

ВЕЙВЛЕТ-ПАКЕТНІ ОЗНАКИ МОВНОГО СИГНАЛУ У ЗАВДАННІ РОЗПІЗНАВАННЯ МОВЦЯ

Анотація- Запропоновано новий підхід до формування вейвлет-пакетного дерева декомпозиції мовних сигналів, що оптимізований до завдання розпізнавання мовця. Експериментально показано, що використання концепції критичних смуг та цілеспрямованого вибору базисної вейвлет функції для виділення ознак мовного сигналу дозволило зменшити рівень рівних помилок розпізнаванні мовця на 1% та 3% в порівнянні з використанням мел-частотних кепстральних коефіцієнтів та вейвлет-пакетних ознак, запропонованих Сарікаєм, Фаруком та Даттою відповідно.

Ключові слова: ознаки мовного сигналу, вейвлет-пакетне перетворення, критична смуга, розпізнавання мовця.

O.O. KORNIIENKO

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

WAVELET AND PACKAGE FEATURES OF VOICE SIGNAL CONCERNING RECOGNITION SPEAKER PROBLEM

Abstract- Feature extraction is an important step of all speech processing systems. Historically, the short-term spectral parameters have dominated as the features of the speech and speaker recognition tasks. These features don't allow to represent the quasi-stationary speech segments such as vowel and consonant sounds, speech-nonspeech or interphonemes shifts. Therefore, we use wavelet-packet transform to address the shortcomings of short-term spectral features in this paper. We created the algorithm of adaptation the wavelet-packet tree to critical bands which were described by the psychoacoustics model – equivalent rectangular bandwidth (ERB). The experimental results confirm that ERB-based speech features outperform Mel Frequency Cepstral Coefficients and sub-band wavelet-packet features, which were adapted to well-known Mel-scale (proposed by Sarikaya) or Bark-scale (proposed by Farooq and Datta) during the speaker verification task. Relative reduction of the equal error rate by at least 1% was observed for the proposed speech features, using i-vector speaker recognition approach. We have achieved EER 10,7% for the proposed features using Daubechies 3 wavelet and all discrete cosine coefficients. It was shown that the using detailed wavelet-packet speech decomposition tree increase total error equal rate by at least 2%. This means that the speaker dependent information preferably located in 0-1 kHz frequency band. The experimental results showed that using the teager-kaiser energy operator for calculating sub-band cepstral coefficients decreases error equal rate instead of traditional energy operator in the speaker recognition task. It was confirmed Farooq's assumption that the increasing total count of discrete cosine components, which represents 90% and 100% of total signal energy, slightly decrease error equal rate by less than 1%.

Keywords: speech features, wavelet-packet decomposition, critical band, speaker recognition.

1 Вступ. Проблема параметризації мовного сигналу у задачі розпізнавання мови й аудіоавтентифікації мовця досі залишається відкритою. Сучасні методи розпізнавання складаються із стадії розрахунку ознак мовного сигналу, призначеної для виділення характеристик голосу мовця, та стадії класифікації, тобто оцінки приналежності вектора ознак до одного з класів моделей голосу мовців. Точність класифікації залежить від типу ознак.

Кепстральні коефіцієнти лінійного передбачення (LPCC, Linear Predictive Cepstral Coefficients) [1] та мел-частотні кепстральні коефіцієнти (MFCC, Mel Frequency Cepstral Coefficient) [2] найчастіше використовуються як ознаки мовного сигналу у задачах розпізнавання мови та мовця. Ці параметри є непридатними для описання локальних властивостей квазістаціонарного мовного сигналу, тому що розраховуються з використанням швидкого перетворення Фур'є [3]. Тому представляє інтерес використання вейвлет-перетворення для виділення ознак мовного сигналу [4].

Відомо три підходи до розрахунку параметрів мовних сигналів з використанням вейвлет-перетворення. Перший підхід полягає в декореляції спектральних ознак замість використання дискретного косинус перетворення (DCT, Discrete Cosine Transform) [5]. Другий підхід полягає в розрахунку високоенергетичних вейвлет-коефіцієнтів як ознак мовного сигналу [4]. Однак такі параметри не є стійкими до нестационарної завади. У третьому підході вейвлет-перетворення використовується для розрахунку енергії частотних смуг сигналу замість застосування мел-частотного банку фільтрів [6]. Зокрема, при розпізнаванні мовця, вейвлет-пакетне перетворення застосовується для обчислення спектра сигналу [7] та як альтернатива ознакам MFCC [6, 8-10]. Структуру вейвлет-пакетного дерева декомпозиції мовних сигналів при розрахунку таких ознак адаптують до шкали мелів або барків [11] та застосовують для обробки сигналів з частотою дискретизації 8 кГц. Це обмежує придатність таких ознак до застосування у сучасних системах обробки мовних сигналів. Отже, актуальною задачею є розробка нового алгоритму апроксимації вейвлет-пакетного дерева декомпозиції [11].

