



DOI 10.28925/2663-4023.2020.7.3142

УДК 004.934

Терейковська Людмила Олексіївна

кандидат технічних наук, доцент, доцент кафедри інформаційних технологій проектування та прикладної математики

Київський національний університет будівництва і архітектури, Київ, Україна

ORCID ID: 0000-0002-8830-0790

tereikovskal@ukr.net

МЕТОД НЕЙРОМЕРЕЖЕВОГО АНАЛІЗУ ГОЛОСОВОГО СИГНАЛУ

Анотація. Стаття присвячена підвищенню ефективності засобів аналізу біометричних характеристик суб'єктів, що взаємодіють з інформаційно-управляючими системами різного призначення. Показано, що з позицій можливості застосування в якості сенсора зчитування біометричних параметрів стандартного периферійного обладнання комп'ютера, широкого використання в інформаційно-управляючих системах голосових повідомлень, їх високої інформативності, складності підробки біометричної інформації, а також можливості проведення прихованого моніторингу широкі перспективи мають засоби аналізу голосового сигналу. Обґрунтовано необхідність вдосконалення методології нейромережевого аналізу голосового сигналу для розпізнавання емоцій і особи. Визначено можливість підвищення ефективності нейромережевих засобів аналізу за рахунок адаптації параметрів нейромережевої моделі до умов використання. Сформовано принципи визначення допустимих нейромережевих моделей і найбільш ефективного виду нейромережевої моделі, що використовується для аналізу голосового сигналу. Розроблено процедуру кодування вхідного сигналу, що дозволяє використовувати в нейронній мережі компактне представлення найбільш інформативних ознак голосового сигналу. Також розроблено процедуру кодування вихідного сигналу нейронної мережі, що дозволяє підвищити ефективність її навчання. Розроблено метод нейромережевого аналізу голосового сигналу, який за рахунок запропонованих принципів адаптації і процедур кодування вхідних і вихідних параметрів, дозволяє впровадити в засоби розпізнавання нейронну мережу, архітектура якої адаптована до очікуваних умов застосування. Ефективність запропонованого методу доведено експериментальним шляхом. Проведені експериментальні дослідження показали, що використання розробленого методу дозволяє забезпечити точність розпізнавання емоцій ідентифікованого диктора, яка дорівнює приблизно 0.94, що відповідає кращим сучасним рішенням в даній області. Запропоновано співвіднести шляхи подальших досліджень з розробкою рішень, які дозволили б аналізувати голосовий сигнал довільної тривалості в умовах дії шуму різного виду.

Ключові слова: розпізнавання емоцій; аутентифікації; голосовий сигнал; нейронна мережа; метод розпізнавання.

1. ВСТУП

Підвищення ефективності засобів аналізу біометричних характеристик людини є одним з основних напрямків розвитку багатьох сучасних інформаційно-управляючих систем. Зазначені засоби вже досить апробовані в сфері кібербезпеки при проведенні біометричної аутентифікації користувачів при вході в систему, а також в процесі її експлуатації [1], [2]. Крім цього, подібні засоби знайшли своє застосування в медицині для діагностування хвороб, в системах розпізнавання впливу на підсвідомість в мультимедійних повідомленнях засобів масової інформації при веденні інформаційних воєн, в маркетингових системах з метою визначення вражень клієнтів від нової продукції, в системах контролю функціонального стану операторів об'єктів критичної



інфраструктури з метою запобігання виникнення позаштатних ситуацій, пов'язаних з людським фактором, а також в системах управління розумним будинком для оптимізації мікроклімату [3], [4].

Крім того, викликає інтерес застосування засобів аналізу біометричних характеристик людини в системах дистанційної освіти для оптимізації навчального навантаження. При цьому очікуваним результатом аналізу біометричних характеристик в більшості випадків є інтегральне розпізнавання особи та емоційного стану індивіда. Разом з тим, практичний досвід і результати науково-прикладних робіт [5], [6] вказують на необхідність суттєвої модернізації таких аналізаторів в напрямку зменшення ресурсоемності, збільшення точності розпізнавання, розширення номенклатури було розпізнати емоцій, скорочення терміну розробки та підвищення рівня адаптації до інших умов застосування, що й обумовлює актуальність досліджень у цьому напрямку.

Постановка проблеми. Вдосконалення методології аналізу біометричних характеристик людини для аутентифікації та розпізнавання емоційного і функціонального стану особи.

