
СИСТЕМИ ТЕХНІЧНОГО ЗОРУ І ШТУЧНОГО ІНТЕЛЕКТУ З ОБРОБКОЮ ТА РОЗПІЗНАВАННЯМ ЗОБРАЖЕНЬ

УДК 681.327.12

М. М. БИКОВ, В. В. КОВТУН

ПІДВИЩЕННЯ ІНФОРМАТИВНОСТІ ОСНОВНОГО ТОНУ ДЛЯ РОЗПІЗНАВАННЯ МОВЦІВ ЗГОРТАЛЬНИМИ НЕЙРОМЕРЕЖАМИ

Вінницький національний технічний університет

21021, вул. Хмельницьке шосе, 95, м. Вінниця, Україна, E-mail: kovtun_v_v@vntu.edu.ua

Анотація. У статті автори розробили метод і алгоритм виділення основного тону мовця із мовного сигналу, який використовує дані вейвлет-перетворення мовного сигналу та його сингулярного аналізу, що дозволяє довільно масштабувати представлення сигналу у частотній області і підвищує достовірність виділення основного тону у випадку присутності шумів у мовному сигналі і використанні для розпізнавання мовця згортальної нейромережі.

Аннотация. В статье авторы представили метод и алгоритм выделения основного тона диктора из речевого сигнала, использующий данные вейвлет-преобразования речевого сигнала и его сингулярного анализа, что позволяет произвольно масштабировать представление сигнала в частотной области и повышает достоверность выделения основного тона в случае присутствия шумов в речевом сигнале и использовании для распознавания диктора сверточной нейросети.

Abstract. In the article the method and algorithm were presented for speaker's pitch extraction from the speech signal, using the data of the wavelet transform and singular analysis of the speech signal. That allows arbitrary scale representation of the signal in the frequency domain and increases reliability allocation pitch in the case of noise present in the speech signal and using a convolution neural network for speaker recognition.

Ключові слова: автоматизована система розпізнавання мовців критичного застосування, розпізнавання образів, цифрова обробка сигналів, сингулярний аналіз, згортальна нейромережа, вейвлет-перетворення.

ВСТУП

Генерація мовних сигналів людиною є складним процесом, в якому одним з основних елементів є голосовий джерело (голосові низки). Результатом його діяльності є пульсуючий повітряний потік, особливістю якого є повторення імпульсів голосового джерела із певною частотою, яку називають частотою (періодом) основного тону. Основний тон є одним з основних параметрів мовного сигналу і може використовуватися для розпізнавання мовців [1]. Слід зауважити, що інформаційна складова основного тону виходить за рамки задачі розпізнавання мовців, наприклад, інформація про траєкторію зміни мелодії частоти основного тону описує інтонацію мовця і може бути оцінкою його емоційного стану. Дослідження показали [2], що частотний діапазон основного тону для чоловіків і жінок різний і становить в середньому 80-200 Гц для чоловіків і 150-320 Гц для жінок, тоді як частотний діапазон мовлення становить 300-4000 Гц, а розбірливою мова залишається при звуженні частотного діапазону до 300-2400 Гц, але втрачає індивідуальність. Таким чином, основний тон присутній в області низьких частот, що робить цю важливу ознаку чутливою до низькочастотних шумів, що ускладнює його екстракцію із зачумлених мовних сигналів.

Задача автоматизованого виділення основного тону має тривалу історію, що робить доцільним узагальнений аналіз існуючих методів його виділення. Актуальна класифікація методів виділення осикового тону включає методи амплітудної селекції, лінійного прогнозу, кореляційні методи, методи частотної або спектральної селекції і методи кепстрального аналізу. Всі ці методи базуються на аналізі мовного сигналу в частотному (спектральному) та часовий діапазонах. Спектральні методи аналізують

енергетичні піки на різних частотах і їх співставлення, а часові – виявляють особливості опису мовного сигналу методами кореляційного аналізу.

У методах амплітудної селекції [1, 2] обчислюється відстань між глобальними максимумами сигналу, яку приймають за період основного тону, при цьому ефективність методів здебільшого визначається значенням порогу ідентифікації максимуму, який повинен бути досить високий, щоб відсіяти локальні максимуми і шуми, але не настільки, щоб пропустити глобальний максимум. Ефективність методів підвищують збільшуючи кількість каналів, але амплітудні методи є залишаються найменш точними та чутливими до впливу шумів.

