

УДК 004.7:621.32

Ю.В. Данюк

Харківський університет Повітряних Сил імені Івана Кожедуба, Харків

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ РОЗПІЗНАВАННЯ МОВНИХ СИГНАЛІВ НА ОСНОВІ СПОСОБУ СПЕКТРАЛЬНОГО ОЦІНЮВАННЯ

Розглянуто задачу розпізнавання мовних сигналів на основі способу спектрального оцінювання за критерієм мінімуму інформаційної неузгодженості з урахуванням фонемної структури слів. Наведено практичні результати моделювання на реальних мовних сигналах. Проведено порівняння з кепстральним способом.

Ключові слова: розпізнавання, мовний сигнал, спектральне оцінювання.

Постановка проблеми

Розпізнаванню мовних сигналів як науковій задачі вже досить багато років, і до цих пір вона залишається актуальною. Один з найперших способів мовного розпізнавання заснований на спектральному аналізі, представлений у вигляді набору алгоритмів [1]. Особливий інтерес викликає універсальний підхід, що зводиться до критерію мінімуму інформаційної неузгодженості (МІН) і реалізується у частотній області з використанням спектральних оцінок [2].

Метою статті є експериментальне дослідження ефективності спектрального підходу до критерію МІН з урахуванням фонемної структури слова для розпізнавання мовних сигналів.

1. Спектральний підхід

Вихідним пунктом при обґрунтуванні методу спектрального підходу є авторегресійна (АР) модель мовного сигналу. Така апроксимація добре описує поведінку мовного тракту при його розгляді у вигляді динамічної труби [3]. Рівняння, що описує АР процес, визначається таким чином [4]:

$$x(n) = \sum_{i=1}^P a(i)x(n-i) + o(n), \quad (1)$$

де $x(n)$ – масив даних розміром n ; $a(i)$ – АР коефіцієнти процесу; P – порядок процесу; $o(n)$ – процес типу гаусівського шуму.

Обчислення параметрів АР моделі мовного сигналу, а саме її вектор АР коефіцієнтів $a(i)$, $i = 1 \dots P$, проводиться з розрахунку на мінімізацію помилки лінійного передбачення за її дисперсією. Високими динамічними характеристиками оцінок відрізняється метод Берга [4], при цьому вибір порядку моделі представляє собою оптимізаційну задачу, виходячи з необхідних якісних характеристик алгоритму і доступного обсягу даних.

Вирішення задачі розрізнення для сімейства сигналів, отриманих шляхом лінійної фільтрації «бі-

лого» гаусівського шуму, визначається у припущенні про АР природу сигналів, що підлягають аналізу, із застосуванням принципу інформаційної неузгодженості. Використовуючи граничний перехід від формулювання критерію мінімуму інформаційної неузгодженості в метриці Кульбака-Лейблера в часовій області через матриці коваріації, визначимо оптимальний алгоритм для розрізнення сигналів в частотній області через їх оцінки спектру [2]:

$$\Gamma_{x,r} = \frac{1}{F} \sum_{f=1}^F \left(\frac{G_x(f)}{G_r(f)} + \ln \frac{G_r(f)}{G_x(f)} \right) \rightarrow \min_{r=1, R}, \quad (2)$$

де $G_x(f)$ – вибіркова оцінка спектральної щільності потужності сигналу x ; $G_r(f)$ – вибіркова оцінка спектральної щільності потужності сигналу r із словника; F – половина частоти дискретизації; R – розмір словника.

Алгоритм реалізується у вигляді набору з R паралельних каналів обчислення вирішальної статистики (2). База даних словника містить R , за кількістю розпізнаваних слів, оцінок спектральної щільності потужності сигналів.

У кожному каналі проводиться обчислення статистики на основі оцінки спектру аналізованого сигналу і оцінки спектру одного сигналу зі словника. Рішення приймається у вирішальному пристрої за критерієм мінімуму вирішальної статистики $\gamma_v = \min$ на виходах обчислювачів статистики в кожному з R каналів.