Аналіз останніх досліджень і публікацій. Відповідно до загальноприйнятої термінології під поняттям біометричної характеристики розуміють біологічні та поведінкові характеристики індивіда, які можуть бути зареєстровані і використані в якості відмінних, повторюваних біометричних ознак для автоматичного розпізнавання індивідів [7], [8]. При цьому функціонування більшості згаданих вище засобів базується на аналізі таких біометричних характеристик, як структура судин кисті руки, динаміка рукописного підпису, клавіатурний почерк, текстура шкіри обличчя, топографія кисті руки, папілярна структура Гальтона, топографія особи, топографія пальця, структура райдужної оболонки ока, папілярна структура долоні, зображення сітківки ока, голос, а також хода індивіда. У той же час в джерелах [9], [10], [11] вказується на те, що з позицій можливості застосування в якості сенсора зчитування біометричних параметрів стандартного периферійного обладнання комп'ютера, широкого використання в інформаційно-управляючих системах голосових повідомлень, їх високої інформативності, складності підробки біометричної інформації, а також можливості проведення прихованого моніторингу широкі перспективи мають засоби аналізу голосового сигналу.

Проведений аналіз науково-практичних робіт, присвячених розробці засобів аналізу голосового сигналу, вказує на те, що традиційно виконуваними завданнями зводяться до розпізнавання окремих слів і особи диктора. Останнім часом коло цих задач істотно розширилось. В першу чергу додалася необхідність розпізнавання емоційного та функціонального стану диктора. Крім цього, вважаються перспективними напрямки визначення психічних відхилень, діагностування серцево-судинних захворювань, оцінки достовірності інформації диктора.

Слід зазначити, що різні аспекти проблеми аналізу голосового сигналу людини розглядалися низкою вітчизняних і зарубіжних вчених: Бенжі О., Вінцюком Т. К., Гусевим М. М., Зайцевим В. Г., Михайленком В. М., Рабінером Л. Р., Шафером Р. В., Хінтоном Д. Однак в роботах зазначених вчених при аналізі голосового сигналу акценти ставилися на підвищення ефективності розпізнавання слів і особи диктора. Інші завдання, в силу своєї новизни, в більшості випадків були залишені без уваги.

Як свідчать дані [12]-[14] в основі функціонування поширених систем аналізу голосового сигналу покладені математичні моделі, що базуються на байєсовському підході, прихованих марківських процесах, машинах опорних векторів, методах

динамічного програмування, а також теорії нейронних мереж. При цьому і дані сучасних літературних джерел [15], [16] і результати дослідження найбільш відомих засобів аналогічного призначення (Google+, Microsoft Office, VoiceNavigator, Siri) дозволяють стверджувати, що найбільш широкі перспективи має нейромережева методологія аналізу голосового сигналу.

При цьому в проаналізованій літературі немає єдиної думки про те, який вид нейромережевої моделі доцільно використовувати для вирішення задачі розпізнавання особи та розпізнавання її емоцій.

Також недостатньо повно висвітлено питання попередньої обробки голосового сигналу перед його подачею на вхід нейронної мережі. При цьому в теоретичних роботах [3], [17], [18] вказується, що обґрунтованість процедури визначення вхідних та вихідних параметрів, правильність вибору архітектури нейромережевої моделі є найважливішими факторами, що зумовлюють ефективність її застосування.

Мета статті. Розробка методу нейромережевого аналізу голосового сигналу, який за рахунок адаптації виду і параметрів нейромережевої моделі дозволить забезпечити ефективне розпізнавання емоцій і особи диктора.

2. РОЗРОБКА ТА ДОСЛІДЖЕННЯ МЕТОДУ НЕЙРОМЕРЕЖЕВОГО АНАЛІЗУ ГОЛОСОВОГО СИГНАЛУ

В якості відправної точки розробки методу нейромережевого аналізу голосового сигналу використана наведена в [19] методологія побудови нейромережевих засобів захисту інформації, описаний в [20] метод нейромережевого аналізу клавіатурного почерку, а також запропоновані в [21] парадигми нейромережевого розпізнавання емоцій.

2.1. Процедура кодування вхідних параметрів

Відповідно до даних [19], перший етап методу нейромережевого аналізу співвіднесено з формуванням множини вхідних і вихідних параметрів нейромережевої моделі, що зумовило створення відповідних процедур кодування цих параметрів.

В основу процедури кодування вхідного сигналу нейросетевой моделі покладені дані [6], [10] які передбачають представлення голосового сигналу у вигляді множини мел-кепстральних коефіцієнтів. Відповідно до цих даних реалізація процедури передбачає виконання таких кроків:

- Вхідний дискретний голосовий сигнал фільтрується. Для цього використовуються вирази:

$$z(n) = x(n) - 0,9x(n-1), \quad (1)$$

$$x(n) = z(n) \left[0,54 - 0,46 \cos \left((i-6) \frac{2\pi}{180} \right) \right], \quad (2)$$

де $x(n)$ - дискретне значення амплітуди голосового сигналу при n -ій реєстрації.

- Розраховується спектр голосового сигналу. Для цього в більшості розглянутих робіт використовується метод дискретного перетворення Фур'є:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-i2\pi kn/N}, \quad (3)$$

де N - кількість реєстрацій голосового сигналу.

