

МОДЕЛЬ СИСТЕМИ РОЗПІЗНАВАННЯ МОВИ В СЕРЕДОВИЩІ MATLAB

Процес розпізнавання мови включає наступні етапи: запис і оцифрування мови, обчислення часових і спектральних характеристик зразків, класифікацію часових фреймів по фонетичних категоріям, підбор відповідних категорій слів, вимір ступеня довіри до ухваленого рішення.

При розпізнаванні використовуються дві групи методів: параметричні, спрямовані на математичне перетворення мовного сигналу з виділенням і стабілізацією основних інформативних ознак (перетворення Фур'є, цифрова фільтрація, вейвлет-аналіз і інш.); лінгвістичні, ціль яким контекстна обробка висловлення (динамічне програмування, сховане марковське моделювання).

Вихідною інформацією для лінгвістичної обробки служать результати параметризації мовного сигналу.

Розглянемо проектування модуля розпізнавання ізольованих слів, що дозволяє перетворювати мову на вході в текст на виході. Як платформа для розробки обрана комплексна система MATLAB, що надає широкі можливості по обробці мовних сигналів і проведенню трудомістких математичних обчислень.

Структурна система модуля розпізнавання ізольованих слів, що використовує функціональний метод, наведена на рисунку 1.

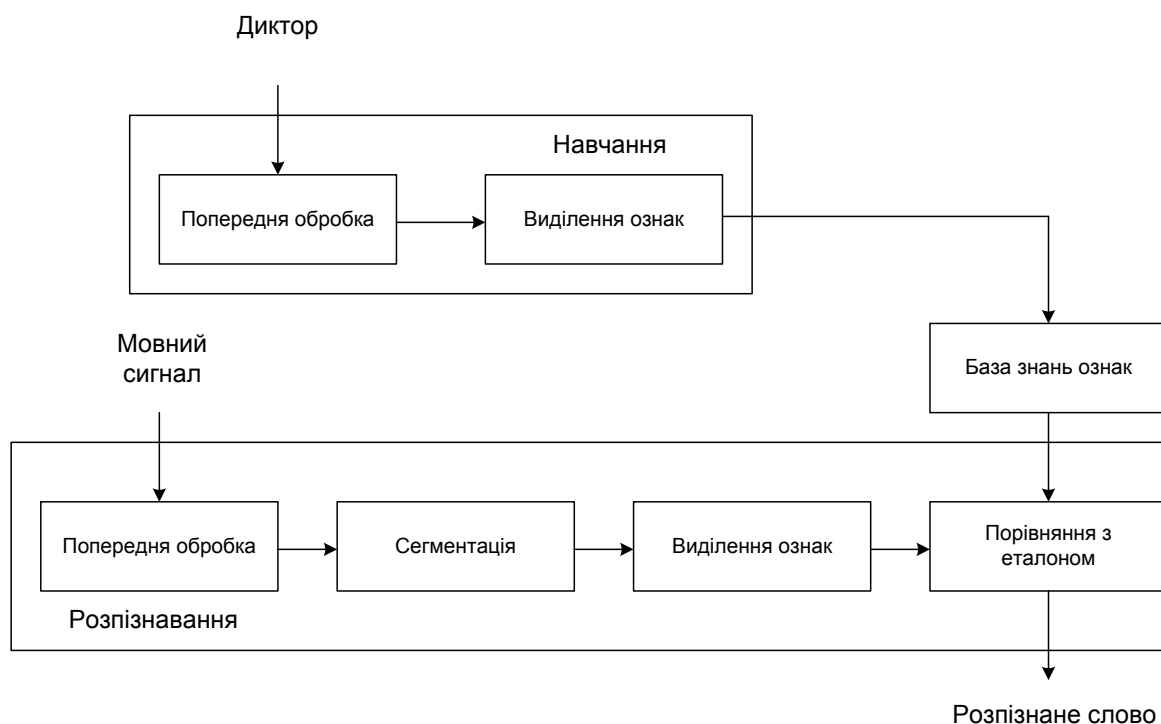


Рис. 1. Загальна схема системи розпізнавання мови

Попередня обробка. На етапі попередньої обробки виконується фільтрація сигналу. Віддаляється шум високочастотних складових спектра сигналу. При використанні вейвлетів можна обмежити рівень коефіцієнтів, що деталізують.