У цій роботі запропоновано новий спосіб формування вейвлет-пакетного дерева декомпозиції мовного сигналу та відповідну структуру дерева декомпозиції.

2 Концепція критичних смуг

Протягом останніх десятиліть досягнуто значних результатів у галузі дослідження слухової системи людини [12, 13, 14]. Периферична слухова система поводить себе так, ніби вона складається з набору

смугових фільтрів з перекриттям смуг пропускання. Частотна вибірковість слухової системи і характеристики відповідних слухових фільтрів можуть бути досліджені шляхом проведення експериментів, заснованих на техніці маскувння. Маскувальний ефект полягає у зменшенні чутливості слухового апарату людини до спектрального складу звуку у межах ширини смуги одного слухового фільтра. Така смуга називається критичною [12].

Відомі спроби оцінки залежності ширини полоси пропускання від центральної частоти слухового фільтра [15]. Розраховано, що критична смуга пропускання є сталою і приблизно рівна 100 Гц для частот діапазону нижче 500 Гц, у той час для більш високих частот вона збільшується пропорційно центральній частоті фільтра [15]. Використовуючи це наближення, у [15] запропоновано міри сприйняття висоти тону - шкалу барків та шкалу мелів.

Поняття еквівалентних прямокутних смуг пропускання (ERB, Equivalent Rectangular Bandwidth) було введено у [13] для визначення форми АЧХ слухових фільтрів й оцінки величини критичних смуг. У [13] показано, що існують протиріччя у висновках [15], зокрема, на частотах нижче 500 Гц ширина критичної смуги продовжує зменшуватися.

Згідно з [13], для будь-якого слухового фільтра існує такий ідеальний фільтр з прямокутною АЧХ, який пропускає білий шум тієї ж потужності, що й розглянутий фільтр. ERB – це ширина полоси пропускання такого ідеального фільтра. Ширина смуги ERB в залежності від центральної частоти діапазону [16] описується співвідношенням:

$$ERB(f) = 24.7 \cdot (0.00437 \cdot f + 1) \quad (1)$$

де значення ширини смуги ERB і центральної частоти смуги f вимірюється в Гц.

Припустімо, що використання концепції критичних смуг ERB при формуванні ознак мовного сигналу дозволить підвищити ефективність систем розпізнавання мовця.

3 Вейвлет-пакетний аналіз мовних сигналів

3.1 Алгоритм адаптації вейвлет-пакетного дерева декомпозиції

Дискретне вейвлет-пакетне перетворення є узагальненням класичного дискретного вейвлет-перетворення та виконується шляхом каскадного застосування пари низько- та високочастотних фільтрів [4].

Структуру вейвлет-пакетного дерева декомпозиції сигналу зображають бінарним деревом [17], де

вузол відповідає вейвлет-пакетним коефіцієнтам розкладу частотного піддіапазону, рух по лівій гілці – розрахунку низькочастотних компонент сигналу, по правій – високочастотних. Шлях по дереву до вузла може бути кодованим бінарною послідовністю довжини L (глибина декомпозиції сигналу). Біт зі значенням 0 послідовності відповідає низькочастотному розкладанню сигналу, 1 - високочастотному. Позиція біту визначає рівень розкладання сигналу. Наприклад, старший біт визначає тип фільтра, що застосовується на рівні декомпозиції 0, молодший - на максимальному рівні L_{max} . Спектр вхідного сигналу з частотою дискретизації f_s розділяється на $2^{L_{max}+1}$ частотні піддіапазони для рівня розкладання L_{max} з шириною смуги пропускання $\frac{f_s}{2^{L_{max}+2}}$. Центральна частота піддіапазону такого вейвлет-пакетного дерева декомпозиції розраховується:

$$f_i = (i + 0,5) \cdot \frac{f_s}{2^{L_{max}+2}} \quad (2)$$

де i – індекс вузла піддіапазону дерева декомпозиції, $i \in [0, 2^{L+1}]$, L – номер рівня декомпозиції $L \in [0, L_{max}]$, f_s - частота дискретизації сигналу.

На рис. 1 зображена блок-схема алгоритму формування структури вейвлет-пакетного дерева декомпозиції мовного сигналу, що апроксимує функцію $B(f)$ залежності ширини критичної смуги психоакустичної моделі від її центральної частоти, наприклад, вираз (1) [18].

