УДК 519.688

Д.т.н., професор Зайцев В.Г., студент Долженко Д.А.

Національний технічний університет України
«Київський політехнічний інститут»

ОСНОВНІ СКЛАДОВІ ПРОЦЕСУ РОЗПІЗНАВАННЯ МОВНОЇ ІНФОРМАЦІЇ

Abstract

Volodymyr G. Zaitsev, prof., Doctor of Science; Denis Dolzhenko, student
Main components the process speech information recognition

This paper concerns the task of recognition of speech information. The paper contains the developed structural block diagram of speech recognition system and the brief description of each functional unit. The main modern methods of recognition and analysis of speech data are considered.

Вступ

Задача розпізнавання мовної інформації є складною задачею, яка використовує такі області науки як цифрова обробка сигналів, розпізнавання образів та лінгвістика. На сьогоднішній день задача розпізнавання мовної інформації в загальному вигляді так і залишається не вирішеною, а найкращі окремі рішення цієї задачі не є публічними.

Постановка задачі

1. Задача полягає в побудові схеми процесу розпізнавання мовної інформації в загальному вигляді, яка б могла навчатись та перенавчатись.
2. Для кожного етапу процесу розпізнавання визначити сучасні методи вирішення конкретної задачі (аналіз акустичного сигналу, його трансформація та розпізнавання).

Термінологія

Розпізнавання мови - процес перетворення мовного сигналу в текстовий потік.

Кепстр – це результат перетворення Фур'є спектра мовного сигналу f(n). Математично кепстр потужності сигналу визначається як

$$K = | F\{\log(| F\{f(n)\}|^2)\}|^2$$

де *F*- символ перетворення Фур'є.

Дискретне косинусне перетворення (ДКП) – одне з ортогональних перетворень дискретного мовного сигналу f(n), варіант косинусного перетворення для вектора дійсних чисел [5].

Прихована Марківська модель (ПММ) – статистична модель, що імітує роботу процесу, схожого на Марківський процес з невідомими параметрами, і завданням ставиться розпізнавання невідомих параметрів на основі спостережуваних (див. рис 1). Отримані параметри використовуються для розпізнавання образів [4].

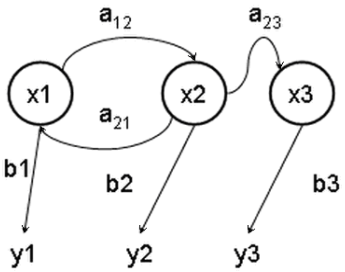

Рис. 1

x_i - прихований стан (мовний кадр або фонема);
y_j - спостережуваний результат (спостережуваний сигнал, який треба розпізнати);
a_{ij} - ймовірність переходу (ймовірність того, що після мовного кадру або фонеми x_i йде мовний кадр або фонема x_j);
b_i – достовірність результату (достовірність того що y_i є відображенням x_i).
В процесі навчання в статистичній моделі змінюються її параметри a_{ij} та x_i, які впливають на розпізнавання образів.

Основні складові процесу розпізнавання мовної інформації

Загальний процес розпізнавання мовної інформації можна подати як показаний на рис 2.

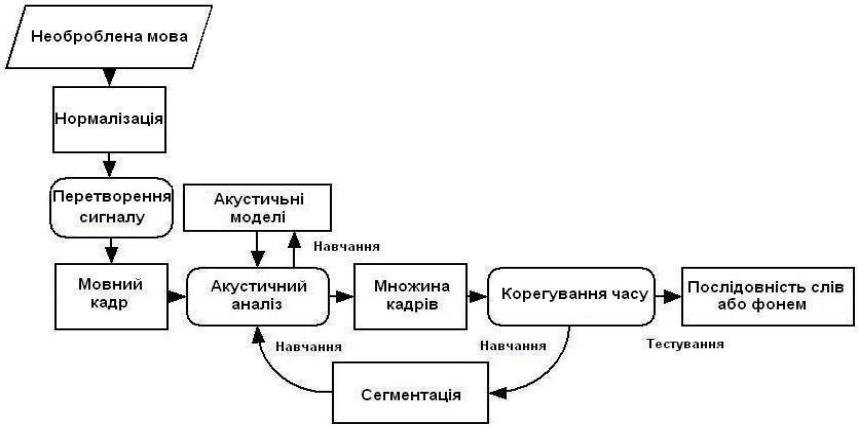

Рис. 2. Основні складові процесу розпізнавання мовної інформації:

• *Необроблена мова*. Аналоговий мовний сигнал з максимальною частотою 20 кГц при запису з мікрофону, або 8 кГц при запису з телефонної лінії (див. рис. 3).