Кореляційні методи [2] виділення періоду основного тону засновані на оцінюванні середнього значення періоду пульсацій квазіперіодичної кореляційної функції, знаходження глобальних максимумів якої набагато простіше, порівняно з методами амплітудної селекції. Втім, кореляційним методам властива нестійка робота при аналізі модульованих по амплітуді сигналів, що характерно для емоційно забарвлених мовних сигналів.

Вокалізовані звуки відповідають пікам в спектрі мовного сигналу на частотах, кратних частоті основного тону – цей факт лежить в основі спектральних методів виділення основного тону [2]. Якщо виконати дискретне перетворення Фур'є з досить малим кроком дискретизації по частоті, то в якості оцінки частоти основного тону можна використовувати частоту, на якій спостерігається максимальне значення енергії спектру. Втім, часто виникає ситуація, коли в цій же частотній смузі присутня і друга гармоніка основного тону, іноді навіть з більшою енергією, яка і буде помилково прийнята за оцінку основного тону.

Кепстральні методи [2] виділення основного тону базуються на обчисленні та аналізі кепстру – зворотного перетворення Фур'є комплексного логарифма спектра потужності мовного сигналу. Як і в методах спектрального аналізу, мовної сигнал розбивається на фрагменти-фрейми. Максимум кепстра із тривалістю від 2 мс до 20 мс дуже точно вказує, що даний фрейм є вокалізованим, а положення максимуму визначає період основного тону. Втім, як і для амплітудної селекції точність кепстральних методів визначається значенням порогової функції для ідентифікації максимуму.

Методи лінійного передбачення [2] засновані на описі мовного сигналу лінійним поліномом, що моделює мовної сигнал, пов'язаний з сигналом збудження. Таким чином, для ефективної роботи методу необхідно підібрати порядок полінома, який достатньо адекватно описуватиме спектр мовного сигналу. При наближенні частоти спектра до частоти основного тону буде спостерігатися все більше його вирівнювання. Складність полягає у визначенні ступеня рівномірності спектра зі збільшенням частоти сигналу, що призводить до падіння точності ідентифікації основного тону методом лінійного прогнозу при зростанні частоти сигналу понад 200 Гц.

ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

Отже, підводячи підсумок проведеному аналізу інформаційних джерел, можна зробити висновок, що існуючі методи екстрагування основного тону не достатньо ефективно працюють за присутності у мовному сигналі шумів і у більшості випадків помилкова ідентифікація піків сигналу як основний тон відбувається через недостатню деталізованість застосовуваних засобів аналізу сигналів, через що втрачається інформація про індивідуальні особливості сигналу. Таким чином необхідно розробити метод аналізу мовного сигналу у частотній області, який би дозволяв розглядати мовний сигнал у довільному масштабі і забезпечував достатню графічну репрезентативність для застосування загортальної нейромережі для розпізнавання мовців.

МЕТОД ВИДІЛЕННЯ ОСНОВНОГО ТОНУ ДЛЯ АВТОМАТИЗОВАНОЇ СИСТЕМИ РОЗПІЗНАВАННЯ МОВЦІВ КРИТИЧНОГО ЗАСТОСУВАННЯ

Сформулюємо математичну модель екстракції частоти основного тону із дискретизованого мовного сигналу S_N , де N – довжина файлу запису, який представимо у вигляді елементарного спектра часових рядів (спектр квазігармонік) $T_{L,N}$, які генеруються кожним резонатором мовного тракту

$$S_N = \sum_{i=1}^L T_{i,N}, L < N. \quad (1)$$

Для отримання спектра мовного сигналу об'єднаємо переваги вейвлет-аналізу [3] (Wavelet analysis) і сингулярного спектрального аналізу [4] (Singular spectrum analysis). Аналіз Фур'є [2] використовує фіксований тригонометричний базис для опису сигналу, що обмежує роздільну здатність одержуваного з його допомогою спектра в частотному і часовому вимірах. Сингулярний спектральний аналіз використовує адаптивний базис, створений із самого аналізованого ряду, що дозволяє

збільшити частотну роздільну здатність одержуваного з його допомогою спектра змінюючи довжину «гусени» SSA, а вейвлет-базис, в свою чергу, практично не обмежений у роздільній здатності у часовому вимірі сигналу.