Вхідними параметрами алгоритму є:

максимальна кількість рівнів розкладання L_{max} , частота дискретизації сигналу f_s та функція залежності ширини критичної смуги від центральної частоти $B(f)$. На першому кроці алгоритму формується індекс вузла піддіапазону вейвлет-пакетного дерева декомпозиції. Далі розраховується центральна частота

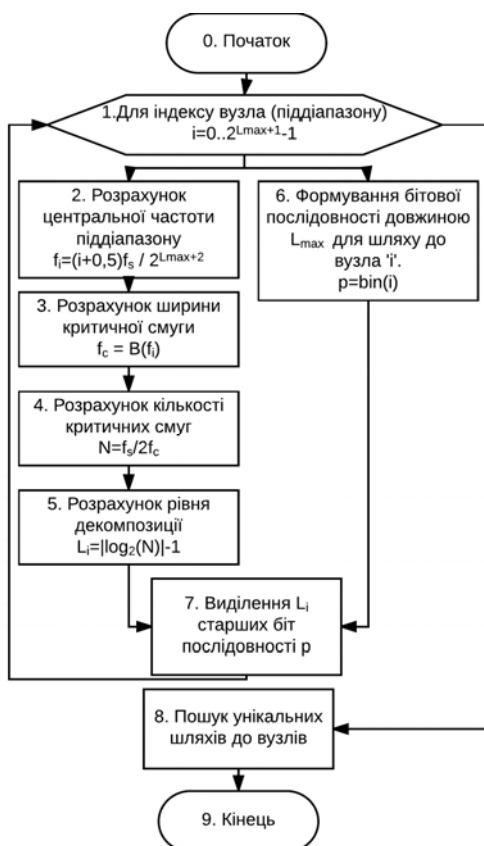


Рис. 1. Блок-схема алгоритму формування структури вейвлет-пакетного дерева декомпозиції сигналу

піддіапазону за виразом (2). На третьому кроці розраховується ширина критичної смуги за виразом (1). Для розрахунку частотної роздільної здатності вейвлет-пакетного перетворення, що відповідає ширині критичної смуги, на четвертому та п'ятому кроці розраховуються максимальна кількість критичних смуг та відповідний рівень декомпозиції. На шостому кроці формується бінарне представлення індексу вузла i довжиною L_{max} . На сьомому кроці виділяються L_i старших біт, що були розраховані на четвертому кроці. На восьмому кроці вибираються унікальні бінарні послідовності шляхів серед сформованих на сьомому кроці.

Результатом роботи алгоритму є масив унікальних бінарних послідовностей шляхів до вузлів піддіапазонів вейвлет-пакетного дерева декомпозиції мовного сигналу, що апроксимують психоакустичну модель $B(f)$. Алгоритм може бути використаний для формування вейвлет-пакетного дерева декомпозиції мовного сигналу у програмних пакетах PyWavelets, Matlab Wavelet Packet Toolbox, Matlab Uvi_Wave Toolbox.

3.2 Структура вейвлет пакетного дерева декомпозиції мовного сигналу

Згідно з концепцією критичних смуг, у [19] сформовано рекомендації щодо створення структури вейвлет-пакетного дерева декомпозиції, а саме для сигналу оцифрованого з частотою дискретизації $f_s=16$ кГц найбільша частотна роздільна здатність 31.25 Гц забезпечується у діапазоні [0, 1000] Гц, 62.5 Гц у діапазоні [1000, 2500], 125 Гц у діапазоні [2500, 8000] Гц. Для перевірки правильності цих рекомендацій відносно параметрів мовного сигналу, розглянемо дерево декомпозиції з подвоєною шириною полоси діапазонів, тобто роздільна здатність дерева становить 31.25 Гц у діапазоні [0, 2000] Гц, 62.5 Гц у діапазоні [2000, 5000] Гц і 125 Гц у діапазоні [5000, 8000] Гц.

Для визначення структури дерева декомпозиції сигналу, при якій забезпечується найбільша точність розпізнавання мовця (оптимальної) у роботі розглянуто апроксимації таких частотно-деформаційних шкал:

- Апроксимація мел-частотної шкали вейвлет-пакетним деревом декомпозиції, запропонованим у [8] (MWPT, Mel-scale wavelet packet tree).
- Апроксимація барк-частотної шкали вейвлет-пакетним деревом, запропонованим у [20] та [21] (BWPT, Bark-scale wavelet packet tree).
- Апроксимація еквівалентної прямокутної смуги пропускання (1) (EWPT-2), запропонована у [11].
- Апроксимація еквівалентної прямокутної смуги пропускання (1) (EWPT-3), запропонована у [19].
- Апроксимація еквівалентних прямокутних смуг пропускання (CWPT, Cascade-scale wavelet packet tree) з подвоєною шириною частотних діапазонів [11].
- Створена нами апроксимація еквівалентних прямокутних смуг пропускання (EWPT-1, ERB-scale wavelet packet tree), описана виразом (1).

На рисунку 2 представлено залежність частотної роздільної здатності від центральної частоти піддіапазону для розглянутих апроксимацій критичних смуг вейвлет-пакетним перетворенням.

