-перетворення (ОХ: 1 відл = $1{,}2*10^{-4}$ с); в) спектрограма обраного відрізка мовного сигналу (ОХ: 1 відл. = $2{,}4*10^{-4}$ с); г) значення коефіцієнтів дискретного вейвлет-перетворення обраного відрізка сигналу (ОХ: 1 відл = $1{,}2*10^{-4}$ с); д) значення коефіцієнтів неперервного вейвлет-перетворення відрізка сигналу (ОХ: 1 відл = $1{,}2*10^{-4}$ с)

На рисунку 1 зміна значень коефіцієнтів вейвлет-перетвоення кореляційної функції частоти основного тону для відповідного диктора для відповідного масштабного коефіцієнту (вісь OY) в часі (вісь OX) передається інтенсивністю забарвлення (зі збільшенням значення коефіцієнтів вейвлет-перетворення колір світлішає).

Кореляційна функція, яка є мірою лінійної залежності між вибірками стаціонарних послідовностей $x(nT)$ (частота основного тону парольної фрази, що аналізується) та $y(nT)$ (еталонна частота основного тону відповідного диктора), $n = 1,\ldots,N$, в різні моменти часу обчислюється за формулою

$$R_{x,y}(mT) = \frac{1}{N-m}\sum_{n=0}^{N-m-1} x(nT)y(nT+mT).$$

Кореляційну функція частоти основного тону для диктора 1, побудована з використанням еталонних записів парольної фрази, представлена у вигляді графіку на рис. 2.

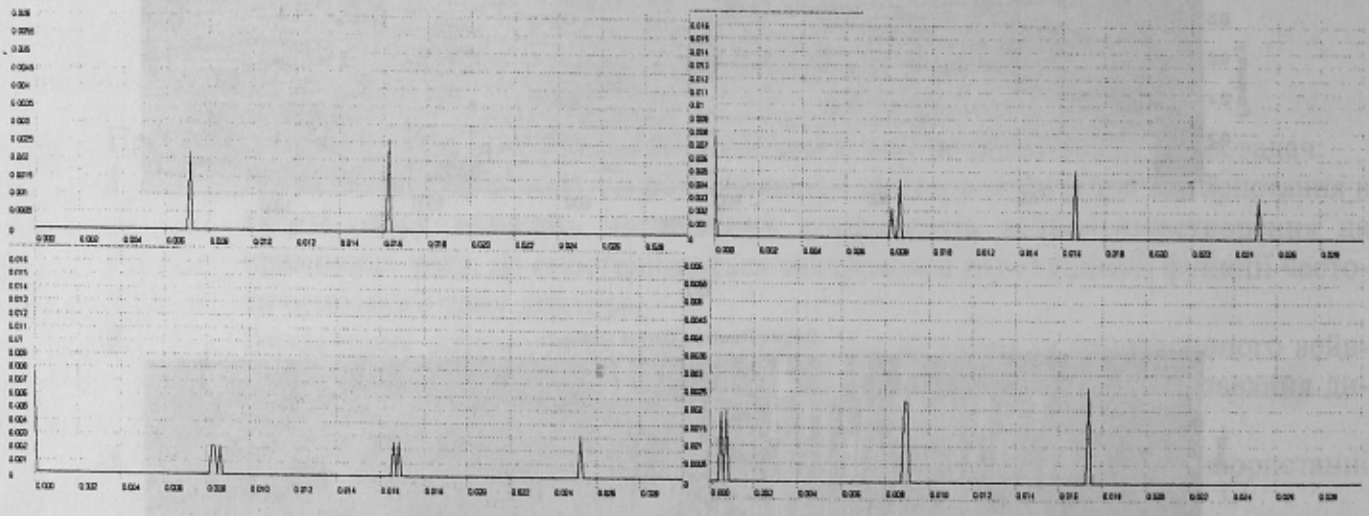


Рис. 2 – Кореляційна функція частоти основного тону дикторів 1-4 для обраної парольної фрази

Отже, в якості ознаки розпізнавання диктора використовуємо значення кореляційної функції частоти основного тону диктора та значення коефіцієнтів вейвлет-перетворення мовного сигналу на ділянках, де спостерігаються екстремуми кореляційної функції. Як свідчать

результати експерименту, використання коефіцієнтів вейвлет-перетворення в якості додаткового критерію дозволило підвищити точність розпізнавання диктора на 1,5% у порівнянні з результатом, одержаним в [1], в результаті точність склала 96,5%. Застосування аналізу мовного сигналу, який полягав в знаходженні екстремумів кореляційної функції частоти основного тону, і застосування процедури вейвлет-перетворення лише до ділянок, на яких присутні екстремуми, дозволило прискорити процедуру ідентифікації на 60-80% в залежності від характеру парольної фрази.

