дифференциальных измерительных устройств периодического сравнения, основанные на выборе начальной фазы коммутирующего напряжения // Отбор и передача информации, вып.59. К., 1980. С.78-83.
3. Бучма І.М. Похибки виділення обвідної методом запам'ятовування амплітудних значень у схемах періодичного порівняння // Вимірювальна техніка та метрологія. 1999. №55. С.25-31.
4. Бучма І. Автоматична корекція похибок методом зразкових сигналів у схемах з періодичним порівнянням // Вимірювальна техніка та метрологія. 2000. №56. С.3-8.
5. Земельман М.А. Автоматическая коррекция погрешностей измерительных устройств. М., 1972.
6. Зыбов В.Н., Мизюк Л.Я., Тетерко А.Я. Принципы построения одноканальных устройств обработки сигналов при двухчастотном вихретоковом контроле // Отбор и передача информации, вып.71. К.,1985. С. 84-89.
7. Приборы для неразрушающего контроля материалов и изделий. Справочник./Под ред. Клюева В.В. кн.2. М., 1986.
8.Проектування засобів вимірювання з періодичним порівнянням. Кн.1 і 2; Навч. посібник /Ю.О.Скрипник, М.О.Присенко, В.О.Дубровний. К., 1977.
9.Семенюк Г.Т., Шаповалов Г.О., Паньчишин О.Я. та інші. Колісний транспортний робот для автоматизації процесів діагностики крупногабаритних зварних споруд // Матеріали доповідей наук.-техн. конф."Фізичні методи та засоби контролю матеріалів та виробів" Київ-Львів, 1996. С.81-82.
10. Yamada T., Suzuki I. Спосіб та електромагнітний пристрій для вимірювання товщини пластини. A=Trans. Jest. Electron. Ins. and Commun. Eng. Jap. A.-1988, v.71, p. 1458-1460.

УДК 621.391.19

ВИКОРИСТАННЯ ДОВЖИН ЗРАЗКІВ ПРИ РОЗПІЗНАВАННІ МОВИ

© Р. Попович
Національний університет "Львівська політехніка"

Запропоновано використання довжин мовних зразків (кількості акустичних векторів спектральних оцінок) для звуження області пошуку класифікатора при розпізнаванні мови.

One has offered using of speech patterns lengths (number of spectral estimation acoustic vectors) for reducing of classifier search domain at speech recognition.

Вступ

Різноманітні мовні технології (стиск та передача мовних сигналів, зміна темпу мовлення, синтез мови, розпізнавання мови, ідентифікація особи за голосом, діагностика ступеня певних захворювань) інтенсивно розвиваються та знаходять усе більше за-

стосування в різних областях діяльності людини. Розпізнавання мови є одним з найскладніших напрямків мовних технологій із можливістю різнопланового застосування.

Постановка задачі

Типова система розпізнавання мови має в собі чотири компоненти [1,2,3,4]: звуковий препроцесор, звукову модель, модель мови та класифікатор.

Загалом, фраза мовлення, яку треба розпізнати, складається з послідовності слів $W = w_1, w_1, ..., w_n$. Робота системи розпізнавання полягає у визначенні найбільш відповідної (ймовірної) послідовності слів $\hat{W}$, маючи звуковий (мовний) сигнал.

На початковому етапі обробки з мовного сигналу вибирається вся необхідна звукова інформація в компактному вигляді. Звуковий препроцесор перетворює мовний зразок на послідовність акустичних векторів $\mathbf{Y} = \mathbf{y}_1, \mathbf{y}_1, ..., \mathbf{y}_n$. Кожен вектор є стислим поданням короткочасного мовного спектру на інтервалі, як правило, близько 25 мс зі зсувом інтервалів на 10 мс.

Звукова модель забезпечує метод утворення зразкових (еталонних) послідовностей акустичних векторів за заданою послідовністю слів.

Метою моделі мови є дати метод оцінки можливості появи (апріорної ймовірності) послідовності слів незалежно від спостереження мовного сигналу. Для цього забезпечується механізм оцінки можливості появи певного слова у фразі, якщо знаємо попередні слова.