Для згладжування спектрограми на кордонах квазістаціонарних фрагментів до значень, отриманих за допомогою виразу (4), застосовується віконна функція Хеммінга:

$$X(k) = X(k)H(k), \quad (4)$$

де $H(k) = [0,54 - 0,46 \cos(2\pi k/N - 1)]$ - функція Хеммінга

- Використовуючи вирази (5-8), розраховуються параметри фільтрів:

$$B(b) = 1125 \ln(1 + b/700), \quad (5)$$

$$B^{-1}(b) = 700(e^{b/1125} - 1), \quad (6)$$

$$f(m) = \left(\frac{N}{F}\right) B^{-1}\left(B(f_1) + m \frac{B(f_h) - B(f_1)}{M + 1}\right), \quad (7)$$

$$Q_m(k) = \begin{cases} 0, & \text{if } k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)}, & \text{if } f(m-1) \leq k < f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)}, & \text{if } f(m) \leq k \leq f(m+1) \\ 0, & \text{if } k > f(m+1) \end{cases}, \quad (8)$$

де m - номер фільтра, M - кількість фільтрів, F - частота дискретизації.

- Далі, використовуючи вирази (9, 10), розраховуються значення мел-кепстральних коефіцієнтів:

$$S(m) = \ln \left(\sum_{k=0}^{N-1} X(k)^2 Q_m(k) \right), \quad 0 \leq m \leq M. \quad (9)$$

$$C(n) = \sum_{m=0}^{M-1} S(m) \cos(\pi n(m + 0,5)), \quad 0 \leq n \leq M. \quad (10)$$

де $C(n)$ - значення n -го мел-кепстрального коефіцієнта.

2.2. Процедура кодування вихідних параметрів

При розробці процедури кодування вихідного сигналу нейромережевої моделі використаний обґрунтований в [14], [17] метод кодування вихідних параметрів для нейромережевих засобів призначених для розпізнавання мережевих кібератак. Цей метод дозволяє врахувати в очікуваному вихідному сигналі мережі близькість еталонів класів кібератак, що мають бути розпізнані. При цьому для визначення близькості еталонів використана дещо модифікована ймовірнісна нейронна мережа типу PNN, структура якої показана на рис. 1.

Така мережа призначена для поділу вхідних образів на три класи: А, В і С. Кількість вхідних параметрів дорівнює K . Дана мережа складається з чотирьох шарів нейронів: вхідного - Ln_{in} , образів - Ln_o , додавання - Ln_s і вихідного - Ln_{out} . Кожен нейрон шару Ln_o співвідноситься з конкретним навчальним прикладом. Кількість нейронів шару Ln_o дорівнює кількості навчальних прикладів.

Структура зв'язків між шарами Ln_{in} та Ln_o повнозв'язна. Кожен нейрон шару Ln_s співвідноситься з окремим розпізнаваним класом, а тому їх кількість дорівнює кількості класів. Слід зазначити, що нейрони шару Ln_o зв'язуються тільки з тим нейроном шару Ln_s , який відповідає класу навчального прикладу. Ваговий коефіцієнт такого зв'язку дорівнює 1, а всі інші вагові коефіцієнти вихідних зв'язків рівні 0 (на рис. 1 ці зв'язки не показані). Для вхідних зв'язків нейронів шару Ln_o вагові коефіцієнти дорівнюють складовим частинам відповідних навчальних прикладів.

