

ПОРІВННЯ МІР ПОДІБНОСТІ ФРАГМЕНТІВ МОВНИХ СИГНАЛІВ

© Ткаченко В.С., 2010

Розглянуто деякі міри подібності фрагментів мовних сигналів, які використовуються під час їх розпізнавання. Розроблено алгоритм та комплекс програм для обробки сигналів, наведено результати роботи програм.

Ключові слова – міра подібності, мовний сигнал, алгоритм.

Some measures of similarity of speech signals fragments, which are used for their recognition are considered. The algorithm and program complex for signal processing are developed.

Keywords – measures of similarity, speech signal, algorithm.

Вступ

Для будь-яких методів розпізнавання мовних сигналів основним є встановлення міри подібності виділеного (на основі попередньої сегментації [1]) фрагмента сигналу з деяким сигналом, прийнятим за еталон, або з сукупністю еталонних сигналів та віднесення на основі цієї міри сигналу до відповідного класу [2, 3]. Нижче розглянуто деякі міри подібності, які використовуються в методах та алгоритмах розпізнавання мовних сигналів, запропоновано алгоритм порівняння та проведено порівняння окремих мір щодо їх ефективності для задач розпізнавання мовних сигналів. Крім того, ефективність використання міри подібності визначається також обсягом обчислювальних затрат, оскільки переважно оброблення та розпізнавання мовних сигналів повинні здійснюватися в реальному часі.

1. Існуючі міри подібності мовних сигналів та їх порівняння

Розглядаються два фрагменти мовних сигналів, один з яких $x[i]$, $i = \overline{1, n}$, отриманий за попередньої сегментації мовного потоку, а інший $y[i]$ є деяким еталоном, з яким порівнюється виділений сигнал. Для забезпечення можливості порівняння сигнали попередньо нормуються в такий спосіб, щоб їх питома енергія була однаковою і дорівнювала 1: $\sum_{i=1}^n x^2[i] = \sum_{i=1}^n y^2[i] = 1$. Як

міри подібності можуть бути використані як універсальні міри, придатні для сигналів різних типів, так і міри, які переважно використовуються під час оброблення та порівнянні мовних сигналів завдяки певним характерним особливостям цих сигналів [4]. Розглянемо деякі з них.

1.1. Сума квадратів відхилень

Ця міра визначається співвідношенням $\Delta_1 = \sum_{i=1}^n (x[i] - y[i])^2$ і знаходиться в межах:

$0 \leq \Delta_1 \leq 2$. Внаслідок коливального характеру мовних сигналів ця міра є ефективною лише у тому разі, якщо частоти основного тону порівнюваних сигналів збігаються і проводиться попередня обробка “вирівнювання фаз” сигналів.

1.2. Коефіцієнт кореляції

Визначається співвідношенням $\Delta_2 = \sum_{i=1}^n x[i] \cdot y[i]$. Аналогічно до попереднього випадку

ефективна лише за таких самих умов.

1.3. Спектральна міра

Ця міра будується на основі коефіцієнта кореляції між модулями спектрів сигналів з попереднім використанням віконної функції, наприклад, функції Хеммінга. Використання цієї міри пов'язане з моделями формування мовних сигналів, насамперед сигналів, що відповідають голосним звукам, спектри яких мають формантну структуру. При цьому фазові співвідношення сигналів не відіграють істотної ролі.

1.4. Спектрально-часова невизначеність

Визначається як добуток тривалості сигналу на ефективну ширину його спектра: $\Delta t \cdot \Delta f$. Ця міра є ефективною лише для фрагментів малої тривалості, насамперед таким, що відповідають голосним звукам (один квазіперіод сигналу), і в деяких випадках є доповненням до спектральної міри.

1.5. Міра на основі кепстральних коефіцієнтів

Ця міра визначається гомоморфною обробкою сигналів, є специфічною для мовних сигналів внаслідок врахування моделей їх формування і широко використовується в системах розпізнавання мовних сигналів на основі нейромережових технологій [4].