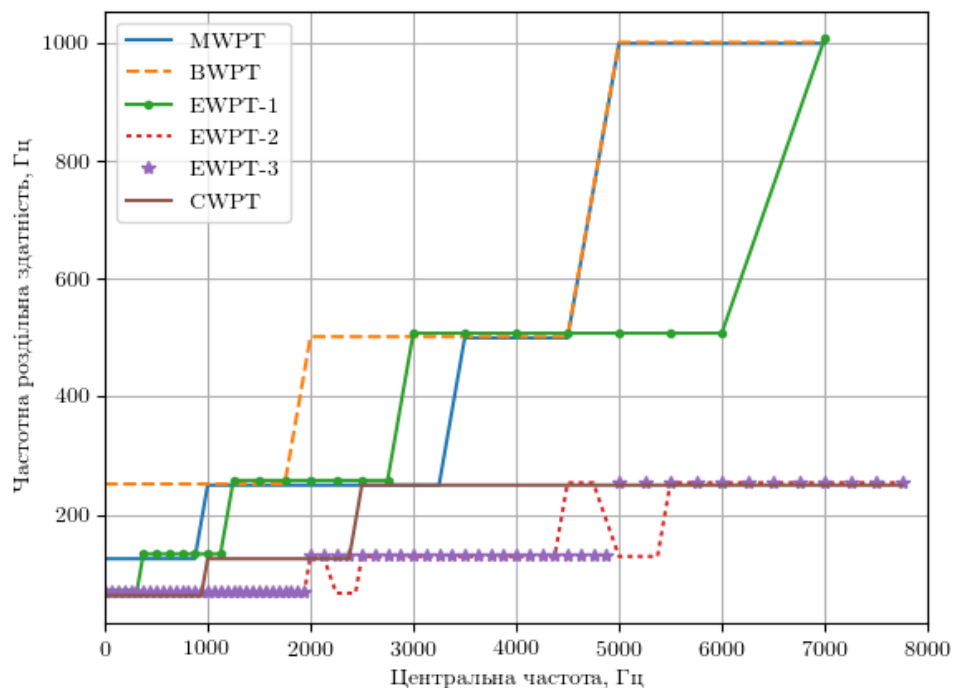


Рис. 2. Залежність частотної роздільної здатності від центральної частоти вейвлет-пакетного піддіапазону

Використано вейвлет-пакетні кепстральні коефіцієнти (SBCC, Subband cepstral coefficients) [8] як ознаки мовного сигналу:

$$SBCC_n = \sum_{j=1}^L \log \left[\frac{1}{N_j} \sum_{k=1}^{N_j} E(\omega_j(k)) \right] \cos \left(\pi \frac{n(j-0.5)}{L} \right) \quad (2)$$

де $\omega_j(k)$ визначає k -й коефіцієнт j -ого піддіапазону вейвлет-пакетного перетворення, $E(\cdot)$ – оператор розрахунку енергії, $j = 1..L$, L – кількість піддіапазонів, $k = 1, 2, \dots, N_j$ – номер коефіцієнта j -ого піддіапазону, N_j – кількість коефіцієнтів j -ого піддіапазону, $n = 1..L$ – номер коефіцієнта DCT.

Згідно з [20, 22] у роботі випробувано два варіанти розрахунку енергії піддіапазону: традиційний та оператор Тігера-Кайзера (TKE, Teager Kaiser Energy).

4 Експериментальні результати

Задля отримання результатів оцінки ефективності ознак мовних сигналів у завданні розпізнавання мовця використано 6561 записи голосу 31 мовця загальною тривалістю 694.4 хв з корпусу VoxForge. Частота дискретизації мовного сигналу становить 16 кГц, тривалість фреймів 32 мс з перекриттям 16 мс. Віконною функцією обрано функцію Хеммінга. Базисними вейвлет-функціями обрано вейвлети родини Добеші (db), Мейєра (dmeu), коіфлет (coif) та біортогональної (bior). Модель мовця створювалась сучасним методом *i-vector* [23] з програмного пакету BOB [24] з параметрами: розмірність “вектору ідентичності” (identity vector, *i-vector*) становить 100, кількість компонент гаусових сумішей розподілів – 256, класифікатором обрано алгоритм PLDA з розмірністю векторів лінійних моделей 50.

Оцінка ефективності ознак мовного сигналу проводилась шляхом зрівняння точності розпізнавання мовця на основі рівня рівних помилок (EER, Error Equal Rate). EER відповідає точці перетину залежностей помилок 1-ого та 2-ого роду від порогу вирішувального пристрою.

Помилка EER для методу ідентифікації мовця з використанням ознак MFCC з 24 мел-частотними фільтрами становить 11.7%.

Графік залежності помилки EER від типу вейвлет-пакетного дерева декомпозиції для базисної функції Добеші 3 представлено на рисунку 3. Застосування концепції критичних смуг ERB, при формуванні ознак, дозволило зменшити EER на 1% порівняно з ознаками BWPT та MWPT. Ми вважаємо, що успіх впровадження таких дерев декомпозиції сигналу полягає у забезпеченні високої частотної роздільної здатності у діапазоні 0-1кГц. Цей діапазон містить інформацію про голос мовця – частоту основного тону та резонансні частоти голосового тракту. З іншого боку, ознаки, сформовані за допомогою розширеного дерева CWPT (див. рис. 3) не показали найкращих результатів так як містять лінгвістичну інформацію.