Короткочасні особливості сигналу, а до них можна віднести й шуми у вигляді безлічі таких особливостей, створюють коефіцієнти, що деталізують, з високим змістом шумових компонентів, що мають більші випадкові викиди значень сигналу. Задавши деякий поріг для їх рівня й зрізавши за рівнем коефіцієнти, що деталізують, можна зменшити рівень шумів.

Для автоматичного видалення шумів у системі MATLAB існує функція $wden(X, TPTR, SORH, SCAL, N, 'wname')$, що повертає очищений від шуму сигнал, отриманий обмеженням вейвлет-коефіцієнтів перетворення вхідного сигналу X .

$TPTR$ – задає правило вибору порога:

- *'heursure'* – евристичний варіант алгоритму Штейну незміщеної оцінки ризику;
- *'sqtwolog'* – для інверсного порога $\sqrt{2 \cdot \log(\cdot)}$;
- *'minimaxi'* – для мінімального/максимального порога.

$SORH$ (*'s'* або *'h'*) – для гнучкого або твердого порога відповідно.

$SCAL$ – визначає мультиплікативне граничне перемасштабування (якщо шум поза межами $[0,1]$ або не "білий").

N – рівень вейвлет-розкладання сигналу.

'wname' – ім'я ортогонального вейвлета.

Найкраща фільтрація досягається застосуванням гнучкого інверсного порога. Кращої ж вейвлет-функцією з тестуємих є вейвлет Добеши-4 при 8 рівнях вейвлет - розкладання.

Сегментація. Сегментація мовного сигналу має на увазі виділення ділянок сигналу, що відповідають окремим структурним одиницям мовного сигналу. Якщо як такі одиниці розглядати фонем, то завдання сегментації зводиться до виявлення міжфонемних переходів. У рамках традиційних підходів рішення цього завдання досить проблематично.

Однак вейвлет-перетворення дозволяє вирішити цю проблему, принаймні, для фонем, що відповідають порівняно протяжним квазістаціонарним ділянкам мовного сигналу.

На міжфонемних переходах сигнал перетерплює значні зміни відразу на багатьох масштабах дослідження, і, відповідно, характеризується зростанням вейвлет-коефіцієнтів для багатьох рівнів деталізації, у той час, як на стаціонарних ділянках фонем вейвлет-коефіцієнти виявляються згрупованими поблизу певних масштабів.

Таким чином, відшукування міжфонемних границь може бути зведене до відшукування моментів збільшення вейвлет-коефіцієнтів на значній кількості рівнів масштабування.

Виділення ознак. Як ознаки, що характеризують мовний сигнал, а саме виділені в результаті сегментації фонем, обрані коефіцієнти деталізації ортогонального вейвлет-перетворення кожного із сегментів.

Функція *wavedec* системи MATLAB виконує багаторівневий одномірний вейвлет-аналіз. Її варіант $[C,L]=wavedec(X,N,'wname')$ повертає вектори вейвлет-розкладання сигналу X на рівні N , використовуючи вейвлет *wname*. Далі, застосовуючи функцію *detcoef* у вигляді $DCELL=detcoef(C,L,'CELLS')$, одержуємо масив коефіцієнтів, що деталізують, на всіх рівнях розкладання від 1 до N .

У процесі попереднього навчання створюється база знань еталонів – фонем. У якості вейвлет-базису використаний вейвлет Добеши-8. База знань організована у вигляді *mat-файлу*, що зберігає коефіцієнти, що деталізують, і завантажується в робочу область при кожному сеансі роботи з розпізнавачем.

Порівняння з еталоном. У ході процесу розпізнавання здійснюється ухвалення рішення про відповідність даного сегмента мовного сигналу еталону фонем з бази. У цей час найбільш успішні алгоритми відповідності використовують як основний інструмент динамічне програмування. Алгоритм динамічного програмування є ефективною обчислювальною процедурою визначення оптимальної траєкторії в безлічі кінцевих станів. Також даний алгоритм може бути використаний для знаходження можливого найкращого тимчасового вирівнювання між зразком невідомого висловлення й еталонним зразком. При розробці системи застосований алгоритм Витерби, що є різновидом динамічного програмування.

На початковому етапі створюється БЗ еталонів, що містить 5 фонем: А, О, У, К, Р. Були проведені експерименти з оцінки якості розпізнавання ізольованих слів. Контрольні слова вимовляються тим же диктором, що робив навчання системи.