

# ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ РОЗПІЗНАВАННЯ ВИМОВЛЕНИХ ЛІТЕР НА ОСНОВІ СПАЙКІНГОВОЇ НЕЙРОННОЇ МЕРЕЖІ

Вінницький національний технічний університет;

## **Анотація**

*Запропоновано інформаційну технологію для розпізнавання вимовлених літер на основі спайкінгової нейронної мережі. Було розроблено архітектуру обраного типу мережі. Було спроектовано програму розпізнавання вимовлених літер, написану мовою програмування Python у середовищі PyCharm з використанням фреймворку PySNN. Розроблена програма має вищу достовірність (61,7%), ніж аналогічна програма (53,3%), а значить достовірність розпізнавання вимовлених літер покращена на 8,4%.*

**Ключові слова:** інформаційна технологія, розпізнавання звуків, спайкінгова нейронна мережа.

## **Abstract**

*An information technology for recognizing spoken letters based on a spiking neural network is proposed. The architecture of the chosen type of network was developed. A spoken letters recognition program written in the Python programming language in the PyCharm environment using the PySNN framework was designed. The developed program has a higher accuracy (61.7%) than the analog program (53.3%), which means that the accuracy of recognizing spoken letters has improved by 8.4%.*

**Keywords:** information technology, spoken letter recognition, spiking neural network.

## **Вступ**

Розпізнавання мови стає все більш популярним; зростають вимоги до достовірності розпізнавання таких систем, у результаті чого зростає попит та потреба у точних і швидких системах розпізнавання мови. Основним недоліком всіх систем є низька достовірність розпізнавання схожих слів, що починаються чи закінчуються приголосним звуком (літерою). Прикладами таких слів є бочка, почка, дочка, точка, кочка, ночка, мочка. Тому ключ до підвищення загальної достовірності систем розпізнавання мови лежить у підвищенні якості розпізнавання приголосних звуків, особливо таких як б, п, д, т, г, к. Саме цій задачі і присвячена дане дослідження.

## **Результати дослідження**

В ході аналізу предметної області розглянуто основні методи розпізнавання звуків та як найбільш перспективний, було обрано нейромережевий метод. Дослідивши відомі нейронні мережі, для інформаційної технології розпізнавання вимовлених літер, як основа, була обрана спайкінгова нейронна мережа [1].

У нашому випадку мережа складається з двох шарів спайкінгових нейронів Іжикевича [2], з'єднаних тренуваними синапсами. Перший шар мережі - це вхідний шар, що складається з 520 вхідних нейронів, кожен нейрон отримує значення амплітуди однієї ознаки як свій струм інжекції. Вихідний шар складається з 6 нейронів, що відповідають кожному з 6 приголосних звуків (літер). Вихідні нейрони містять ту саму конфігурацію параметрів, що і нейрони вхідного шару. Кожен вхідний нейрон підключений до кожного з вихідних нейронів через тренований синапс, в результаті чого кожен вихідний нейрон має 520 вхідних синапсів. Синапси були ініціалізовані з вагами, рівними 0,5. Навчання для цих синапсів відбувається за методом STDP.

Структура програми розпізнавання вимовлених літер на основі спайкінгової нейронної мережі представлена на рис. 1. Основні блоки структури:

- 1) Формувач мел-кепстральних коефіцієнтів;
- 2) Спайкінгова нейромережа;
- 3) Детектор співпадінь.

Крім блоків, також потрібні інформаційні ресурси:

- 1) Навчальна вибірка;
- 2) Тестова вибірка;
- 3) Прототипи спайк-шлейфів.

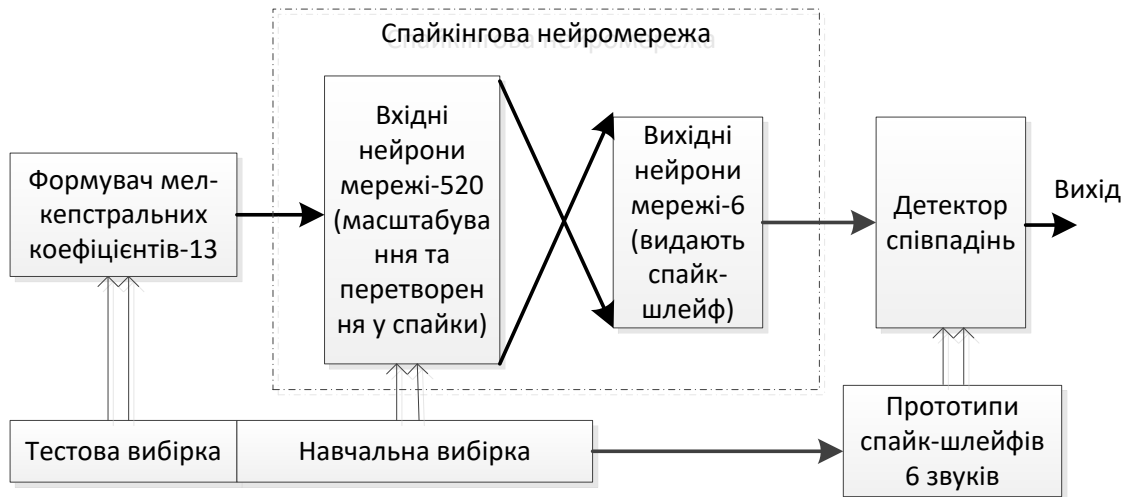


Рисунок 1 - Структура програми розпізнавання вимовлених літер на основі спайкінгової нейронної мережі

Формувач мел-кепстральних коефіцієнтів розбиває звуковий файл тривалістю 0,5 сек на 40 фреймів і для кожного фрейму формує 13 мел-кепстральних коефіцієнтів. Всього отримується 520 чисел, які перетворюються вхідними нейронами мережі [3] у частоту від 2 до 38 Гц.