Неперервне вейвлет-перетворення мовного сигналу S_N можна описати як

$$WLT(a, b) = \sqrt{a\Delta t} \sum_{k=n}^{n+[aN]+1} S(k) c_k(a, b), \quad (2)$$

де $c_k(a, b) = \Psi\left(\frac{k+1}{a} - \frac{b}{a\Delta t}\right) - \Psi\left(\frac{k}{a} - \frac{b}{a\Delta t}\right)$, Δt – період дискретизації, a – масштабний коефіцієнт вейвлета, b – параметр зсуву вейвлета, $\Psi(t)$ – базис обраного вейвлета з компактним носієм в діапазоні $[0, N-1]$, $n = b/\Delta t$.

Для виділення детермінованих складових мовного сигналу, характерних для гармонік основного тону, використаємо абсолютні значення розрахованих вейвлет-коефіцієнтів.

Прийнявши значення масштабного коефіцієнта $a = a_0$, вейвлет-коефіцієнти перетворення $WLT(a_0, b)$ можна розглядати як дійсний ненульовий часовий ряд $B = (b_0, \dots, b_i, \dots, b_{N-1})$, де $i = \overline{0, N-1}$.

Виконаємо процедуру конвертації вихідного часового ряду B на послідовність багатовимірних векторів використовуючи вікно із L відліків, $1 < L < N$. Результатом процедури конвертації буде $K = N - L + 1$ векторів $X_i = (b_{i-1}, \dots, b_{i+L-2})^T$, $1 \leq i \leq K$ розмірності L , які організуємо в матрицю

$$X = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} b_1 & b_2 & \dots & b_{K-1} \\ b_2 & b_3 & \dots & b_K \\ \vdots & \vdots & & \vdots \\ b_{L-1} & b_L & \dots & b_{N-1} \end{pmatrix}. \quad (3)$$

Матриця (3) є ганкелевою матрицею, тому що матриця $x_{ij} = b_{i+j-2}$ і матриця X мають однакові елементи на діагоналях $i + j = const$.

Виконаємо сингулярний розклад матриці X часового ряду B , для чого розрахуємо матрицю $S = XX^T$. Якщо $\lambda_1, \lambda_2, \dots, \lambda_L$ – власні числа матриці S , взяті в неспадному порядку ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$), U_1, \dots, U_L – ортонормована система власних векторів матриці S , які відповідають власним числам, $d = \max\{i : \lambda_i > 0\}$ і $V_i = X^T U_i / \sqrt{\lambda_i}$, $i = \overline{1, d}$, то сингулярний розклад матриці формалізується як:

$$X = \sum_{i=1}^d X_i, = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T. \quad (4)$$

Далі використовуючи розклад (4) виконаємо процедуру групування, яка розділить множину індексів $\overline{1, d}$ на m підмножин I_1, \dots, I_m , які не перетинаються. Якщо $I = \{i_1, \dots, i_p\}$, то результуюча матриця X_I , яка відповідає групі I , визначається як

$$X_I = X_{i_1} + \dots + X_{i_p}. \quad (5)$$

Аналогічні описаним для отримання матриці (5) обчислення повторюємо для $I = \{i_1, \dots, i_p\}$, що дозволяє сформулювати узагальнений розклад (4):

$$X = X_{I_1} + \dots + X_{I_m}. \quad (6)$$

Далі виконаємо перетворення матриці згрупованого розкладу (6) на новий ряд b_n з N елементів. Нехай Y – матриця розмірністю $L \times K$ з елементами y_{ij} , $1 \leq i \leq L$, $1 \leq j \leq K$. Якщо $L^* = \min(L, K)$, $K^* = \max(L, K)$, $N = L + K - 1$ і $y_{ij}^* = y_{ij}$ (якщо $L < K$) або $y_{ij}^* = y_{ji}$ (якщо $L \geq K$). Діагональне усереднення перетворює матрицю Y в ряд g_0, \dots, g_{N-1} за формулою

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2}^*, & \text{якщо } 0 \leq k \leq L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2}^*, & \text{якщо } L^* \leq k \leq K^*, \\ \frac{1}{N-k} \sum_{m=1}^{N-K^*+1} y_{m,k-m+2}^*, & \text{якщо } K^* \leq k \leq N. \end{cases} \quad (7)$$

Вираз (7) відповідає усередненню елементів матриці Y здовж діагоналей $i + j = k + 2$, тобто при $k = 0$ $g_0 = y_{11}$, при $k = 1$ $g_1 = (y_{12} + y_{21})/2$ і т. д. Якщо матриця Y є траєкторною матрицею деякого ряду (h_0, \dots, h_{N-1}) , то $g_i = h_i$ для всіх i .