З рисунка 3 видно, що мінімальні помилки EER для розроблених ознак EWPT- 1,2,3 досягнуто при використанні базисних вейвлет функцій родин біортогональної 1.5 і Добеші 3, 4 та 7 (рис. 4). Найменша помилка EER 10,7% отримана для ознак EWPT-1 з використанням вейвлету родини Добеші 3. Це пояснюється частотно-часовими особливостями вибіркості вказаних базисних вейвлет-функцій.

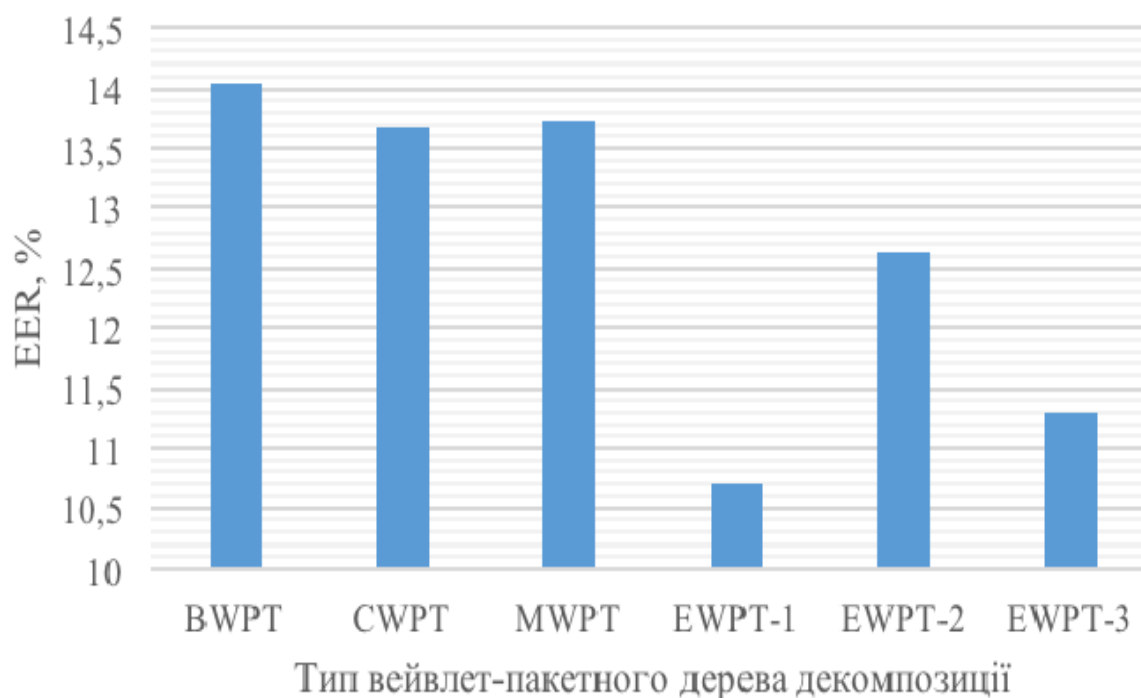


Рис. 3. Залежність рівня рівних помилок EER розпізнавання мовця від типу вейвлет-пакетного дерева декомпозиції

