

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ РЕАЛІЗАЦІЇ ОБЧИСЛЕННЯ MFCC КОЕФІЦІЄНТІВ НА АПАРАТНІЙ ПЛАТФОРМІ STM32

студент Ткачов Д.В., доцент, к.т.н. Зубков О.В.,
Харківський національний університет радіоелектроніки,
кафедра мікропроцесорних технологій і систем, м.Харків, Україна
e-mail: denys.tkachov@nure.ua, oleh.zubkov@nure.ua

Abstract. The relevance of calculating cepstral coefficients on the STM32 hardware platform for subsequent recognition by neural networks is demonstrated. An algorithm for calculating the coefficients, requiring minimal RAM usage, was developed. The algorithm takes into account the features of the librossa library to achieve identical calculations in Python and C++ for STM32. The execution time of the algorithm blocks was measured and the possibility of calculating the MFCC in real time was proven.

Ключові слова: кепстральні коефіцієнти, мікроконтролер, STM32, нейронна мережа, алгоритм, DSP, час обчислення.

Вступ. Аналіз звукових сигналів і виділення інформативних ознак є базовою задачею у багатьох прикладних системах: розпізнавання мовлення та ключових слів (keyword spotting), класифікація звуків оточення, виявлення подій безпеки та біометрична ідентифікація [1]. Однією з найбільш поширених та перевірених у практиці характеристик для цих задач є набір Mel-Frequency Cepstral Coefficients (MFCC), який забезпечує компактне і інформативне представлення спектрально-перцептивних властивостей сигналу [2,3]. Оскільки дедалі більше застосунків переміщується на «граничні» пристрої (edge) — з міркувань енергоефективності, безпеки й латентності — виникає необхідність виконувати обчислення MFCC безпосередньо на мікроконтролерах з обмеженими ресурсами, зокрема на сімействі STM32. Це вимагає адаптацій алгоритму (спрощення фільтрбанків, фіксована арифметика, оптимізовані FFT/DFT реалізації) та використання апаратних оптимізацій (DSP-інструкції й бібліотеки), щоб забезпечити реальноремне виконання та прийнятне енергоспоживання на пристроях із обмеженою пам'яттю і тактовою частотою.

Поява все більшої кількості застосунків «завжди-увімк» (always-on) для голосових інтерфейсів, моніторингу довкілля та IoT-пристроїв робить критично важливим забезпечення стабільного й енергоефективного фронтенду для виділення ознак. Сучасні дослідження показують практичні підходи до спрощення та апаратно-специфічної оптимізації MFCC для вбудованих платформ, а також демонструють, що при розумних компромісах точності можна домогтися прийнятної продуктивності на MCU класу Cortex-M. Крім того, наявні програмні комплекти (наприклад, ARM CMSIS-DSP) і приклади реалізацій для STM32 дозволяють

прискорити розробку та підвищити ефективність обчислень при збереженні коректності спектральних перетворень.

Метою цього дослідження є розробка та апробація ефективної реалізації обчислення MFCC-коефіцієнтів на платформах STM32 з урахуванням апаратних обмежень (об'єм RAM/FLASH, пропускна здатність ADC та енергоспоживання). Крім того обчислення коефіцієнтів MFCC є першим кроком в задачах розпізнавання акустичних сигналів і оцінка часу обчислення коефіцієнтів, а також ресурсів на реалізацію обчислення є дуже важливим з точки зору подальшої інтеграції з нейронними мережами.

Основна частина.

Перед обчисленням MFCC коефіцієнтів аналоговий сигнал дискретизується у часі з частотою f_s та перетворюється у кодові комбінації за амплітудою. Потік відліків розбивається на перекриваючі кадри з розміром N відліків. Сам алгоритм обчислення коефіцієнтів MFCC складається з шести етапів [3]. Спочатку, для зменшення спектральних витіків виконується віконна обробка кадра (наприклад, вікно Хеммінга), тобто відліки аналогового сигналу помножуються на значення віконної функції $w[k]$ за формулою 1. Математично ця операція визначається формулою 2.

$$w[k] = 0.54 - 0.46 \cos\left(\frac{2\pi k}{N-1}\right), k = 0 \dots N-1 \quad (1)$$

$$s[k] = y_{frame}[k] \cdot w[k] \quad (2)$$

Стандартна DSP бібліотека для мікроконтролерів STM32 не містить вбудованих функцій віконної обробки. Тому реалізація віконної обробки – це послідовне помноження відліків сигналу на відліки вікна у циклі. На другому етапі виконується перетворення Фур'є (FFT). Реалізація FFT мінімізує час обчислень та витрати оперативної пам'яті. В результаті отримуємо масив спектральних компонент $X[k], k = 0 \dots N_{FFT} - 1$, а після цього обчислюється спектральна потужність

$$P[k] = \frac{|X[k]|^2}{N_{FFT}} \quad (3)$$

Стандартна DSP бібліотека містить вбудовану оптимізовану за швидкістю функцію перетворення Фур'є `arm_cfft_f32`, що видає комплексні значення на виході. За цими значеннями у циклі реалізується обчислення потужності $|X[k]|^2$. Наступний етап – це обчислення шкали частот Фішера за формулою

$$mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad (4)$$

Далі для шкали задаються $M+2$ точок, що відповідають M трикутним фільтрам, а за ними розраховуються нормалізовані значення частот (біни)

