

## ОПТИМІЗАЦІЯ АЛФАВІТУ ІНФОРМАТИВНИХ ОЗНАК ДЛЯ АВТОМАТИЗОВАНОЇ СИСТЕМИ РОЗПІЗНАВАННЯ МОВЦІВ КРИТИЧНОГО ЗАСТОСУВАННЯ

Вінницький національний технічний університет

### *Анотація*

*В результаті проведених досліджень виконано математичну постановку задачі ідентифікації оптимального алфавіту інформативних ознак для застосування у автоматизованих системах розпізнавання мовців критичного застосування.*

**Ключові слова:** автоматизована система розпізнавання мовців критичного застосування, інформативні ознаки, загоральна нейронна мережа, сингулярний розклад, розклад Карунена-Лоева.

### *Abstract*

*As a result of research is performed mathematical formulation of optimal information features alphabet problem for use in the automated speaker recognition system of critical use.*

**Keywords:** automated speaker recognition system of critical use, information features, convolution neural network, a singular transform, Karhunen-Loeve transform.

### **Вступ**

В задачах розпізнавання образів знайшли своє застосування методи зменшення розмірності простору ознак, основними з яких є метод Карунена-Лоева (Karhunen-Loeve transform, KLT), метод сингулярного розкладу (Singular value decomposition, SVD) та метод головних компонент (Principal component analysis, PCA). Спільною рисою цих методів є застосування математичного перетворення, базисними функціями якого є власні вектори коваріаційної матриці вхідного сигналу, з метою виділення найменш корельованих його компонент.

Метою дослідження є отримання математичної постановки задачі ідентифікації оптимального алфавіту інформативних ознак для застосування у автоматизованих систем розпізнавання мовців критичного застосування.

### **Результати досліджень**

Авторами запропоновано метод і алгоритм оптимізації алфавіту барк-кепстрального аналізу записів мовних сигналів, використовуючи який модуль прийняття рішень автоматизованої системи розпізнавання мовців критичного використання здійснюватиме процес розпізнавання. В якості бази еталонних записів використано безкоштовну базу даних NOIZEUS – 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 мовця, частота дискретизації записів складає 25 кГц, але задля додавання шуму була зменшена до 8 кГц) та записів типових побутових та техногенних шумів.

Для отримання кепстральних коефіцієнтів вхідний мовний сигнал поділявся на кадри, тривалістю 20 мс, з кожного з яких виділялися 12 кепстральних коефіцієнтів, 12 дельта-коефіцієнтів (перша похідна) і 12 подвійних дельта кепстральних коефіцієнтів (друга похідна). В результаті роботи запропонованого алгоритму оптимізації алфавіту інформативної ознаки проводиться селекція частотних смуг за принципом мінімізації коефіцієнтів матриці міжсмугової кореляції, що дозволяє відсіяти менш інформативні смуги і проводити навчання класифікатору на максимально інформативному матеріалі [1]. Для розпізнавання мовців за оптимізованим алфавітом інформативної ознаки використано згоральну нейромережу. Практичну реалізацію загоральної нейромережі глибокого навчання виконано засобами кросплатформних бібліотек Caffe із відкритим програмним кодом. На розпізнавання мовців нейромережу було навчено із використанням алгоритму стохастичного градієнтного спуску (Stochastic Gradient Descent [2]).

Для навчання нейромережі використано 60% обсягу бази аудіозаписів, в яку увійшли екземпляри записів без шумів та із різним рівнем шум/сигнал (5, 10, 15 дБ) відповідно. Тестуюча вибірка складала решту 40% аудіо записів [3]. Узагальнені результати експерименту представлено на рисунку 1.