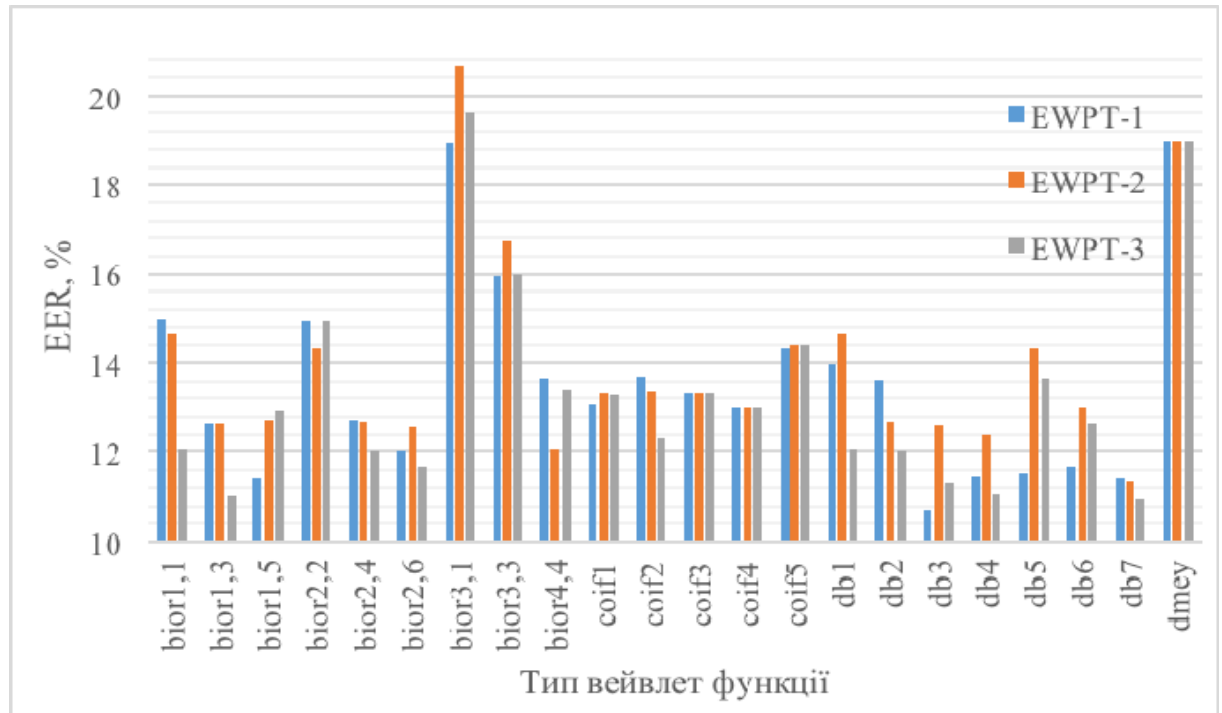


Рис. 4. Залежність рівня рівних помилок EER розпізнавання мовця від типу материнської вейвлет функції

На рис. 5 зображено гістограми залежностей помилок EER від типу ознак, сформованих із використанням традиційного оператора розрахунку енергії та оператора Тігера-Кайзера. Виявлено, що використання оператора ТКЕ зменшує точність системи розпізнавання мовця.

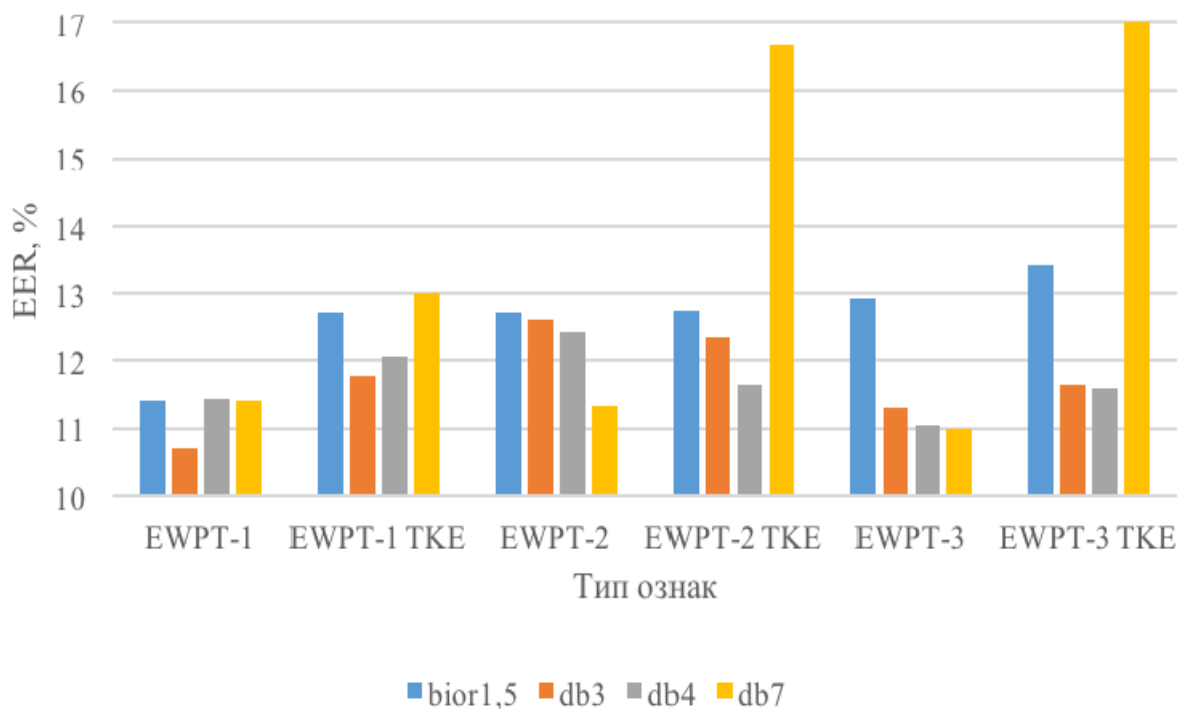


Рис. 5. Залежність рівня рівних помилок EER розпізнавання мовця від типу ознак мовного сигналу

Залежність помилки EER від кількості компонент DCT представлено на рис. 6. Оцінка точності розпізнавання мовця проведена для ознак, розрахованих без використання DCT та для ознак з кількістю компонент, що відповідають 70-100% (таблиця 1) енергії мовного сигналу (див. таблицю 1).