Застосовуючи діагональне усереднення (4) до результуючих матриць X_{I_k} , отримуємо ряди $B^{<k>} = (b_0^{<k>}, \dots, b_{N-2}^{<k>})$, і, відповідно, вихідний ряд (b_0, \dots, b_{N-1}) розкладається в суму m рядів:

$$B_j^n = b_n = \sum_{k=1}^m b_n^{<k>}, \quad (8)$$

де індекс j – порядковий номер коефіцієнта a_j вейвлет-розкладу мовного сигналу S_N .

Частотний ряд претендентів в частоту основного тону (f_1, f_2, \dots, f_L) визначимо із співвідношення

$$f_n = \max[WL T(B_{L,N})] = \frac{p}{N\Delta t},$$

$$p = \left(k, \left\| \frac{1}{N} \sum_{j=1}^N B_j^n e^{-\frac{2\pi i k j}{N}} \right\|_k \subseteq MAX, k = \overline{1, N} \right), n = \overline{1, L}, \quad (9)$$

де Δt – період дискретизації, p – індекс елемента в ряду $B_i^{<n>}$, який відповідає максимальній амплітуді вейвлет-перетворення в n -й квазігармонії.

Далі проріджуємо ряд (f_1, f_2, \dots, f_L) до розмірності K , обмеживши пошук частоти основного тону діапазоном $[f_{\min}, f_{\max}]$:

$$f_j = f_n \in [f_{\min}, f_{\max}], j = \overline{1, K}, K < L. \quad (10)$$

Виберемо з діапазону (10) кандидата f_0 за критерієм найменшої кратності

$$f_0 = f_j \in (\min(f_j), 2 \min(f_j), \dots, M \min(f_j)) \quad (11)$$

Введемо допоміжний ряд H_i , який чисельно описує кінцево-імпульсну характеристику фільтра низьких частот:

$$H_i = \begin{cases} 2\pi f_c W_i, & \text{якщо } i = 0, \\ \frac{\sin(2\pi f_c i)}{2\pi f_c}, & \text{якщо } i > 0, \end{cases} \quad (12)$$

де $f_c = f_0 + \Delta f$ – частота зрізу, $W_i = 0.5 \left(1 - \cos\left(\frac{2\pi i}{N-1}\right) \right)$ – ряд, отриманий із використанням віконної функції Ханна.

Далі отримуємо ряд $T0_n$ застосовуючи операцію згортання:

$$T0_n = \sum_{i=1}^N H_i \times S_{n-i}, n = \overline{1, N}. \quad (13)$$

Так як ряд $T0_n$ це квазігармонічна траєкторія, що відповідає основному тону на заданому часовому проміжку із урахуванням модуляції, то його середня частота $F0$ і його середня амплітуда $T0$ можуть бути знайдені усередненням:

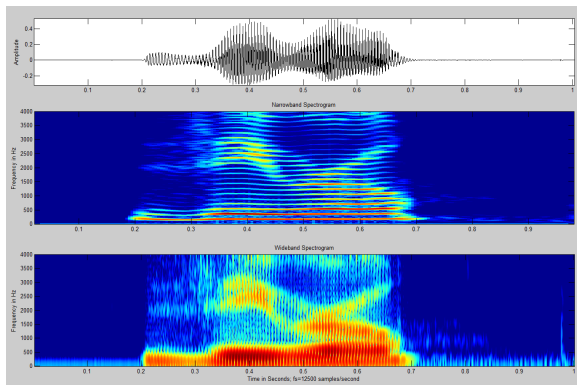
$$F0 = \frac{f_0^1 + f_0^2 + \dots + f_0^m}{m-1} = \frac{\sum_{i=1}^m \frac{1}{(k_i - k_{i-1})\Delta t}}{m-1}, \quad (14)$$

