

УДК 004.89:004.934]:621.3.049.77

РЕАЛІЗАЦІЯ ТА ОПТИМІЗАЦІЯ АЛГОРИТМІВ ВИДІЛЕННЯ АУДІООЗНАК (MFCC) ДЛЯ СИСТЕМ EDGE AI НА БАЗІ МІКРОКОНТРОЛЕРІВ STM32

Ткачов Д.Р.

e-mail: denys.tkachov@nure.ua

Науковий керівник – ст. викладач Зайченко О.Б.

Харківський національний університет радіоелектроніки, каф. ПЕЕА
м. Харків, Україна

The rapid advancement of Edge AI requires the deployment of complex audio processing and machine learning algorithms directly on resource-constrained microcontrollers. This thesis focuses on the implementation and optimization of Mel-Frequency Cepstral Coefficients (MFCC) feature extraction algorithms specifically for STMicroelectronics STM32 platforms. It comprehensively describes the theoretical foundations of the MFCC pipeline alongside practical hardware-software co-design strategies utilizing CMSIS-DSP and X-CUBE-AI tools

Швидкий розвиток периферійних обчислень (Edge Computing) зумовив необхідність розгортання складних алгоритмів цифрової обробки сигналів (DSP) та машинного навчання (ML) безпосередньо на мікроконтролерах (MCU). Обробка аудіоданих на периферійних пристроях (наприклад, для розпізнавання мови або діагностики обладнання) вимагає перетворення сирих аудіосигналів у формат, зрозумілий для нейронних мереж. Найбільш критичним методом у цій сфері є виділення Мел-частотних кепстральних коефіцієнтів (MFCC), які слугують компактним акустичним «відбитком» звуку [1].

У цій статті в основному використовуються наступні п'ять кроків для отримання MFCC-характеристик аудіо:

1. Попереднє підкреслення та сегментація кадру аудіосигналу;
2. Розрахунок спектра потужності кожного кадру;
3. Фільтрування за допомогою банку фільтрів Мела;
4. Дискретне косинусне перетворення (DCT);
5. Збереження коефіцієнтів 2 – 13 після DCT як характеристик звуку.

Процес вилучення MFCC показано на рис. 1 [1].

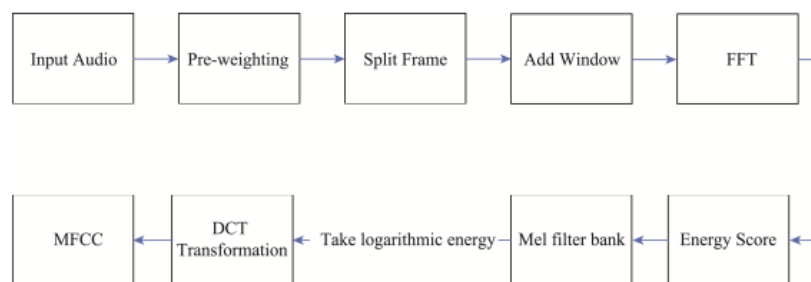


Рисунок 1 – Блок-схема характеристик MFCC

Алгоритм MFCC стискає ці дані у представлення, що імітує нелінійне сприйняття звуку людським вухом. Процес складається з таких послідовних кроків та обчислень.

Компенсує падіння енергії на високих частотах у людській мові за допомогою фільтра верхніх частот [1]

$$y[n] = x[n] - \alpha x[n-1], \quad (1)$$

де α типово дорівнює 0,97.

Сигнал розбивається на короткі кадри (20–40 мс). До кожного кадру застосовується віконна функція (вікно Хеммінга) для зменшення спектрального витоку

$$w[n] = 0,54 - 0,46 \cdot \cos\left(\frac{2\pi n}{N-1}\right). \quad (2)$$

Швидке перетворення Фур'є (FFT) – сигнал переводиться у частотну область

$$X[k] = \sum_{n=0}^{N-1} x[n](e^{-j2\pi kn/N}). \quad (3)$$

Гребінка Мел-фільтрів: частота конвертується у Мели за формулою

$$Mel(f) = 2595 \left(1 + \frac{f}{700}\right). \quad (4)$$

Спектр пропускається через набір трикутних фільтрів $H_m[k]$, нерівномірно розподілених за цією шкалою

$$M_m = \sum_k P[k]H_m[k]. \quad (5)$$

Практична реалізація: конвеєр розпізнавання ключових слів (KWS) Ефективність платформи підтверджується створенням системи розпізнавання команд («yes», «no», «stop», «go») на базі мікроконтролера STM32F429ZGT6. Після виділення MFCC-спектрограм дані подаються на легку CNN-архітектуру, що включає шари Conv2D, MaxPooling2D та повнозв'язні шари.