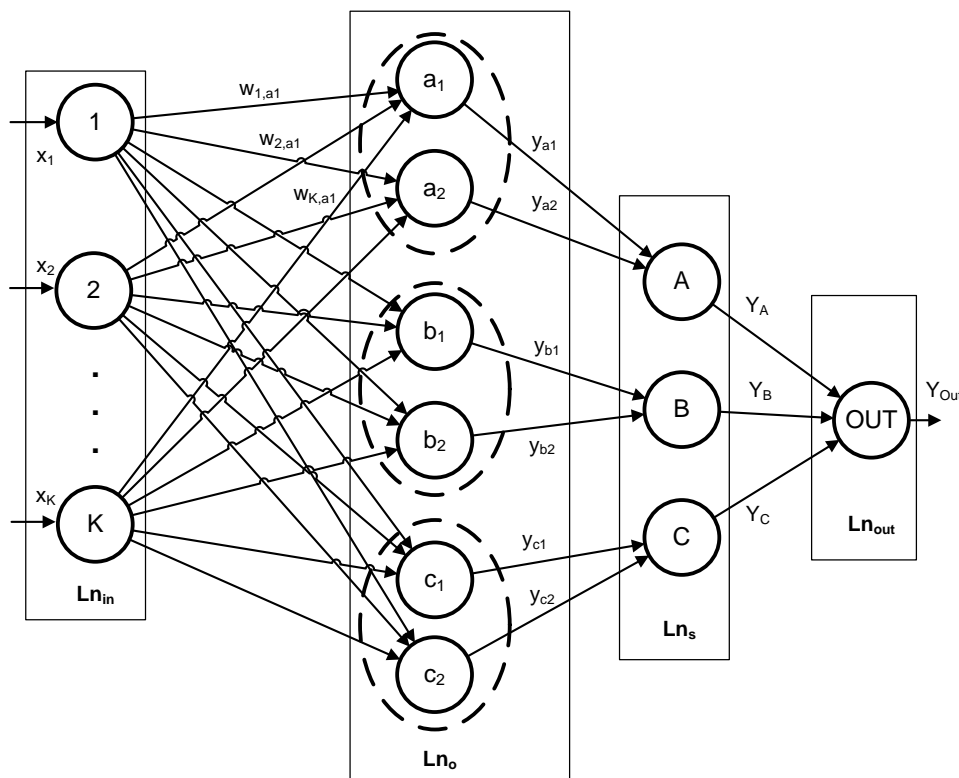


Рис 1. Структура нейронної мережі типу PNN

Процедура кодування вихідного сигналу передбачає реалізацію наступних кроків:

1. Визначити множину класів, що мають бути розпізнані.
2. Сформувати навчальну вибірку, що складається з еталонів класів, які повинні бути розпізнані. Для цього можливо використовувати такі вирази:

$$y(n) = \bar{y}(n)/N = n/N, \quad (11)$$

$$y_n(n) = 1 \wedge y_k(n) = 0, \quad (12)$$

де $\bar{y}(n)$ - необроблене значення вихідного сигналу для еталона n -го класу, що має бути розпізнаний, n - номер класу, що має бути розпізнаний, N - кількість класів, що мають бути розпізнані, k - номер вихідного нейрону.

Відзначимо, що результатом розпізнавання є особа диктора і/або емоція диктора. Вираз (11) слід використовувати для нейромережевої моделі з одним вихідним нейроном, в вираз (12) - для нейромережевої моделі, у якій кількість вихідних нейронів дорівнює кількості класів.

3. Побудувати мережу PNN і провести її навчання. Для цього:

3.1. Визначити множину вхідних нейронів, що відповідають множини вхідних параметрів моделі.

3.2. Визначити множину нейронів шару додавання, що відповідає множині класів, які повинні бути розпізнані.

3.3. Визначити множину нейронів шару образів, які відповідають множині навчальних прикладів.

3.4. Для кожного з нейронів шару образів визначити вагові коефіцієнти вхідних зв'язків.

3.5. Для кожного з нейронів шару образів визначити зв'язок з відповідним нейроном шару додавання.

4. По черзі подати на вхід PNN еталони класів і, використовуючи вирази (13, 14), для кожного з еталонів розрахувати величину вихідного сигналу кожного з нейронів шару додавання. Саме ці величини і є очікуваними вихідними сигналами для кожного з еталонів класів, що мають бути розпізнані.

$$y_j = \sum_{k=1}^K \exp(-\sigma^{-2}(w_{k,j} - x_k)^2), \quad (13)$$

де y_j - вихідний сигнал j -го нейрону шару L_{n_0} , x_k - k -ий вхідний параметр невідомого образу, $w_{k,j}$ - ваговий коефіцієнт зв'язку між k -им нейроном шару $L_{n_{in}}$ і j -им нейроном шару L_{n_0} , K - кількість вхідних параметрів, σ - радіус функції Гаусса.

$$Y_n = \frac{1}{I} \sum_{i=1}^I y_i, \quad (14)$$

Y_n - вихідний сигнал довільного n -го нейрона шару L_{n_s} , I - кількість нейронів шару L_{n_0} , які пов'язані з n -им нейроном шару L_{n_s} , y_j - вихідний сигнал i -го нейрона шару L_{n_0} , який пов'язаний з n -им нейроном шару L_{n_s} .

5. При необхідності отримані величини вихідних сигналів мають бути масштабовані [3], [19]. Масштабовані величини і будуть очікуваним вихідним сигналом для еталонів класів.

2.3. Принципи адаптації параметрів нейронної мережі

Розробка процедур кодування вхідних і вихідних сигналів дозволила перейти до формування наступних етапів методу побудови нейромережевого аналізу голосу, які, згідно з результатами [20], повинні співвідноситися з адаптацією параметрів нейронної мережі до умов поставленого завдання. В основу адаптації покладено групу принципів, що впливають з обґрунтованих в [21] парадигм нейромережевого розпізнавання емоцій.

Принцип 1. Вид та параметри нейромережевої архітектури повинні адекватно відповідати значущим умовам завдання розпізнавання емоцій. Аналітичне визначення вказаного принципу відповідає виразам (15, 16).

$$h_{D1}(v_i) \rightarrow \max, v_i \in \{V\}_I, \quad (15)$$

$$h_{D2}(v_{\max}, m_k) \rightarrow \max, m_k \in \{M\}_K, v_{\max} \in \{V\}_I, \quad (16)$$

де v_i - i -ий вид нейромережевої моделі, $\{V\}_I$ - множина допустимих видів нейромережевих моделей, I - кількість допустимих видів нейромережевих моделей, v_{\max} - найбільш ефективний вид нейромережевої моделі, m_k - k -та комбінація параметрів найбільш ефективного виду нейромережевої моделі, $\{M\}_K$ - множина допустимих комбінацій параметрів.