$$T0 = \frac{1}{m} \sum \max(T0_n), n = \overline{1, m},$$

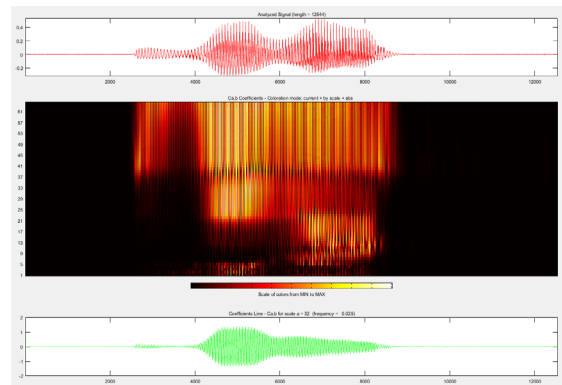
де $k_i = \{n, T0_n \in \max, n = \overline{1, N}\}, i = \overline{1, m}$ – номер індексу в точці максимуму.

АЛГОРИТМ ВІЗУАЛІЗАЦІЇ ЧАСТОТИ ОСНОВНОГО ТОНУ МОВЦЯ ДЛЯ ЗГОРТАЛЬНОГО НЕЙРОМЕРЕЖЕВОГО КЛАСИФІКАТОРА

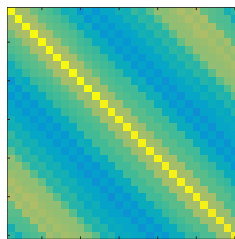
Сформуємо Matlab-орієнтований алгоритм візуалізації частоти основного тону мовця для згортального нейромережевого класифікатора на основі сформульованого вище метода екстрагування частоти основного тону із застосуванням вейвлет-аналізу та сингулярного спектрального аналізу (для наочності супроводжуватимемо основні етапи алгоритму демонстраційними рисунками обробки запису мовного сигналу у вигляді wav-файлі із частотою дискретизації 8 кГц):



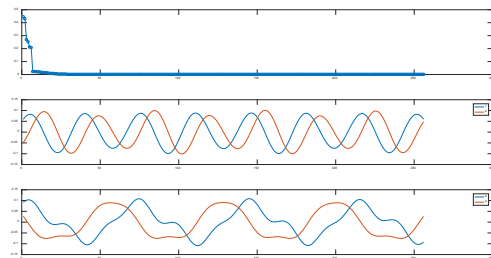
а)



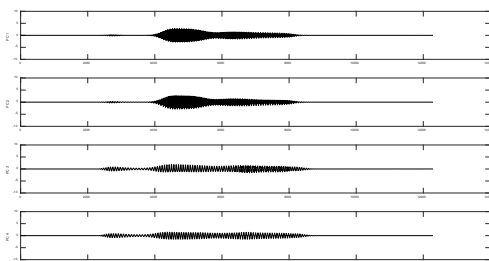
б)



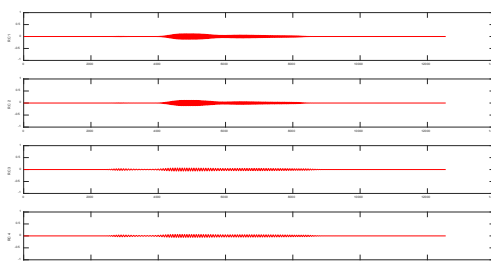
в)



г)



д)



е)