Аналіз за допомогою X-CUBE-AI показує надзвичайну ефективність. CMSIS-DSP надає ф-ції перетворення Фур'є та матричних множень, а Q15 дозволяє прискорити обчислення [3].

Обчислювальна складність складає 240 900 операцій множення з накопиченням (MACC) [1]. Модель займає лише 48.54 КБ Flash-пам'яті (з 1024 КБ доступних) та потребує 9.63 КБ оперативної пам'яті (RAM).

Точність розпізнавання на мікроконтролері сягає 90,58%, що практично відповідає результатам на ПК, див. рис. 2.

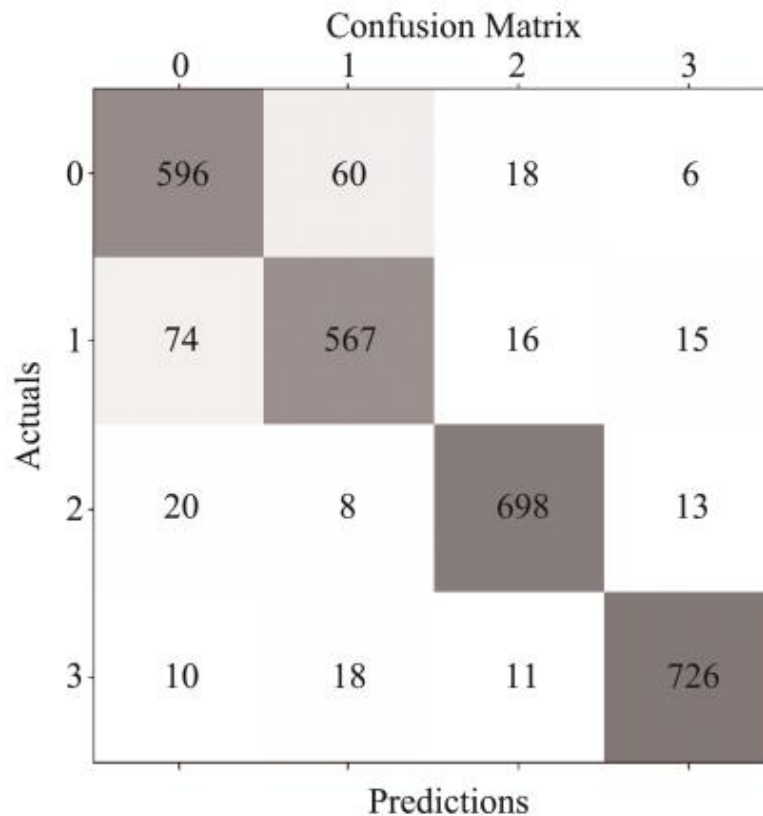


Рисунок 2 – Модель класифікації STM32F429 Матриця плутанини

Комбінація потужних мікроконтролерів (зокрема серій STM32F4 та H7), оптимізованих бібліотек обробки сигналів (CMSIS-DSP, PDM2PCM), засобів ШІ-конвертації (X-CUBE-AI) та цілеспрямованого спрощення аудіоознак дозволяє створювати надійні та енергоефективні системи Edge AI.

Список використаних джерел:

1. Kuang, W., Luo, W. Based on STM32 of CNN Speech Keyword Command Recognition System. Instrumentation. URL: <https://instrumentationjournal.com/index.php/instr/article/download/19/16/36> (дата звернення: 12.03.2026);
2. Vreča, J., Pilipović, R., Biasizzo, A. Hardware–Software Co-Design of an Audio Feature Extraction Pipeline for Machine Learning Applications. MDPI Electronics. URL: <https://www.mdpi.com/2079-9292/13/5/875> (дата звернення: 12.03.2026);
3. AN4841 Application note: Digital signal processing for STM32 microcontrollers using CMSIS. STMicroelectronics. URL: https://www.st.com/resource/en/application_note/an4841-digital-signal-processing-for-stm32-microcontrollers-using-cmsis-stmicroelectronics.pdf (дата звернення: 12.03.2026).