Розв´язання 2-ої поставленої задачі, тобто вибір оптимального вейвлет-базису, здійснювалось з використанням ентропійного критерію імовірнісного розподілу вейвлет-коефіцієнтів [4, 5]. Ентропія функції f по відношенню до вейвлет-базису відображає кількість істонтніх членів в розкладі $f = \sum_k s_{j_0,k}\varphi_{j_0,k} + \sum_{j\geq j_0,k} d_{j,k}\psi_{j,k}$, де крефіцієнти $d_{j,k}$ несуть інформацію саме про флуктуації мовного сигналу в часі, і визначається як $H(f) = \exp\left(-\sum_{j,k}\left|d_{j,k}\right|^2 \log\left|d_{j,k}\right|^2\right)$. Результати дослідження викладено в таблиці 1.

Таблиця 1. Вибір оптимального вейвлет-базису для неперервного та дискретного вейвлет-перетворення мовного сигналу по ентропійному критерію

Вид вейвлет-базису	Ентропія вихідної функції
Дискретні вейвлети	
Вейвлет Хаара	0,041
Вейвлет Добеші	0,033
Койфлети	0,037
Неперервні вейвлети	
Вейвлет Гауса	0,021
МНАТ-вейвлет	0,019
Вейвлет Морле	0,025

Отже, найкращі результати одержано при використанні вейвлету Добеші (дискретнбе вейвлет-перетворення) та МНАТ-вейвлету (неперервне вейвлет-перетворення) для обробки мовного матеріалу. Також можна стверджувати, що використовуючи коефіцієнти неперервного вейвлет-перетворення можна більш точно відтворити мовний сигнал, проте цього недоліку можна завадити за рахунок збільшення масштабного коефіцієнта. Потрібно також зауважити, що найшвидшим серед порівнюваних виявилось койфлет-перетоворення.

При розв´язанні 3-ої поставленої задачі, тобто дослідження залежності ефективності ідентифікації диктора від типу класифікатора розглядались три основних методи класифікації образів з вчителем: статистичний, по мінімуму відстані та нейромережевий [6-8].

Нехай, $W(i,j,x,y)$ - вектор з навчальної вибірки, де i - номер класу, j - номер прикладу, (x,y) - координати в векторі властивостей; $X(x,y)$ - вектор з тестової вибірки; $P(C_j,X)$ - імовірність належності X до класу C_j при обраному розподілі ймовірностей, тоді функція класифікації $f_{class}(X)$ для статистичному методу запишеться у вигляді:

$$f_{class}(X) = \arg\max_j\left[\prod_{x,y} P\left(C_j \mid X_{x,y}, \theta^j_{x,y}\right)\right],$$

де $M^j_{x,y} = \frac{1}{n_j}\sum_{i=1}^{n_j} W^j_{i,x,y}$, $\sigma^j_{x,y} = \frac{1}{n-1}\sum_{i=1}^{n}\left(M^j_{x,y} - W^j_{i,x,y}\right)^2$, $\theta_{norm} = \{M,\sigma\}$.

Функція, за якою здійснюється класифікація, при використання правила мінімуму відстані приймає такий вид:

$$f_{class}(X) = \arg\min_i \left[\min_j \sum_{x,y} \rho\left(X_{x,y}, W^i_{j,x,y}\right) \right].$$

де ρ - евклідова відстань.

Функція класифікації для нейромережевого методу має такий вигляд:

$$f_{class}(X) = \arg\min_i \left[\min_j \sum_{x,y} \rho\left(EVec\left(Ann(X)\right), Vix(X)_{j,x,y}\right) \right],$$

$$EVec(x) = \left((y_1, y_2, \ldots, y_n) \,|\, y_i = \begin{cases} 1, i = \arg\max_j (x_j), \\ 0. \end{cases} \right),$$

де $Vix(X)$ - функція виходу нейромережі.