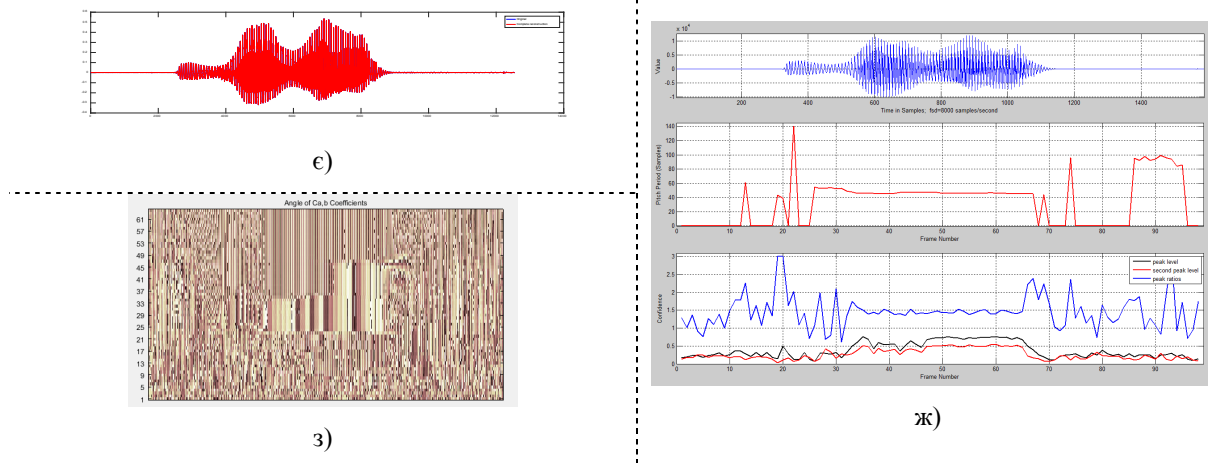


Рис. 1. Алгоритм візуалізації частоти основного тону мовця для згортального нейромережевого класифікатора: а) Спектрограма мовного сигналу; б) Дискретне вейвлет перетворення мовного сигналу; в) Коваріаційна матриця C ; г) Графічне представлення власних значень Λ і власних векторів $Rho1-4$; д) Головні значення $PC1-4$ для власних векторів $Rho1-4$; е) Відновлювані значення $RC1-4$ для власних векторів $Rho1-4$; є) Оцінювання якості сингулярного розкладу мовного сигналу; ж) Узагальнені результати виділення основного тону; з) Результуюча матриця із даними вейвлет-перетворення на періодах основного тону.

1. Завантажуємо wav-файл із мовним сигналом у програмне середовище Matlab використовуючи функцію *audioread* та візуалізуємо його;
2. Отримаємо спектрограму мовного сигналу (рис. 1,а)
3. Виконаємо дискретне вейвлет-перетворення мовного сигналу використовуючи функцію *dwt* із базисом *haar* (рис. 1,б)
4. Обчислюємо коваріаційну матрицю C (рис. 1,в) для фреймів мовного сигналу (256 відліків – приблизно 20 мс) використовуючи функції *xcorr* та *toeplitz*. Довжину «гусені» встановимо рівною четвертій частині від тривалості фрейму;
5. Розраховуємо власні значення λ та власні вектори ρ (рис. 1,г) для коваріаційної матриці C із допомогою функції *eig*, яка повертає матрицю власних векторів Rho , розташованих у стовпчиках матриці, та матрицю власних значень Λ , розташованих на діагоналі матриці;
6. Розраховуємо головні значення PC (рис. 1,д), які є результатом скалярного добутку вектора Y (вектор X із часовою затримкою T) і власних векторів Rho ;
7. Обчислюємо відновлені значення RC (рис. 1,є) для вхідного сигналу X усереднюючи значення з діагоналі матриці, обчисленої добутком вектора Y із інвертованою матрицею власних;
8. Контролюємо якість сингулярного розподілу (рис. 1,ж) оригінального сигналу (синій) відновлюваними значеннями $RC1-4$ (червоний);
9. Використовуючи сформульований вище метод виділення частоти основного тону, узагальнений у виразі (14), виконаємо екстракцію основного тону для фреймів мовного сигналу (узагальнений для всього файлу результат наведено на рис. 1,з);
10. Проріджуємо отриману в п.3 діаграму вейвлет-перетворення мовного сигналу, залишаючи лише відрізки сигналу де спостерігається період основного тону.

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

Проведемо дослідження отриманої із застосуванням створеного алгоритму інформативної ознаки. Для прийняття рішень застосовуватимемо згортальний нейромережевий класифікатор [5]. Архітектура загортальної нейромережі, створеної для автоматизованої системи розпізнання мовців критичного застосування із використанням значень вейвлет-коефіцієнтів на періодах основного тону, наведена на рисунку 2.