Класифікатор зводить воєдино дані від трьох раніше описаних компонент. Він бере послідовності слів, дозволені моделлю мови, утворює за цими послідовностями, користуючись звуковою моделлю, зразкові послідовності акустичних векторів та порівнює їх з послідовністю акустичних векторів, утворених звуковим препроцесором. Найближча до утвореної звуковим препроцесором зразкова послідовність акустичних векторів вказує на найбільш відповідну послідовність слів, яка і є результатом роботи класифікатора.

Як правило, усі можливі послідовності слів відслідковуються паралельно. Цей підхід спирається на принцип оптимальності Белмана (динамічного програмування) і його часто називають алгоритмом Вітербі [2,5,6].

Через складність сучасних систем розпізнавання мови суттєвим завданням є звуження області пошуку найбільш відповідної послідовності слів.

Відомо два способи вирішення цього завдання: використання різних моделей мови та обмеження ширини потоку (поняття активного стану). У випадку, коли немає ніяких обмежень на послідовність слів у фразі, модель мови незастосовна (наприклад, розпізнавання вимовлених номерів телефонів чи номерів автомобілів). Використання поняття активного стану описане далі.

Обчислення за алгоритмом Вітербі ведуться послідовно (рекурсивно) в часі. Для обмеження ширини потоку введено поняття активного стану. Оцінка $V(t)$ відповідності послідовності слів у момент часу t обчислюється, коли вона досяжна з активного стану в момент часу t-1. Активні стани - це такі стани, для яких оцінка $V(t\text{-}1)$ близька до max $V(t\text{-}1)$. Якщо вдало вибрати поріг, який задає близькість оцінок $V(t\text{-}1)$ та max $V(t\text{-}1)$,то цей евристичний прийом зменшує обсяг обчислень при погіршенні точності розпізнавання.

Завдання даної роботи – дати ще один додатковий прийом для звуження області

пошуку класифікатора.

Отримання спектральних векторів

Принципове припущення, яке робиться в сучасних розпізнавачах [1,2,3] є те, що мовний сигнал розглядається як стаціонарний (тобто спектральні характеристики відносно постійні) на інтервалі в кілька десятків мілісекунд. Тому основною функцією попередньої обробки є розбити вхідну мову на інтервали [2,3] і для кожного інтервалу отримати згладжену спектральну оцінку.

У даній роботі бралися мовні зразки, дискретизовані з частотою 16 КГц та розрядністю 16 біт. Дискретизована мова розбивалася на інтервали тривалістю 25,6 мс, тобто 409 відліків. Інтервали перекривалися зі зсувом на 10 мс (160 відліків).

Як звичайно, застосовувалось високочастотне підсилення, щоб компенсувати послаблення, спричинене розсіюванням від губ. Для цього інтервали пропускалися через фільтр першого порядку

$$T(1) = 0;\ T(n) = S(n) - S(n - 1),\ \ n = 2\ldots409.$$

Для обробки такого типу до кожного інтервалу застосовується функція вікна, у даному випадку бралося вікно Хемінга.

Блоки доповнювалися справа нулями до довжини 512 точок. Потім у результаті використання швидкого перетворення Фур'є отримувалися 256 комплексних значень у спектральній області.

Фур'є спектр згладжувався додаванням 255 останніх спектральних коефіцієнтів (перший коефіцієнт - постійна складова спектру ігнорувався) у межах "трикутних" частотних смуг, розташованих на нелінійній (подібній до логарифмічної) *Mel*-шкалі [2]. Для граничної частоти мови, що дорівнює 16 КГц, беруть 24 таких частотних смуги. Mel-шкала введена для наближення частотного розділення людського вуха, яке є лінійним до 1000 Гц та логарифмічним понад 1000 Гц.

У результаті описаних дій отримувався 24-елементний акустичний вектор.

Виконувалася нормалізація акустичних векторів у межах одного зразка. Для цього знаходилася максимальна норма вектора та всі вектори множилися на величину, обернену до цього значення.