Принцип 2. Формування множини допустимих видів нейромережевих моделей доцільно здійснювати з позицій забезпечення їх ефективного навчання.

2.4. Етапи методу нейромережевого аналізу голосового сигналу

Інтеграція методології побудови нейромережевих систем з розробленими процедурами кодування вхідних і вихідних параметрів і запропонованими принципами адаптації дозволила запропонувати метод нейромережевого аналізу голосового сигналу, який передбачає виконання 7 етапів.

Етап 1. Визначити множину вхідних параметрів нейромережевої моделі. Для цього слід визначити тривалість голосового сигналу і реалізувати описану вище процедуру кодування вхідних параметрів.

Етап 2. Визначити множину емоцій і множину дикторів. Елементи цих множин будуть відповідати вихідним нейронам мережі. Таким чином, кількість вихідних нейронів дорівнює сумі потужностей зазначених множин.

Етап 3. Сформувати навчальну та тестову вибірки. Використовуючи запропоновану процедуру кодування, визначити очікуваний вихідний сигнал для кожного з прикладів.

Етап 4. Базуючись на другому запропонованому принципі адаптації, визначити множину допустимих видів нейромережевих моделей.

Етап 5. Базуючись на першому запропонованому принципі адаптації, визначити найбільш ефективний вид нейромережевої моделі.

Етап 6. Базуючись на результатах [16], провести оптимізацію архітектурних параметрів обраного виду нейромережевої моделі.

Етап 7. Провести експериментальні дослідження і визначити похибку розпізнавання особи і емоцій диктора. У разі недостатньої точності слід перейти до третього етапу методу.

2.5. Експериментальні дослідження

Для верифікації запропонованого методу проведено комп'ютерні експерименти, спрямовані на розпізнавання емоцій одного диктора на основі його голосового сигналу. Прийнято наступні умови застосування нейромережевої моделі:

- Розпізнаванню підлягають чотири емоції - нейтральність, радість, смуток і злість.

- Голосовий сигнал характеризується за допомогою таких параметрів, як тривалість - 7,008 с, частота дискретизації 16000 Гц, тривалість квазістаціонарного фрагмента - 16 мс, кількість мел-кепстральних коефіцієнтів - 26.

- Обсяг навчальної вибірки - 760 прикладів.

- Обсяг тестової вибірки - 76 прикладів.



Відзначимо, що зазначені умови застосування отримані виходячи з використаної бази даних голосових сигналів Emotion_Recognition, доступної на сайті www.kaggle.com.

Відповідно до запропонованого методу нейромережевого розпізнавання визначено, що найбільш ефективним видом нейромережевої моделі є двошаровий перцептрон з такими архітектурними параметрами:

- Кількість вхідних нейронів - 22750.
- Кількість вихідних нейронів - 4.
- Кількість прихованих нейронів - 70.
- Функція активації прихованих нейронів sigmoid.
- Функція активації вихідних нейронів softmax.

Модель двошарового перцептрона реалізована програмно з використанням мови програмування Python і бібліотек Keras, Numpy, Soundfile і Matplotlib. Експерименти проводилися на персональному комп'ютері з процесором Intel Core i7-8700 (3.2-4.6 ГГц), об'ємом оперативної пам'яті 16 ГБ, відеокартою nVidia GeForce GTX 1660Ti, який функціонував під управлінням операційної системи Microsoft Windows 10.

Після навчання тривалістю 100 епох точність розпізнавання склала приблизно 0.94. Отримана точність відповідає кращим сучасним рішенням в даній області [9], [10], а отже, результати експериментів підтверджують ефективність запропонованого методу нейромережевого аналізу голосового сигналу.

7. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Визначено перспективи вдосконалення інтерфейсних модулів інформаційно-управляючих систем за рахунок підвищення ефективності засобів аналізу голосового сигналу з метою розпізнавання емоцій і особи диктора. Обґрунтовано можливість підвищення ефективності нейромережевих засобів аналізу за рахунок адаптації параметрів нейромережевої моделі до умов використання. Сформовано принципи визначення допустимих нейромережевих моделей і найбільш ефективного виду нейромережевої моделі, що використовується для аналізу голосового сигналу. Розроблено процедуру кодування вхідного сигналу, що дозволяє використовувати в нейронній мережі компактне представлення найбільш інформативних ознак голосового сигналу. Також розроблена процедура кодування вихідного сигналу нейронної мережі, що дозволяє підвищити ефективність її навчання. Розроблено метод нейромережевого аналізу голосового сигналу, який за рахунок запропонованих принципів адаптації і процедур кодування вхідних і вихідних параметрів, дозволяє впровадити в засоби розпізнавання нейронну мережу, архітектура якої адаптована до очікуваних умов застосування. Проведені експериментальні дослідження показали, що використання розробленого методу дозволяє забезпечити точність розпізнавання емоцій ідентифікованого диктора, яка дорівнює приблизно 0.94, що відповідає кращим сучасним рішенням в даній області. Запропоновано співвіднести шляхи подальших досліджень з розробкою рішень, які дозволили б аналізувати голосовий сигнал довільної тривалості в умовах шуму.



СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Gnatyuk S. Critical Aviation Information Systems Cybersecurity, Meeting Security Challenges Through Data Analytics and Decision Support, NATO Science for Peace and Security Series, D: Information and Communication Security. - IOS Press Ebooks, Vol.47, №3, pp. 308-316, 2016. doi: 10.3233/978-1-61499-716-0-308.
- [2] Терейковский И., Корченко А. «Модели эталонов лингвистических переменных для обнаружения сниффинг-атак», Захист інформації, Том 19, №3, сс. 228-242, 2017. doi: 10.18372/2410-7840.19.11902.
- [3] Иванов А. И. Нейросетевые алгоритмы биометрической идентификации личности. Москва, 144 с., 2004.
- [4] Gnatyuk S., Sydorenko V., Aleksander M. Unified data model for defining state critical information infrastructure in civil aviation, Proceedings of the 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, Ukraine, May 24-27, 2018, pp. 37-42. doi:10.1109/dessert.2018.8409095.
- [5] Tereikovska L., Tereikovskiy I., Mussiraliyeva S., Akhmed G. Recognition of emotions by facial Geometry using a capsule neural network. International Journal of Civil Engineering and Technology (IJCIET). Vol. 10, Issue 04, April 2019, pp. 270-279.
- [6] Akhmetov B., Tereykovsky I., Doszhanova A., Tereykovskaya L. Determination of input parameters of the neural network model, intended for phoneme recognition of a voice signal in the systems of distance learning. International Journal of Electronics and Telecommunications. Vol 64, No 4 (2018), 425-432. doi: 10.24425/123541.
- [7] ISO/IEC 19794-5:2011/AMD 2:2015 [ISO/IEC 19794-5:2011/AMD 2:2015] Information technology - Biometric data interchange formats - Part 5: Face image data - Amendment 2: XML encoding and clarification of defects. [Електронний ресурс]. Доступно: <https://www.iso.org/standard/61589.html>.
- [8] ISO/IEC 24713-1:2008 Information technology - Biometric profiles for interoperability and data interchange - Part 1: Overview of biometric systems and biometric profiles. [Електронний ресурс]. Доступно: <https://www.iso.org/standard/38823.html>.
- [9] Лебедева Н.Н. и Каримова Е.Д. «Акустические характеристики речевого сигнала как показатель функционального состояния человека». Успехи физиологических наук, №1, сс. 57-95, 2014.
- [10] Ajinkya N. Jadhav, Nagaraj V. Dharwadkarю. A Speaker Recognition System Using Gaussian Mixture Model, EM Algorithm and K-Means Clustering, International Journal of Modern Education and Computer Science, Vol.10, No.11, pp. 19-28, 2018. doi: 10.5815/ijmecs.2018.11.03.
- [11] Altincay H. (2003). Speaker identification by combining multiple classifiers using Dempster-Shafer theory of evidence. Speech Communication, v.41, N4, 531-547. doi.org:10.1016/S0167-6393(03)00032-3
- [12] Toliupa S., Tereikovskiy I., Dychka I., Tereikovska L., Trush A.. The Method of Using Production Rules in Neural Network Recognition of Emotions by Facial Geometry. 3rd International Conference on Advanced Information and Communications Technologies (AICT). 2019, 2-6 July 2019, Lviv, Ukraine, pp. 323 – 327. doi: 10.1109/AIACT.2019.8847847.
- [13] Campbell W., Sturim D., Reynolds D. (2006). Support vector machines using GMM supervectors for speaker verification. IEEE Signal Process. Lett., v.13, N5, pp. 308-311. doi: 10.1109/LSP.2006.870086.
- [14] Tereykovska L., Tereykovskiy I., Ayt Khozhaeva E., Tynymbayev S., Imanbayev A. Encoding of neural network model exit signal, that is devoted for distinction of graphical images in biometric authenticate systems (2017). // News of the national academy of sciences of the republic of kazakhstan series of geology and technical sciences. Volume 6, Number 426 (2017), pp. 217 – 224.
- [15] Akhmetov, B., Tereikovskiy, I., Tereikovska, L., Adranova, A. Neural Network User Authentication by Geometry of the Auricle // Recent Developments in Data Science and Intelligent Analysis of Information Proceedings of the XVIII International Conference on Data Science and Intelligent Analysis of Information, June 4-7, 2018, Kyiv, Ukraine, pp.11-19. doi: 10.1007/978-3-319-97885-7_2.
- [16] Tereikovskiy I, Chernyshev D., Tereikovska L.A., Mussiraliyeva S, Akhmed G. The Procedure for the Determination of Structural Parameters of a Convolutional Neural Network to Fingerprint Recognition. Journal of Theoretical and Applied Information Technology, No 8, Volume 97, 2019, pp. 2381-2392.
- [17] Oksiuk O., Tereikovska L. and Tereikovskiy I. Determination of Expected Output Signals of the Neural Network Model Intended for Image Recognition. 4th International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T) 2017, pp. 596-599. doi: 10.1109/INFOCOMMST.2017.8246471.