Дослідження ефективності класифікаторів різного типу за умови використання вибраних ознак полягало в виконанні низки етапів. На першому етапі для кожного з дикторів формувалася навчальна вибірка, в яку входили еталонні записи вимови парольної фрази для кожного з дикторів і проводилося навчання класифікатора з використанням навчальних вибірок. Після цього проводилося тестування на тестовій вибірц і результати тестування якості ідентифікації дикторів класифікаторами порівнювалися з результатами ідентифікації цих дикторів тренованим експертом. І, в решті решт, за одержаними результатами розраховувалася точність розпізнавання кожного диктора окремо та середня загальна точність розпізнавання. Результати дослідження наведені в таблиці 2.

Таблиця 2. Залежність імовірності правильної ідентифікації диктора від метода класифікації та кількості еталонів для кожного з дикторів в навчальній вибірці

Метод класифікації		Кількість еталонів в навчальній вибірці		
		6	10	16
По мінімуму відстані	Сума модуля	59,2	81,8	83,8
	Макимум модуля	60,9	81,3	83,6
	Сума квадратів	61,8	80,5	84,7
Статистичний	Розподіл Гауса	63,7	64,2	69,8
	Розподіл Коші	65,9	66,0	70,3
Нейромережевий	3-шаровий персептрон	86,6	92,1	96,2

Як видно з результатів досліджень, наведених в таблиці 2, найбільш ефективним для прийняття рішень в задачі розпізнавання дикторів виявився нейромережевий класифікатор (3-х шаровий персептрон). До недоліків даного класифікатора можна віднести експоненціальну залежність часу навчання класифікатора при слабкій корельованості. З таблиці 2 видно, що збільшення обсягу тестової вибірки після деякого значення не приводить до істотних змін точності розпізнавання, і що для формування тестової вибірки достатньо 12-16 еталонних записів парольної фрази для кожного з дикторів.

Висновки

Для підвищення ефективності ідентифікації диктора в статті запропоновано використовувати в якості ознак значень коефіцієнтів вейвлет-перетворення сигналу парольної фрази на ділянках, де спостерігаються екстремуми кореляційної функції. Як показали результати експериментів, запропоноване вдосконалення дозволило підвищити точність розпізнавання диктора на 1,5% у порівнянні з результатом, одержаним в [1], в результаті точність склала 96,5%.

Також в статті проведено оцінку найпоширеніших вейвлет-базисів щодо їх використання при побудові систем розпізнавання диктора за ентропійним критерієм імовірнісного розподілу вейвлет-коефіцієнтів. Найкращі результати одержано при використанні вейвлету Добеші (дискретне вейвлет-перетворення) та MHAT-вейвлету (неперервне вейвлет-перетворення).

Проведене в статті порівняння існуючих класифікаторів стосовно їх ефективності для вирішення задачі розпізнавання диктора за запропонованою ознакою дозволяє стверджувати, що найбільш ефективним для прийняття рішень при вирішенні задачі розпізнавання диктора за запропонованими ознаками є 3-х шаровий персептрон.

Література

1. Биков М.М., Ковтун В.В. Аналіз ефективності ідентифікації диктора за частотою основного тону. Вісник Технологічного університету Поділля – Хмельницький: ХДУ - №2. - Ч.1. – Т.2. – С.20-24
2. Петухов А.П. Введение в теорию базисов всплесков. СПб.: Изд. СПбГТУ, 1999 – 131 с.
3. Воробьев В.И., Грибунин В.Г. Теория и практика вейвлет-преобразования. СПб.: ВУС, 1999, 203 с.
4. Carmona R., Hwang W-L., Torresani B. Practical Time-Frequency Analysis. San Diego: Academic Press, 1998.
5. Mayer Y. Wavelets: Algorithms and Aplications. Philadelphia: SIAM, 1993.
6. Ту Д.Т, Гонсалес Р.К. Принципы распознавания образов / пер. с англ. И.Б. Гуревич; Под ред. Р.И. Журавлева.– М.: Мир, 1978.– 411 с.
7. Hebb D. O. The Organization of Behavior: A Neuropsychological Theory.– New York: Wiley, 1949.– 358 p.
8. Минский М., Пейперт С. Персептроны.– М.: Мир, 1971.– 261 с.