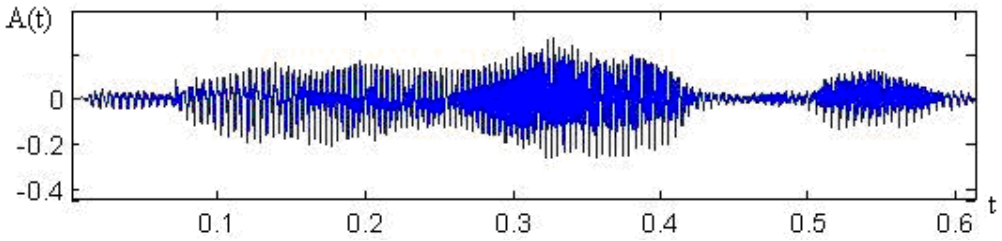

Рис. 3

Необроблений мовний сигнал.

• *Нормалізація*. Вирівнювання рівня сигналу до певної величини. Це дозволяє зменшити похибки розпізнавання, пов'язані з тим, що диктор може вимовляти слова з різним рівнем гучності.

• *Перетворення сигналу*. Нормалізований сигнал повинен бути спочатку перетворений та “ущільнений” для спрощення (з точки зору швидкодії) подальшої обробки (див. рис. 4). Ущільнення відбувається за рахунок фільтрації дискретного сигналу, тобто відкидаються високочастотні гармоніки спектру.

Найбільш поширені методи перетворення мовного сигналу:

1. Аналіз Фур'є (ДКП).
2. Лінійне передбачення мови [3].
3. Кепстральний аналіз.

- *Мовні кадри*. Результатом перетворення сигналу є послідовність мовних кадрів. Кожен мовний кадр - це результат перетворення сигналу на визначеному відрізку часу. Для поліпшення якості розпізнавання в кадри може бути додана інформація про спеціальні додаткові ознаки сигналу, наприклад, частота основного тону.

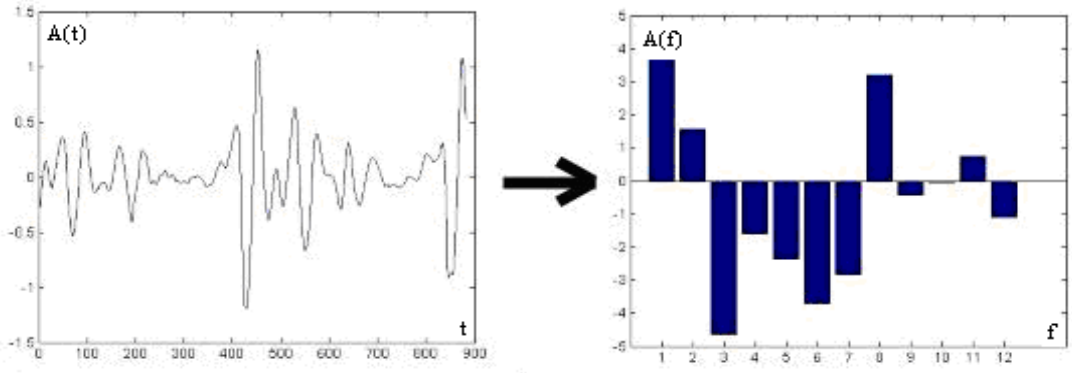

Рис. 4

Перетворення Фур'є(ДКП) мовного сигналу.

- *Акустичний аналіз.* На етапі акустичного аналізу виконується порівняння мовних кадрів з еталонними моделями кадрів відповідного класу, які отримуються під час навчання.

Найбільш поширеними методами та засобами порівняння є:

1. Кореляційний метод.
2. Нейроні мережі.
3. Прихована Марківська модель.

- *Корегування часу.* Використовується для обробки часової варіативності, що виникає при вимові слів (наприклад, "розтягування" або "з'їдання" голосних звуків).

- *Послідовність слів або фонем*. В результаті розпізнавання команди видається послідовність слів або фонем, які найбільш точно відповідають вхідному потоку мови.

Висновки

Наведена схема процесу розпізнавання мовної інформації дозволяє використати існуючі методи попередньої обробки мовного сигналу (див. найбільш поширені методи перетворення мовного сигналу), та розпізнавання образів для розпізнавання мовної інформації. Особливість цієї схеми полягає у тому, що вона здатна перенавчатись (див. рис 2), тобто запам'ятовує нові акустичні моделі, які в подальшому порівнюються з мовними кадрами. В результаті цього з'являється можливість зміни мовних команд при новому конкретному застосуванні.

Література

1. Huang Xuedong. Spoken language processing: a guide to theory, algorithm and system development. –New Jersey: Prentice Hall PTR, 2001, 910 p.
2. McAllaster, D., Gillick, L., Scattone, F. and Newman, M. Fabricating conversational speech data with acoustic models: A program to examine model-data mis-match // *ICSLP-98*. Sydney, 1998. Vol. 5, pp. 1847–1850.
3. Маркел, Джон Д.; Грэй, Августин. Линейное предсказание речи /пер. с англ. под ред. Ю.Прохорова, В.С. Звездина - М.:Мир,1983. – 308 с., ил.
4. Xuedong Huang, M. Jack, and Y. Ariki (1990). Hidden Markov Models for Speech Recognition. Edinburgh University Press. ISBN 0748601627., 276 p.
5. N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform", IEEE Trans. Computers, Jan 1974, pp. 90-93.