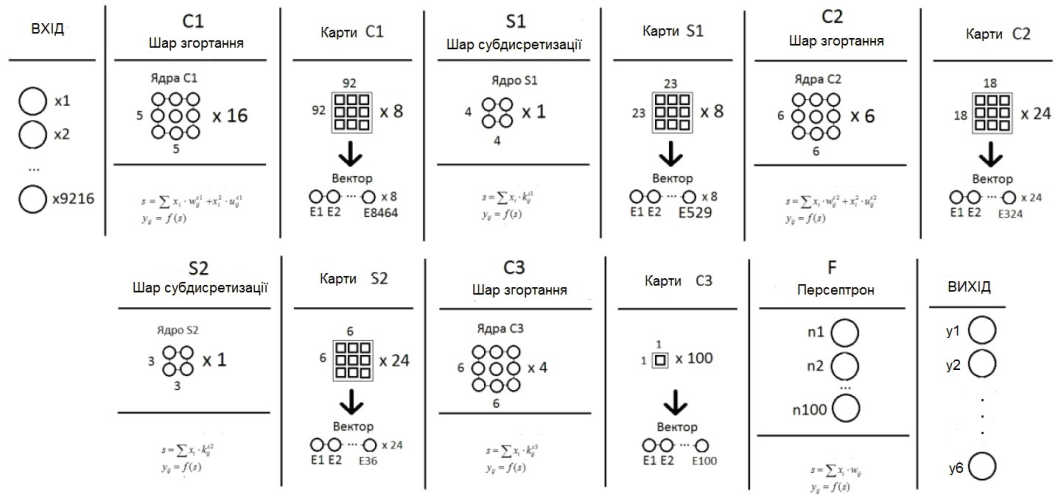


Рис. 2. Архітектура згортальної нейромережі для автоматизованої системи розпізнавання мовців критичного застосування із використанням значень вейвлет-коефіцієнтів на періодах основного тону

Для проведення процедури розпізнавання створено загортальну нейромережу глибокого навчання з використанням кросплатформних бібліотек Caffe [6] з відкритим програмним кодом. На розпізнавання мовців нейромережу було навчено із використанням алгоритму з адаптивною швидкістю навчання (Adaptive Gradient Learning Rate Algorithm, AdaGrad) [7].

В якості бази еталонних записів, які оброблювалися для екстрагування інформативної ознаки, використано безкоштовну базу даних NOIZEUS [8] – спеціалізовану базу даних Школи інжинірингу та комп'ютерних наук Еріка Джонсона при Університеті Техасу в Далласі, США, яка використовується для дослідження алгоритмів покращення звуку і складається з 30 речень англійської розмовної мови, вимовлених трьома чоловіками та трьома жінками (по 5 на кожного диктора, частота дискретизації записів складає 25 кГц, але задля додавання шуму була зменшена до 8 кГц) та записів типових побутових та техногенних шумів. В ході експерименту автоматизовану систему розпізнавання диктора навчали як записами чистих паролних фраз, так і паролними фразами із додаванням шуму. В першому випадку навчальна вибірка містила 18 паролних фраз, в другому – 576, де до чистого сигналу додавався штучний шум з рівнями сигнал/шум 0 дБ, 5 дБ, 10 дБ, 15 дБ відповідно.

Для навчання нейромережі використано 60% обсягу бази аудіозаписів, в яку увійшли екземпляри записів без шумів та із різним рівнем сигнал/шум (5, 10, 15 дБ) відповідно. Тестуючи вибірка складала решту 40% аудіозаписів. Узагальнені результати експерименту представлено на рисунку 8, де імовірність правильного розпізнавання розраховувалася за формулою

$$P = \frac{\sum (N_{p_i})}{N}, \quad (15)$$

де N_{p_i} – кількість правильних результатів розпізнавання i -го диктора, N – загальна кількість експериментів.

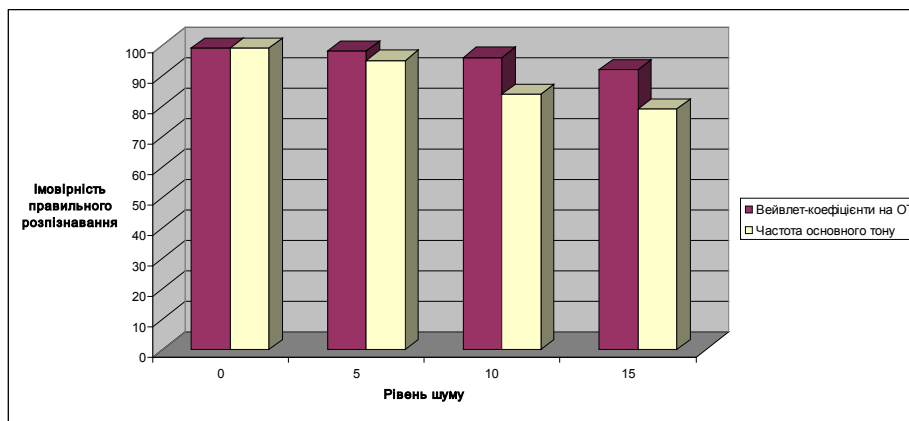


Рис. 3. Результати автоматизованої системи розпізнавання мовців критичного застосування