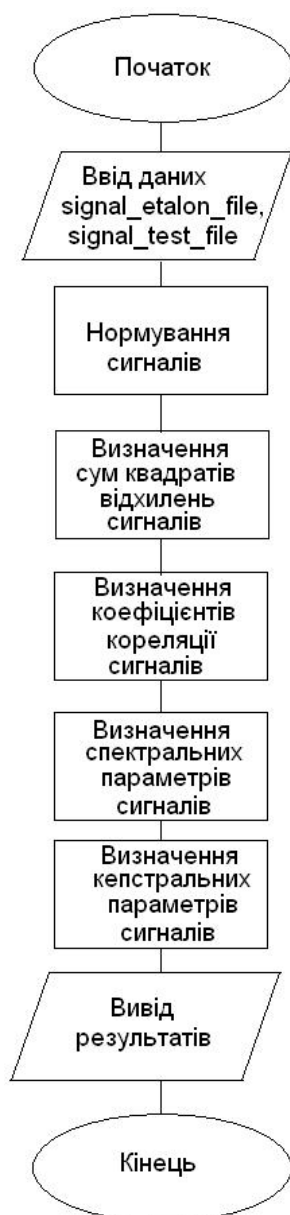


Рис. 1. Блок-схема алгоритму обробки мовних сигналів

1.6. Міра Махаланобіса

Дає змогу визначити міру подібності фрагмента мовного сигналу до сукупності еталонних сигналів, зокрема, з врахуванням можливості побудови множин різних еталонів для однакових елементів мовних сигналів [5]. Ця міра використовується в теорії нейронних мереж, відповідна функція є в пакеті прикладних програм Neural Network Toolbox системи MATLAB. Нехай вектори μ_i та μ_j визначають середні значення векторів x_i та x_j відповідно, тобто $\mu_i = E[x_i]$, де E – оператор математичного сподівання. Для вимірювання відстані між двома множинами можна використати відстань Махаланобіса, квадрат якої визначається співвідношенням

$$d_{ij}^2 = (x_i - \mu_i)^T \Sigma^{-1} (x_j - \mu_j),$$

де Σ^{-1} – матриця, обернена до матриці коваріації Σ . Припускається, що матриця коваріації для обох множин одна й та сама, тобто

$$\Sigma = E[(x_i - \mu_i)(x_i - \mu_i)^T] = E[(x_j - \mu_j)(x_j - \mu_j)^T].$$

У частковому випадку, коли $x_i = x_j$, $\mu_i = \mu_j = \mu$ і $\Sigma = I$, де I – одинична матриця, відстань Махаланобіса вироджується в Евклідову відстань між вектором x_i і вектором математичного сподівання μ .

2. Алгоритм обробки мовних сигналів та результати роботи програми

Для реалізації алгоритму обробки мовних сигналів з використанням вищезазначених мір вибрано систему моделювання MATLAB. Вона дає можливість ефективно керувати даними та отримувати результати у графічному вигляді. Блок-схему роботи алгоритму зображено на рис. 1.

Як вхідні дані використовуються тестовий та еталонний сигнали, пошук яких здійснюється у тестовому.

При знаходженні фрагмента, найближчого до еталону, виконувалися визначення мір подібності, що дало такі результати:

сума квадратів відхилень $\Delta_1 = 4.35 \times 10^{-4}$, коефіцієнт кореляції $\Delta_2 = 0.2137$, спектральні характеристики (рис. 2) та кепстральні характеристики (рис. 3).