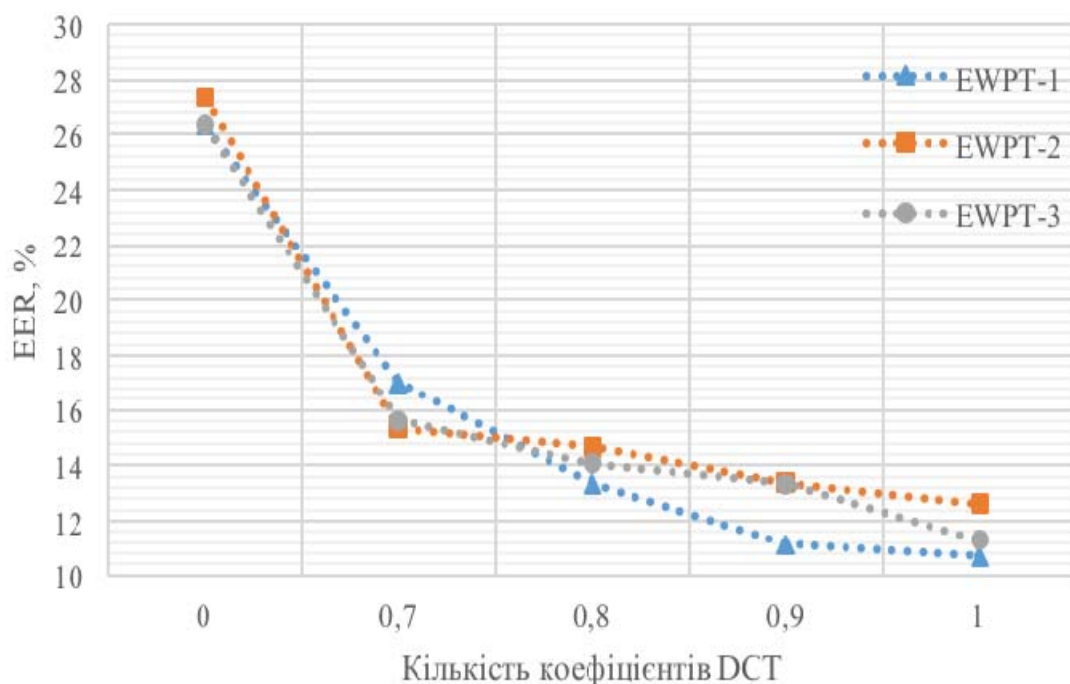


Рис. 6. Залежність рівня рівних помилок EER розпізнавання мовця від кількості компонент DCT

Виявлено, при збільшенні кількості декорельованих ознак мовного сигналу, що відповідають 90% та 100%, спостерігається незначний приріст точності розпізнавання мовця, що підтверджується висновками з [8, 19].

Таблиця 1

Залежність кількості DCT компонент мовного сигналу від нормованої енергії мовного сигналу для різних типів дерев декомпозиції

Тип дерева	0	0.7	0.8	0.9	1
EWPT-1	18	8	12	16	18
EWPT-2	35	7	12	24	35
EWPT-3	39	7	14	25	39

Розроблені ознаки показали схожу точність розпізнавання мовця (рис. 6), при цьому кількість коефіцієнтів DCT запропонованого дерева EWPT-1 є меншою (таб. 1).

5 Висновок

Запропоновано загальний метод формування дерева декомпозиції для виділення ознак мовного сигналу у завданні розпізнавання мовця. Експериментально встановлено:

1. Використання концепції критичних смуг, що описуються еквівалентними смугами пропускання (ERB) зменшує помилку EER розпізнавання мовця щонайменше на 1% порівнянно з вейвлет-пакетними ознаками MWPT, BWPT та загальновідомими ознаками MFCC.

2. Найменша помилка EER 10,7% отримана для запропонованих ознак EWPT-1 з використанням вейвлету родини Добеші 3 для усіх коефіцієнтів дискретного косинус перетворення.

3. Застосування дерева декомпозиції мовного сигналу з подвоєною шириною частотних діапазонів, побудованих з використанням концепції ERB пропускання не дозволило підвищити точність розпізнавання мовця.

4. Використання оператора ТКЕ при розрахунку енергії вейвлет-пакетних піддіапазонів зменшує точність системи розпізнавання мовця в порівнянні із застосуванням звичайного оператора.

5. При збільшенні кількості декорельованих ознак мовного сигналу, що відповідають 90% та 100% спостерігається незначний приріст (не більше 1%) точності системи розпізнавання мовця.

Представляє інтерес розробки нового методу автоматичного формування вейвлет-пакетного дерева декомпозиції сигналу з використанням генетичного алгоритму та згорткових нейронних мереж для виділення найбільш інформативних ознак замість застосування концепції критичних смуг. Запропонований алгоритм формування вейвлет-пакетних ознак мовного сигналу та відповідна структура дерева декомпозиції можуть бути використані для підвищення ефективності систем обробки мовних сигналів зокрема для аудіоавтентифікації мовця.