Використання довжин зразків

Вимагається при розпізнаванні мови дотримуватися таких умов:

- попередньо виділено мовний сигнал (слово або фразу) на фоні пауз і шумів. Для цього можна скористатися, зокрема, підходом з робіт [7,8].
- як зразки беруться якнайкоротше вимовлені слова (зразки).

При виконанні цих двох умов мовний сигнал, який треба розпізнати, порівнюється не зі всіма зразками (зразками слів у випадку розпізнавання ізольованих слів та послідовностями зразків слів у випадку розпізнавання суцільної мови), а тільки з тими, для яких справджується:

1) $time(n) \le time(1)$,

де $time(n)$, $time(1)$- тривалість (кількість акустичних векторів) відповідно n-го (склад-

ного) зразка й мовного сигналу, який треба розпізнати. Така умова повинна виконуватися, бо як зразки взяті якнайкоротше вимовлені слова.

2) $time(1) \le a\ time(n)$,

де a - коефіцієнт можливого сповільнення темпу мови порівняно із зразками. Для розпізнавання мови в нормальному темпі коефіцієнт a може дорівнювати 2.

В експериментах з розпізнавання ізольованих слів було взято 10 слів (назви цифр від 0 до 9). Розпізнавання виконувалося на основі принципу динамічного часового вирівнювання *DTW* (*dynamic time warping*). При цьому допускалося лише однократно повторювати кожен другий акустичний вектор зразка. Отримані довжини мовних сигналів слів, що на 100% розпізнавалися після навчання на конкретного диктора, зведені в таблицю. У другому стовпці таблиці наведено діапазони довжин сигналів для відповідних слів. Перше число у другому стовпці таблиці - це мінімальна довжина, тобто довжина зразка.

Слово	Довжина
ноль	22-40
один	38-58
два	22-39
три	22-38
чотири	44-65
п'ять	33-48
шість	46-65
сім	38-62
вісім	39-63
дев'ять	46-65

Наприклад, сигнал довжиною 50 векторів, що відповідав слову "один", порівнювався лише зі зразками 4,5,6,7,8,9. Сигнал довжиною 25 векторів, що відповідав слову "три", порівнювався лише зі зразками 0,2,3.

Такий же підхід може бути і при розпізнаванні з використанням КДП-підходу [1]. Для кожного зразка рахується довжина як сума максимально можливих чисел повторюваності зразкових векторів (не враховуючи еталонні вектори пауз).

При розпізнаванні суцільної мови контроль за співвідношенням між тривалістю сигналу й утворюваного складного зразка слід проводити рекурсивно в часі при побудові дерева можливих послідовностей слів.

Висновки

Використання довжин мовних зразків (кількості векторів спектральних оцінок) дозволяє звузити область пошуку класифікатора при розпізнаванні мови. Разом з тим порівняння довжин зразків можна використовувати поряд з двома іншими підходами для досягнення цієї мети: моделлю мови та обмеженням ширини потоку на основі поняття активного стану.

1. Вінцюк Т.К. Аналіз, розпізнавання й інтерпретація мовних сигналів, К., 1987.
2. Young S. Large vocabulary continuous speech recognition. IEEE Signal Processing Magazine, 13(5), 1996, pp.45-57.
3. Kapadia S. Discriminative training of hidden Markov models. PhD thesis, Cambridge University, 1998.
4. Hain T., Woodland P.C., Evermann G., Povey D. The CU-HTK March 2000 Hub 5E transcription system. Proc. Speech Transcription Workshop. College Park, 2000.
5. Young S. The HTK hidden Markov model toolkit: design and philosophy. Technical Report CUED/F-INFENG/TR152,Cambridge University Engineering Department, 1994.
6. Young S., Russell N., Thornton J. Token passing: a simple conceptual model for connected speech recognition systems. Technical Report CUED/F-INFENG/TR38, Cambridge University Engineering Department, 1989.
7. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. Львів, 1997.
8. Rabiner L.R., Schafer R.W. Digital processing of speech signals. Prentice-Hall, Englewood Cliffs, NJ, 1978.