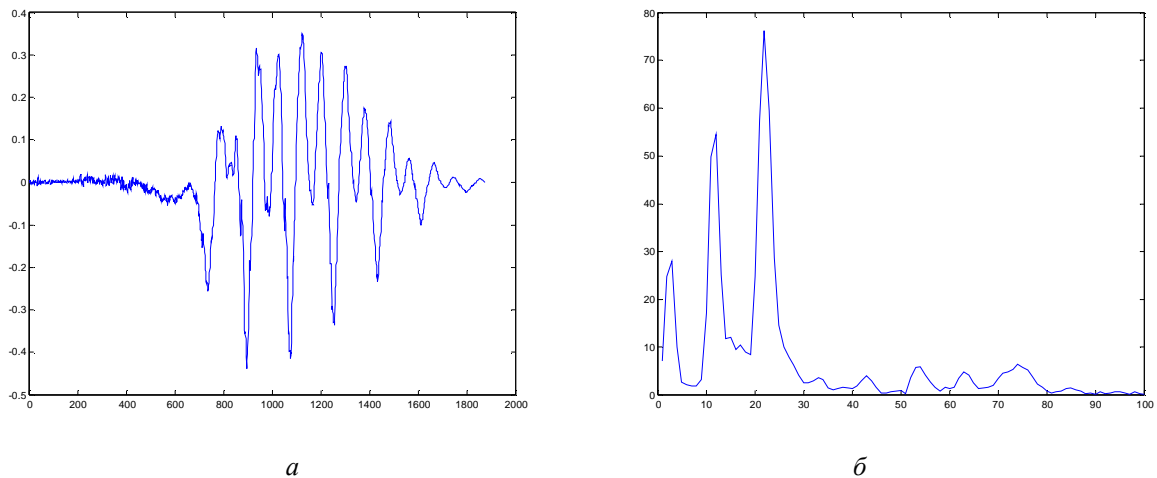


Рис. 2. Вигляд еталонного сигналу:

а – еталонний сигнал, домножений на вікно Хемінга; б – спектр еталонного сигналу

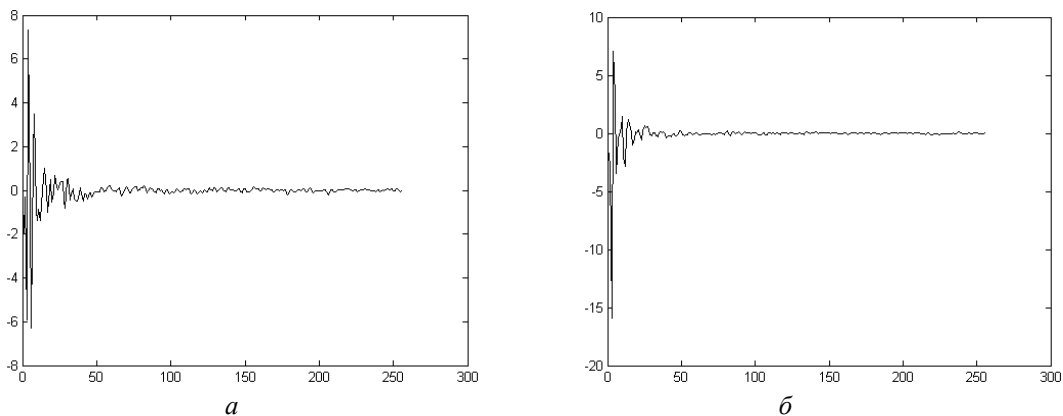


Рис. 3. Кепстри сигналів:

а – кепстр еталонного сигналу; б – кепстр фрагмента сигналу, найближчого до еталонного

Висновки

Складність мовних сигналів, з точки зору їх автоматизованого розпізнавання, особливо в реальному часі, зумовлює необхідність використання комбінованих методів, які ґрунтуються на врахуванні властивостей елементів мовних сигналів різного характеру. При розпізнаванні цих елементів після попередньої сегментації доцільно використовувати різні міри подібності виділених фрагментів еталонним сигналам, враховуючи як ефективність цих мір для конкретних еталонів, так і обчислювальні витрати під час їх застосування запропонованим методом.

1. Каркульовський В.І., Ткаченко В.С. Особливості методів сегментації мовленнєвих сигналів // Вісн. Нац. ун-ту «Львівська політехніка». «Комп'ютерні системи проектування. Теорія і практика». – 2009. – № 651. – С. 141–144. 2. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – К.: Наук. думка, 1987. – 264 с. 3. Фланган Дж. Анализ, синтез и восприятие речи: пер. с англ. / под ред. А.А. Пирогова. — М.: Связь, 1968. — 396 с. 4. Romanyshyn Y., Tkachenko V. Comparison of similarity measures of speech signals fragments //IEEE MEMSTECH'2010. – Polyana, 2010. – P.184. 5. Хайкин С. Нейронные сети: полный курс; 2-е изд. / пер. с англ. – М.: Издательский дом «Вильямс», 2006. – 1104 с.