

РОЗРОБКА І ДОСЛІДЖЕННЯ МОДУЛЯ ЕФЕКТИВНОГО РОЗПІЗНАВАННЯ ЗВУКІВ В СИСТЕМАХ ГОЛОСОВОГО УПРАВЛІННЯ

Биков Микола, канд. техн. наук, доцент, професор кафедри
комп'ютерних систем управління,
Алексєєнко Анна, студентка групи 2АКІТ-16м (іт),
Вінницький національний технічний університет, Україна

Традиційна модель автоматичного розпізнавання мови припускає, що шляхом відстеження акустичних параметрів і застосування одного із засобів пошуку по набору еталонів звукових сегментів можна встановити фонематичні ряди. Потім ці ряди можуть застосовуватися для виділення слів, фраз і сенсу висловів. При автоматичному розпізнаванні мови великі труднощі представляють собою процеси виявлення та ідентифікації деяких груп фонотипів. Тому для більш ефективного розпізнавання голосових команд запропоновано розподілити цей процес на декілька етапів.

Першим етапом є сегментація на склади, тобто кожне слово чи група слів спочатку поділяється на менші одиниці мови — склади. Такий підхід дозволяє застосувати ієрархічну стратегію розпізнавання, тобто спочатку намагатися проводити розпізнавання на складовому рівні, а потім, при необхідності, переходити на рівні інших фонотипів і одночасно зберігати в пам'яті параметри (тривалість, енергію тощо), які в подальшому будуть використовуватися. На відміну від високоенергетичних щільових звуків мови, енергія яких зосереджена в високочастотному діапазоні мовного сигналу, основним параметром для визначення складового сегмента є енергія в більш низькочастотному діапазоні. Тому за ознаку виділення складів у мовному потоці вибрано кореляцію енергій в перших двох формантних діапазонах. Наступним і кінцевим етапом є розрахунок інтегральної міри подібності між цілими словами. Рішення про розпізнавання приймається за мінімумом відстані між реалізацією і однією з еталонних стрічок [1]. В покращених алгоритмах для рядків однакової довжини обчислювальна складність має оцінку $O(n^2/\log n)$.

Для зменшення обчислювальної складності, а отже, підвищення швидкості розпізнавання, автори даній роботі запропонували метод співставлення з опорою на реперні (надійні) точки, в якості яких використовуються центри складів. Ідеєю цього методу є розбиття матриці відстаней на сукупність прямокутних підматриць, розмір яких визначається розміщенням опорних точок в еталонній послідовності звуків слова і в послідовності, що розпізнається. Це дозволяє отримати значний вииграш у складності, яка в даному випадку буде оцінюватися, як $O(m+k)$, де m — кількість реперних точок, k — розмір підматриці.

Аналіз складової інформації здійснюється згідно алгоритму надійної сегментації неперервного мовного сигналу на склади, запропонованого авторами раніше, і виділення таких ознак складів, як їх кількість у слові,

тривалість і часові інтервали центрів складів. Використання для фонетичного опису параметрів сигналу в центрах складів дозволяє мінімізувати контекстну варіацію і, таким чином, створити надійні опорні точки для процедури порівняння еталонних описів і описів фрагментів вхідної мови. Іншим недоліком методу динамічного програмування є те, що для визначення вказаної відстані потрібно, окрім кодів марок, зберігати ще й коди відстаней між ними, що збільшує апаратні ресурси і зменшує швидкість розпізнавання. В роботі [2] для усунення цього недоліку запропонований метод розпізнавання, що ґрунтується на представленні марок різницевиими кодами, які усувають необхідність запам'ятовування кодів відстаней.

В роботі авторами описується модуль розпізнавання, оснований на методі динамічного програмування з використанням представлення марок кодами, що зберігають ранги відстаней (DRP-codes) [3]. DRP — кодом є відображення $i \rightarrow B_i$ множини $M = \{1, 2 \dots m\}$ в множину $\{0, 1^n\}$ двійкових послідовностей довжини n таке, що

$$\forall_{i,j} (R(d_{ij})=r \Rightarrow R(h_{ij})=r), r=1, m_r, i, j \in M. \quad (1)$$

У виразі (1) $R(d_{ij})$ — ранг відстаней d_{ij} між об'єктами i і j в просторі об'єктів; $R(h_{ij})$ — ранг відстані h_{ij} в просторі двійкових кодів; r — ціле число, конкретне значення рангу; m_r — максимальна величина рангу. Далі під кодованими об'єктами розумітимемо марки фонотипів.

Відстань між марками, закодованими DRP-кодом, під час їх порівняння визначається за допомогою порозрядної логічної операції “Г”, однак при цьому можуть виникати помилки, зумовлені невиконанням аксіоми ідентичності під час визначення відстані між однаковими марками. Тому запропоновано модифікувати дану логічну операцію так, щоб для неї виконувалась аксіома ідентичності. Для цього сформульовано математичний опис операції та розроблена відповідна логічна схема, яка скидає результат порівняння однакових марок до “0”. Модифікована операція була названа логічною операцією ““Г” з самоблокуванням”.

Таким чином, використання в розробленому модулі розпізнавання звуків методу динамічного розпізнавання з опорою на реперні точки і їх маркування кодами, що зберігають ранги відстаней, дозволило підвищити надійність і швидкість розпізнавання та розширити склад словника мовних образів.

Список використаної літератури

1. Ли У. Методы автоматического распознавания речи: в 2-х книгах. Пер. с англ. / Под ред. У. Ли. — М.: Мир, 1983. — Кн. 1. 328 с.
2. Glave R.D. The David speech recognition system // Glave R.D., Vander Giet. - Proc. IEEE Int. Conf. ASSP.-Tulsa, 1978, p.429-432.
3. Биков М. М. Універсальний метод представлення інформації в інтелектуальних еволюційних системах // Биков М. М. / Відбір і обробка інформації. - 2006.- Вип. 24(100). С. 35-42.