$b_i = \lfloor (N_{FFT} + 1)f_i/f_s \rfloor$ і формуються імпульсні характеристики трикутних фільтрів $m=1 \dots M$ за формулою 5.

$$H_m(k) = \begin{cases} 0, k < b_{m-1} \\ \frac{k-b_{m-1}}{b_m-b_{m-1}}, b_{m-1} < k < b_m \\ \frac{b_{m+1}-k}{b_{m+1}-b_m}, b_m < k < b_{m+1} \\ 0, k > b_{m+1} \end{cases} \quad (5)$$

За допомогою прямого фільтруючого підходу обчислюється енергія сигналу на виході кожного фільтра

$$S_m(k) = \sum_{k=0}^{N_{FFT}/2} P[k] \cdot H_m(k) \quad (6)$$

Реалізація трикутних фільтрів у стандартній DSP бібліотеці відсутня. Тому для зменшення вимог до використання оперативної пам'яті розрахунок імпульсних характеристик проводиться заздалегідь та їх значення розміщуються у FLASH пам'яті. Так, як у STM32 читання FLASH пам'яті відбувається за 1 такт, то зменшення швидкодії не відбувається.

На останньому етапі виконується логарифмування значень енергії та обчислення дискретного косинусного (DCT) перетворення для розрахунку значень K коефіцієнтів MFCC c_n .

$$c_n = \sum_{m=1}^M S(m) \cdot \cos\left[\frac{\pi n}{M}(m - 0.5)\right], n = 0..K - 1$$

На цьому етапі використовується оптимізована функція бібліотеки DSP.

Результати досліджень.

Практична реалізація накопичення відліків аналогового сигналу після АЦП вимагає мінімального буфера розміром $N+N_{\text{пер}}$, де $N_{\text{пер}}$ – кількість відліків перекриття кадрів. Сам процес накопичення оптимізується за часом завдяки використанню кільцевого буфера та вбудованого DMA контролера, що заповнює цей буфер паралельно з роботою ядра процесора.

Для зменшення використаної оперативної пам'яті значення відліків віконної функції та імпульсних характеристик фільтрів розміщуються у флеш пам'яті, що дозволяє економити $2N$ комірок оперативної пам'яті, а, з урахуванням, що всі операції виконуються з числами з плаваючою комою, економія складає $8N$ байт. Додаткові вимоги до флеш пам'яті несуттєві, бо стандартні розміри програми для контролерів серії STM32F4 не перевищують 128кбайт, а контролери мають від 512кбайт флеш пам'яті.

Використання бібліотеки DSP для FFT та DCT передбачає, що вхідні та вихідні масиви повинні містити числа з плаваючою комою, що потребує по 4 байти на кожний елемент масиву. Використання обчислень з цілими числами виявило формування суттєвих похибок обчислень коефіцієнтів MFCC до 11%, особливо у випадку слабких акустичних сигналів.

Так, як для навчання сучасних нейронних мереж використовується мова Python, а для обчислення коефіцієнтів MFCC застосовується бібліотека librossa, було виконано перевірку результатів обчислення коефіцієнтів MFCC у контролері та з використанням бібліотеки librossa. Виявлена суттєва розбіжність між цими результатами була пов'язана з відмінністю алгоритма бібліотеки librossa у формулі (3), бо ділення відбувалось на N_{FFT}^2 . Після усунення розбіжності результатів обчислення коефіцієнтів MFCC було виконано вимірювання часу виконання основних операцій. При вимірюваннях використовувалась частота дискретизації 44100Гц та розмір кадра 1024 відліка, що відповідає часу 23мс. Результати вимірювань наведені у таблиці 1

Таблиця1 – Значення часу виконання основних операцій обчислення MFCC

Операція	$s[k]$	$P[k]$	c_n
Час обчислення, мс	0,3	0,61	4,2

Висновки. Реалізований алгоритм обчислення коефіцієнтів MFCC на мікроконтролері STM32F407 дозволяє виконувати всі операції у реальному часі при тактовій частоті процесора 168МГц. Алгоритм відповідає стандартній бібліотеці librossa, що широко використовується у сукупності з алгоритмами навчання нейронних мереж. Це дозволяє гарантувати вірну роботу вбудованих нейронних мереж у контролерах STM32 при їх попередньому навчанні на ПК.

Список використаних джерел

1. Gupta N., Ahlawat P. A Review of Audio Classification Techniques. 2024 International BIT Conference. Dhanbad, India. 07-08 December 2024. P. 1–16.
2. Wang, M., Chu, Z., Entzminger, C., Ding, Y. and Zhang, Q. Visualization and Interpretation of Mel-Frequency Cepstral Coefficients for UAV Drone Audio Data. 13th International Conference on Data Science, Technology and Applications., Dijon, France, 9 July 2024. P. 528-534
- Hosseinzadeh M, Haider A, Malik MH, Adeli M, Mzoughi O, Gemeay E, et al. Enhanced heart sound classification using Mel frequency cepstral coefficients and comparative analysis of single vs. ensemble classifier strategies. PLoS ONE. 2024. Vol.19, Iss.12. P. 1–17.