- [18] Tereikovskiy, I., Mussiraliyeva, S., Kosyuk, Y., Bolatbek, M., Tereikovska, L. An experimental investigation of infrasound influence hard drives of a computer system // International Journal of Civil Engineering and Technology. 2018. Volume 9, Issue 6, June 2018, pp. 1558–1566.
- [19] Dychka I., Chernyshev D., Tereikovskiy I., Tereikovska L., Pogorelov V. (2020) Malware Detection Using Artificial Neural Networks. Advances in Computer Science for Engineering and Education II. ICCSEEA 2019. Advances in Intelligent Systems and Computing, vol 938. Springer, pp. 3-12. doi:10.1007/978-3-030-16621-2_1.
- [20] Терейковська Л. «Метод нейромережевого аналізу клавіатурного почерку». Комп'ютерно-інтегровані технології: освіта, наука, виробництво, 2019, №37, сс. 53-59. doi: 10.36910/6775-2524-0560-2019-37-8.
- [21] Михайленко В. М., Терейковская Л.А. «Парадигмы нейросетевого распознавания эмоций». Управління розвитком складних систем. 2019. №39. сс. 179-186.

**Liudmyla A. Tereikovska**

Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Information Technologies of Design and Applied Mathematics, Kyiv National University of Construction and Architecture, Kyiv, Ukraine

ORCID ID: 0000-0002-8830-0790

tereikovskal@ukr.net

METHOD OF NEURAL NETWORK ANALYSIS OF VOICE SIGNAL

Abstract. The article is devoted to increase of efficiency of the means of analysis of biometric characteristics of subjects, interacting with information-control systems of various purpose. It is shown that from the standpoint of the possibility of using as a sensor the reading of the biometric parameters of the standard peripheral equipment of the computer, the widespread use in the information-control systems of voice messages, their high informativeness, the complexity of falsification of biometric information, as well as the possibility of carrying out hidden monitoring, the prospects have analysis tools voice signal. The necessity of improvement of methodology of neural network analysis of voice signal for recognition of emotions and person is grounded. Possibility of increase of efficiency of neural network means of analysis due to adaptation of parameters of neural network model to the conditions of use is determined. The principles of determination of valid neural network models and the most effective type of neural network model used for voice signal analysis have been formed. A coding procedure for the input signal is developed, which allows to use in the neural network a compact representation of the most informative features of a voice signal. A procedure for encoding a neural network output signal has also been developed to increase the efficiency of its learning. The method of neural network analysis of the voice signal is developed, which due to the offered principles of adaptation and procedures of coding of input and output parameters, allows to introduce into neural means a neural network whose architecture is adapted to the expected conditions of use. The effectiveness of the proposed method has been proven experimentally. Experimental studies have shown that the use of the developed method allows to ensure the accuracy of recognition of emotions of the identified speaker, which is approximately 0.94, which corresponds to the best modern decisions in this field. It is suggested to correlate the ways of further research with the development of solutions that would allow to analyze the voice signal of arbitrary duration under the conditions of noise of different kind.

Keywords: recognition of emotions; authentication; voice signal; neural network; recognition method.

REFERENCES (TRANSLATED AND TRANSLITERATED)

- [1] Gnatyuk S. Critical Aviation Information Systems Cybersecurity, Meeting Security Challenges Through Data Analytics and Decision Support, NATO Science for Peace and Security Series, D: Information and Communication Security. - IOS Press Ebooks, Vol.47, №3, pp. 308-316, 2016.
- [2] Tereikovskiy I. and Korchenko A. «Modeli etalonov lingvisticheskikh peremennykh dlya obnaruzheniya sniffing-atak», Protection of information, Vol. 19, №3, pp. 228-242, 2017. (In Russian).
- [3] Ivanov A. I. Neyrosetevyye algoritmy biometricheskoy identifikatsii lichnosti. Moscow, 144 p., 2004. (In Russian).
- [4] Gnatyuk S., Sydorenko V., Aleksander M. Unified data model for defining state critical information infrastructure in civil aviation, Proceedings of the 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), Kyiv, Ukraine, May 24-27, 2018, pp. 37-42.
- [5] Tereikovska L., Tereikovskiy I., Mussiraliyeva S., Akhmed G. Recognition of emotions by facial Geometry using a capsule neural network. International Journal of Civil Engineering and Technology (IJCIET). Vol. 10, Issue 04, April 2019, pp. 270-279.
- [6] Akhmetov B., Tereykovskiy I., Doszhanova A., Tereykovskaya L. Determination of input parameters of the neural network model, intended for phoneme recognition of a voice signal in the systems of distance learning. International Journal of Electronics and Telecommunications. Vol 64, No 4 (2018), 425-432.