В режимі навчання спочатку завантажується навчальна вибірка, яка являє собою 120 файлів формату .wav (по 20 звукових файлів на кожному з 6 приголосних літер б, п, д, т, г, к). Потім ініціалізується спайкінгова нейронна мережа, яка складається з 520 вхідних нейронів та 6 вихідних нейронів. Тобто матриця синапсів має розмір 520x6. Далі відбувається процес навчання мережі. Алгоритм навчання – STDP, а ваги навченої мережі зберігаються.

В режимі тестування на вхід мережі може бути поданий будь-який .wav файл із тестової вибірки. Для цього файлу формуються мел-кепстральні коефіцієнти, які масштабуються і перетворюються у спайки (імпульси) вхідними нейронами. Ці спайки далі надходять на вихідні нейрони через натреновані синапси. На виходах вихідних нейронів формуються послідовності імпульсів, які ми називаємо спайк-шлейфами і які є оригінальними для кожної вимовленої літери. Далі ці спайк-шлейфи перевіряються детектором співпадінь на співпадіння із прототипами спайк-шлейфів літер і для кожного виходу мережі формується число в діапазоні 0...1, яке вказує ступінь відповідності тестового сигналу відповідному прототипу вимовленої літери.

Підготовка даних.

База даних, з якої отримані дані для навчальної та тестової вибірки, - це база даних ISOLET Spoken Letter Database, яка містила 7800 вимовлених літер, відібраних із 150 ораторів, половина з яких – чоловіки, а інша половина - жінки, записаних у форматі .wav. Навчальна вибірка має 120 зразків (по 20 зразків для кожної приголосної літери).

Здійснено програмну реалізація інформаційної технології розпізнавання вимовлених літер на основі спайкінгової нейронної мережі на мові програмування Python, у середовища розробки PyCharm з використанням спеціалізованої бібліотеки для створення та навчання спайкінгових нейронних мереж PySNN.

Було перевірено достовірність роботи розробленої програми у порівнянні с аналогом. Для тестування взято вибірку звуків обсягом 60 (по 10 звуків на кожному із 6 літер). Достовірність розпізнавання вимовлених літер розробленої програми та аналога наведена у табл. 1.

Таблиця 1 - Достовірність розпізнавання вимовлених літер розробленою програмою у порівнянні з аналогом

	Достовірність розпізнавання, %						Середня достовірність
	Тестовий звук (літера)						
	Б	Д	Г	К	П	Т	
Аналог	50	60	40	50	70	50	53,3%
Розроблена програма	60	70	50	60	70	60	61,7%

Із табл. 1 видно, що розроблена програма має вищу достовірність розпізнавання вимовлених літер (61,7%), ніж аналогічна програма (53,3%), а значить достовірність розпізнавання вимовлених літер покращена на 8,4% за рахунок використання спайкінгової нейронної мережі.

### Висновки

Таким чином, запропоновано інформаційну технологію розпізнавання вимовлених літер на основі спайкінгової нейронної мережі.. Було здійснено програмну реалізацію розпізнавання вимовлених літер, для чого було використано середовищі PyCharm, мову програмування Python, фреймворк PySNN. Аналіз результатів роботи програми показав, що розроблена програма має вищу достовірність розпізнавання вимовлених літер (61,7%), ніж аналогічна програма (53,3%), а значить достовірність розпізнавання вимовлених літер покращена на 8,4%.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Neurocomputer architecture based on spiking neural network and its optoelectronic implementation / Oleh K. Kolesnytskyj; Vladislav V. Kutsman; Krzysztof Skorupski; Mukaddas Arshidinova, Proc. SPIE 11176, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2019, 1117609 (6 November 2019); doi: 10.1117/12.2536607.
2. Колесницький О. К. Математична модель імпульсного нейроелемента на біспін-приладі / О. К. Колесницький, І. В. Бокоцей // Вісник Хмельницького національного університету. – 2011. – №5. – С.141-149.
3. Колесницький О.К. Моделювання імпульсної нейронної мережі у задачі розпізнавання багатовимірних імпульсних послідовностей / О.К. Колесницький, С.М. Богатчук, М.В.Крещенецька, С.С.Яремчук // Вісник Вінницького політехнічного інституту. – Вінниця: УНІВЕРСУМ-Вінниця, 2008. – №5, С.62-66.

**Семчук Вячеслав Олександрович** — студент групи 2КН-21м, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, м. Вінниця, email: malpiretv@gmail.com.

**Милосердов Дмитро Андрійович** — аспірант кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, email: malpiretv@gmail.com

**Колесницький Олег Костянтинович** — доцент кафедри комп'ютерних наук ВНТУ, Вінницький національний технічний університет, м. Вінниця.

**Semchuk Vladislav O.** — Department of Intelligent Information Technologies and Automatization, Vinnytsia National Technical University, Vinnytsia, email: malpiretv@gmail.com

**Myloserdov Dmytro A.** — PhD student of the Computer Science Dpt., Vinnytsia National Technical University, Vinnytsia, email: malpiretv@gmail.com

**Kolesnytskyj Oleg K.** — docent of the Computer Sciences Dpt., Vinnytsia National Technical University, Vinnytsia