Література

1. Atal B. S. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification / B. S. Atal // *Journal of the Acoustical Society of America*. – 1974. – Vol. 55(6). – P.1304–1312.
2. Davis S. B. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences / S. B. Davis, P. Mermelstein // *IEEE Transactions on Acoustic, Speech and Signal Processing*. – 1980. – Vol. 28(4). – P. 357–366.
3. Ziolk B. Wavelet method of speech segmentation / B. Ziolk, S. Manandhar, R. Wilson // *Proceedings of 14th European Signal Processing Conference EUSIPCO*. – 2006.
4. Long C.J. Wavelet based feature extraction for phoneme recognition / C.J. Long, S. Datta // *Proc. of 4th Int. Conf. of Spoken Language Processing*. – Philadelphia, USA. – 1996. – Vol. 1. – P. 264–267.
5. Tufekci Z., Feature extraction using discrete wavelet transform for speech recognition / Z. Tufekci, J. N. Gowdy, // *In Proceedings of the IEEE SoutheastCon 2000*. — 2000. — P. 116–123.
6. Sarikaya R. High resolution speech feature parameterization for monophone-based stressed speech recognition / R. Sarikaya, H. L. Hansen // *IEEE Signal Processing Letters*. – 2000. – Vol. 7(7). – P. 182–185.
7. Erzin E. Subband analysis for speech recognition in the presence of car noise / E. Erzin, A. E. Cetin, Y. Yardimci // *In Proceedings of the IEEE international conference on acoustics, speech, and signal processing (ICASSP-95)*. – 1995. –Vol. 1. – P. 417–420.
8. Sarikaya R. Wavelet packet transform features with application to speaker identification / R. Sarikaya, L. Bryan, B.L. Pellom, H.L Hansen // *Proc. of IEEE Nordic Signal Processing Symp*. – 1998. – P. 81-84.
9. Farooq O. Mel filter-like admissible wavelet packet structure for speech recognition / O.Farooq, S. Datta // *IEEE Signal Processing Letters*. – 2001. – Vol. 8(7). – P. 196–198.
10. Farooq O. Mel-scaled wavelet filter based features for noisy unvoiced phoneme recognition / O.Farooq, S. Datta // *In Proceedings of the 7th international conference on spoken language processing (ICSLP 2002)*. – 2002. – P. 1017–1020).
11. Siafarikas M. Wavelet packet approximation of critical bands for speaker verification / M. Siafarikas, T. Ganchev, N. Fakotakis // *International Journal of Speech Technology*. - 2007. - Vol. 10, Issue 4. - pp. 197-218.
12. Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, 12, 47–65.
13. Moore B. J. An introduction to the psychology of hearing / B. C. J. Moore. – San Diego: Academic Press, 2003. – (5).
14. Zwicker E. Subdivision of the audible frequency range into critical bands / E. Zwicker // *Journal of the Acoustical Society of America*. – 1961. – Vol. 33. – P. 248–249.
15. Zwicker E. Analytical expressions for critical band rate and critical bandwidth as a function of frequency / E. Zwicker, E. Terhardt // *The Journal of the Acoustical Society of America*. – 1980. – Vol. 68, P. 1523–1525.
16. Glasberg B.R. Derivation of auditory filter shapes from notched-noise data / B.R. Glasberg, B.C.J. Moore // *Hearing Research*. – 1990. – Vol. 47, Issues 1-2. – P. 103-138.
17. Cody M.A. The Wavelet Packet Transform / M.A. Cody // *Dr. Dobb's Journal*. – 1994. – Vol 19. – P. 44-46, 50-54.
18. Корнієнко О.О. Алгоритм апроксимації вейвлет-пакетного дерева декомпозиції / О.О. Корнієнко // *19-th International Conference SAIT 2017*. – Ukraine, Kyiv. – 2017. – С. 289.
19. Siafarikas M., Ganchev T. Wavelet packet approximation of critical bands for speaker verification / M. Siafarikas, T. Ganchev, N. Fakotakis, G. Kokkinakis // *International Journal of Speech Technology*. – 2006.
20. Huan Zhao. An Improved Speech Enhancement Method based on Teager Energy Operator and Perceptual Wavelet Packet Decomposition / Huan Zhao; Xiujuan Peng; Lian Hu; Gangjin Wang; Fei Yu; Chen Xu // *Journal of Multimedia*. – 2011. – Vol. 6 Issue 3. – P. 308-315.
21. Deshpande M.S. Speaker Identification Using Admissible Wavelet Packet Based Decomposition / M.S. Deshpande, R.S. Holambe // *World Academy of Science, Engineering and Technology*. – 2010. – Vol. 10. – P. 83-86
22. Quatieri T.F. Energy Onset Times for Speaker Identification / T. F. Quatieri, C. R. Jankowski, D. A. Reynolds // *IEEE Signal Process. Lett*. – 1994. – Vol.1(11). – P.160-162.
23. Kenny P. PLDA for speaker verification with utterances of arbitrary duration / P. Kenny, T. Stafylakis, P. Ouellet, M. J. Alam, P. Dumouchel // *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. – Vancouver, Canada. – 2013.
24. Khoury E. Spear: An open source toolbox for speaker recognition based on Bob / E. Khoury, L. El Shafey, S. Marcel // *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. – 2014.

Рецензія/Peer review : 13.4.2017 р.

Надрукована/Printed :23.6.2017 р.

Стаття рецензована редакційною колегією