- [7] ISO/IEC 19794-5:2011/AMD 2:2015 [ISO/IEC 19794-5:2011/AMD 2:2015] Information technology - Biometric data interchange formats - Part 5: Face image data - Amendment 2: XML encoding and clarification of defects. [Електронний ресурс]. Доступно: <https://www.iso.org/standard/61589.html>.
- [8] ISO/IEC 24713-1:2008 Information technology - Biometric profiles for interoperability and data interchange - Part 1: Overview of biometric systems and biometric profiles. [Електронний ресурс]. Доступно: <https://www.iso.org/standard/38823.html>.
- [9] Lebedeva N.N. and Karimova E.D. «Akusticheskiye kharakteristiki rechevogo signala kak pokazatel' funktsional'nogo sostoyaniya cheloveka». *Advances in Physiological Sciences*, №1, pp. 57-95, 2014. (In Russian).
- [10] Ajinkya N. Jadhav, Nagaraj V. Dharwadkarю. A Speaker Recognition System Using Gaussian Mixture Model, EM Algorithm and K-Means Clustering, *International Journal of Modern Education and Computer Science*, Vol.10, No.11, pp. 19-28, 2018.
- [11] Altincay H. (2003). Speaker identification by combining multiple classifiers using Dempster-Shafer theory of evidence. *Speech Communication*, v.41, N4, 531-547.
- [12] Toliupa S., Tereikovskiy I., Dychka I., Tereikovska L., Trush A.. The Method of Using Production Rules in Neural Network Recognition of Emotions by Facial Geometry. 3rd International Conference on Advanced Information and Communications Technologies (AICT). 2019, 2-6 July 2019, Lviv, Ukraine, pp. 323 - 327.
- [13] Campbell W., Sturim D., Reynolds D. (2006). Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Process. Lett.*, v.13, N5, pp. 308-311.
- [14] Tereikovska L., Tereikovskiy I., Ayt Khozhaeva E., Tynymbayev S., Imanbayev A. Encoding of neural network model exit signal, that is devoted for distinction of graphical images in biometric authenticate systems (2017). // *News of the national academy of sciences of the republic of kazakhstan series of geology and technical sciences*. Volume 6, Number 426 (2017), pp. 217 - 224.
- [15] Akhmetov, B., Tereikovskiy, I., Tereikovska, L., Adranova, A. Neural Network User Authentication by Geometry of the Auricle // *Recent Developments in Data Science and Intelligent Analysis of Information Proceedings of the XVIII International Conference on Data Science and Intelligent Analysis of Information*, June 4-7, 2018, Kyiv, Ukraine, pp.11-19.
- [16] Tereikovskiy I, Chernyshev D., Tereikovska L.A., Mussiraliyeva S, Akhmed G. The Procedure for the Determination of Structural Parameters of a Convolutional Neural Network to Fingerprint Recognition. *Journal of Theoretical and Applied Information Technology*, No 8, Volume 97, 2019, pp. 2381-2392.
- [17] Oksiuk O., Tereikovska L. and Tereikovskiy I. Determination of Expected Output Signals of the Neural Network Model Intended for Image Recognition. 4th International Scientific-Practical Conference Problems of Infocommunications Science and Technology (PIC S&T) 2017, pp. 596-599.
- [18] Tereikovskiy, I., Mussiraliyeva, S., Kosyuk, Y., Bolatbek, M., Tereikovska, L. An experimental investigation of infrasound influence hard drives of a computer system // *International Journal of Civil Engineering and Technology*. 2018. Volume 9, Issue 6, June 2018, pp. 1558-1566.
- [19] Dychka I, Chernyshev D., Tereikovskiy I, Tereikovska L., Pogorelov V. (2020) Malware Detection Using Artificial Neural Networks. *Advances in Computer Science for Engineering and Education II. ICCSEEA 2019. Advances in Intelligent Systems and Computing*, vol 938. Springer, pp. 3-12.
- [20] Tereikovska L. «Metod neyromerezhevoho analizu klaviaturnoho pocherku». *Computer-integrated technologies: education, science, production*, 2019, №37, pp. 53-59. (In Ukrainian).
- [21] Mihaylenko V. M. and Tereikovska L A. «Paradigmy neyrosetevogo raspoznavaniya emotsiy». *Management of complex systems development*. 2019. №39. pp. 179-186. (In Russian).