ВИСНОВКИ

Отже, отримані авторами результати дозволяють стверджувати, що застосування комплексної інформативної ознаки – значень вейвлет-коефіцієнтів на періодах основного тону дозволяє ефективно розпізнавати мовців (імовірність правильного розпізнавання на рівні 99% для 6 мовців при співвідношенні шум/сигнал 10 дБ) із застосуванням загортального нейромережевого класифікатора глибокого навчання. Авторами запропоновано метод та алгоритм отримання зображень значень вейвлет-коефіцієнтів та періодах основного тону із застосуванням вейвлет-аналізу та сингулярного спектрального аналізу, який дозволяє отримувати спектр мовного сигналу без обмежень роздільної здатності в частотному та часовому просторах, на відміну від Фур'є-аналізу. Проте, до недоліків створено методу можна віднести значну обчислювальну складність.

СПИСОК ЛІТЕРАТУРИ

1. Биков М. М., Ковтун В. В. Аналіз ефективності ідентифікації мовця за частотою основного тону // Вісник Хмельницького національного університету. – 2004. – № 2. – Ч. 1. – Т. 2(60). – С. 20-23.
2. Рабинер Л., Шафер Р. Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981.– 496 с.
3. Новиков Л.В. Основы вейвлет-анализа сигналов. Учебное пособие. – СПб: ООО "МОДУС+", 1999. – 152 с.
4. Голяндина Н. Э. Метод “Гусеница”-SSA: Анализ временных рядов: учеб. пособие / Н. Э. Голяндина. – СПб: Изд-во СПбГУ, 2004. – 76 с.
5. CS231n: Convolutional Neural Networks for Visual Recognition [Електронний ресурс] – Режим доступу: <http://cs231n.github.io/convolutional-networks/>
6. Caffe | Deep Learning Framework [Електронний ресурс] – Режим доступу: <http://caffe.berkeleyvision.org/>.
7. An overview of gradient descent optimization algorithms [Електронний ресурс] – Режим доступу: <http://sebastianruder.com/optimizing-gradient-descent/>.
8. NOIZEUS: Noisy speech corpus – Univ. Texas-Dallas [Електронний ресурс] – Режим доступу: <http://ecs.utdallas.edu/loizou/speech/noizeus/>.

REFERENCES

1. Bykov M. M., Kovtun V. V. Analiz efektyvnosti identyfikatsiyi movtsya za chastotoyu osnovnoho tonu // Visnyk Khmelnytskoho natsional'noho universytetu. – 2004. – № 2. – CH. 1. – T. 2(60). – S. 20-23.
2. Rabyner L., Shafer R. Tsyfrovaya obrabotka rechevykh syhnalov.– М.: Radyo y svyaz, 1981.– 496 s.
3. Novykov L. V. Osnovy veyvlet-analyza syhnalov. Uchebnoe posobyе. – SPb: ООО "MODUS+", 1999. – 152 s.
4. Holyandyna N. É. Metod “Husenytса”-SSA: Analiz vremennykh ryadov: ucheb. posobyе / N. É. Holyandyna. – SPb: Yzd-vo SPbHU, 2004. – 76 s.
5. CS231n: Convolutional Neural Networks for Visual Recognition [Електронний ресурс] – Режим доступу: <http://cs231n.github.io/convolutional-networks/>
6. Caffe | Deep Learning Framework [Електронний ресурс] – Режим доступу: <http://caffe.berkeleyvision.org/>.
7. An overview of gradient descent optimization algorithms [Електронний ресурс] – Режим доступу: <http://sebastianruder.com/optimizing-gradient-descent/>.
8. NOIZEUS: Noisy speech corpus – Univ. Texas-Dallas [Електронний ресурс] – Режим доступу: <http://ecs.utdallas.edu/loizou/speech/noizeus/>.

Надійшла до редакції 09.02.2017р.

БИКОВ МИКОЛА МАКСИМОВИЧ – к.т.н., доцент, професор кафедри комп'ютерних систем управління, Вінницький національний технічний університет, м. Вінниця, Україна.

КОВТУН В'ЯЧЕСЛАВ ВАСИЛЬОВИЧ – к.т.н., доцент, доцент кафедри комп'ютерних систем управління, Вінницький національний технічний університет, м. Вінниця, Україна.