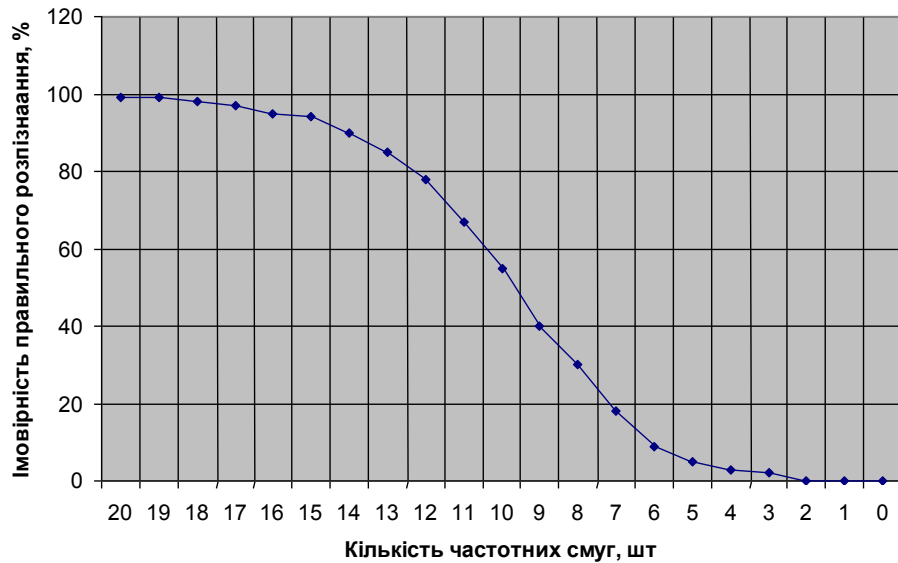


Рис. 1. Залежність імовірності правильного розпізнавання мовців від кількості частотних смуг, інформативні ознаки з яких використовувалися для розпізнавання

Як видно з рисунка, імовірність правильного розпізнавання мовців перевищувала 90% при використанні для розпізнавання 14 частотних смуг із початкових 20 і перевищувала 95% при кількості частотних смуг рівній 16. Така інформація дозволяє зменшити обсяг навчальної вибірки на 30% або 20% відповідно із прогнозованою втратою імовірності правильного розпізнавання [4]. Такою виявилось, що визначені алгоритмом як менш інформативні для розпізнавання мовців є частотні смуги розташовувалися в 4200-8000 Гц.

### Висновки

Експериментальні дослідження довели працездатність запропонованого алгоритму на прикладі оптимізації алфавіту барк-кепстру мовних сигналів, коефіцієнти якого використовувалися для розпізнавання мовців загортальною нейромережею. Застосування алгоритму дозволило виявити залежність між імовірністю правильного розпізнавання мовця і кількістю частих смуг, в яких виділялася інформативна ознака, а також визначити менш інформативні частотні смуги.

### Література

1. Фурман М.С. Дослідження нейромережевого класифікатора цифрових зображень / М.С. Фурман, В.В. Ковтун ; Нац. Ун-т «Вінницький національний технічний університет». – Вінниця : Вид-во Нац. Ун-ту «Вінницький національний технічний університет», 2016. [Електронний ресурс]: Інформаційний портал. Режим доступу: <http://ir.lib.vntu.edu.ua/handle/123456789/10974>.
2. Биков М.М. Метод оптимізації процесу навчання нейромережі в задачі розпізнавання мовців / М.М. Биков, В.В. Ковтун, А. Раїмі ; Нац. Ун-т «Вінницький національний технічний університет». – Вінниця : Вид-во Нац. Ун-ту «Вінницький національний технічний університет», 2015. [Електронний ресурс]: Інформаційний портал. Режим доступу: <http://ir.lib.vntu.edu.ua/handle/123456789/3584>.
3. Биков М.М. Метод нормалізації тривалості звучання паролних фраз для системи розпізнавання мовців / М.М. Биков, В.В. Ковтун, А. Раїмі ; Нац. Ун-т «Вінницький національний технічний університет». – Вінниця : Вид-во Нац. Ун-ту «Вінницький національний технічний університет», 2015. [Електронний ресурс]: Інформаційний портал. Режим доступу: <http://ir.lib.vntu.edu.ua/handle/123456789/3585?locale-attribute=en>.
4. Биков М.М. Метод підвищення ефективності роботи пам'яті в системах пошуку ключових слів у мовному сигналі / М.М. Биков, В.В. Ковтун, К. Конате ; Нац. Ун-т «Вінницький національний технічний університет». – Вінниця : Вид-во Нац. Ун-ту «Вінницький національний технічний університет», 2015. [Електронний ресурс]: Інформаційний портал. Режим доступу: <http://visnyk.vntu.edu.ua/index.php/visnyk/article/view/1221>.

**Берега Артём Олексійович** — студент групи АВ-15мс, факультет комп'ютерних систем і автоматики, Вінницький національний технічний університет, Вінниця, e-mail: [artembereza2017@gmail.com](mailto:artembereza2017@gmail.com)

**Ковтун В'ячеслав Васильович** — доцент кафедри комп'ютерних систем управління Вінницького національного технічного університету, Вінниця.

**Bereza Artem O.** — Department of Automation and computer control systems, Vinnytsia National Technical University, Vinnytsia, e-mail: artembereza2017@gmail.com

**Kovtun Vyacheslav V.** — docent of the Department of Computer control systems, Vinnytsia National Technical University, Vinnytsia