При аналізі природного мовного сигналу часто використовується підхід, при якому обмежений у часі нестационарний сигнал розглядають як часткову реалізацію нескінченно довгого стаціонарного сигналу. Однак безпосереднє застосування до мовних сигналів результату (2) не дозволяє досягти максимальної ефективності алгоритму у зв'язку з нестационарністю мовних сигналів [5]. На практиці можна говорити лише про кусочно-часову стаціонарність мовних сигналів при проголошенні слів, що обумовлено самою фізичною природою мовотворення і неможливістю миттєвої фізіологічної артикуляції

голосового апарату людини. Теорія і практика показують, що період стаціонарності мовного тракту має тривалість близько 20 мсек. [6].

З урахуванням сказаного вирішальна статистика (2) вимагає видозміни, а саме обліку часової структури стаціонарності слів, тобто переходу до фонемного аналізу. Поділ слів на фонемні частини як частина фонетичного аналізу є окремим складним завданням практичної лінгвістики. Ефективним способом реалізації фонемного підходу є метод сегментного аналізу, який полягає в роздільному аналізі ділянок слова з певною довжиною і розрахунок характеристик з урахуванням результатів, які отримано на кожній з цих ділянок. Найбільш простим алгоритмом є поділ слова на суміжні сегменти фіксованої довжини, які не перекриваються. При цьому апріорна база даних словника буде містити не один, а кілька векторів спектральних коефіцієнтів кожного слова, обчислених за відповідними сегментами. Обчислення вирішальної статистики кожного слова буде проводитись шляхом усереднення вирішальних статистик, обчислених за сегментами.

Остаточно, з урахуванням фонемної структури, модифікована вирішальна статистика перетвориться до виду:

$$\Gamma_{x,r} = \frac{1}{L} \sum_{i=1}^L \left(\frac{1}{F} \sum_{f=1}^F \left(\frac{G_x^{(i)}(f)}{G_r^{(i)}(f)} + \ln \frac{G_r^{(i)}(f)}{G_x^{(i)}(f)} \right) \right) \rightarrow \min_{r=1, R}, \quad (3)$$

де $G_x^{(i)}(f)$ – вибіркова оцінка спектральної щільності потужності сегменту i сигналу x ; $G_r^{(i)}(f)$ – вибіркова оцінка спектральної щільності потужності сегменту i сигналу r із словника; F – половина частоти дискретизації; R – розмір словника; L – кількість сегментів обробки.

Після такої модифікації отримуємо можливість застосовувати метод як цілком до слова, встановлюючи кількість сегментів обробки L , що дорівнює одиниці, так і до фонемної структури слова, усереднюючи інформаційну неузгодженість сегментів сигналів по всій його довжині $L \gg 1$. Згідно вирішальної статистики згенеруємо алгоритм обробки сигналів.

2. Синтез алгоритму

Відповідно до критерію (3) алгоритм прийняття рішення за вибіркою даних x довжиною N відліків може бути записаний у вигляді наступної послідовності операцій.

Сегментування проводиться шляхом ділення сигналу на суміжні сегменти однакової довжини. Оцінка спектральної щільності потужності кожного сегмента проводиться процедурою швидкого перетворення Фур'є (ШПФ) з попереднім зважуванням відліків сигналу у вікні обробки. Для спектрального аналізу мовних сигналів традиційно застосовується

зважування вікном Хеммінга [7]:

$$w(k) = 0.54 - 0.46 \cos \left(\frac{2\pi k}{n-1} \right), k = \overline{1, n}, \quad (4)$$

де k – номер відліку вікна; n – довжина вікна.

Алгоритм ковзаючого вікна ШПФ для отримання оцінки спектра потужності сегмента сигналу записується як [4]:

$$G(f) = \frac{1}{w} \sum_{k=1}^w f \{ x_w(k); k = \overline{1, l+1} \}, \quad (5)$$

де $x_w(k)$ – відліки даних сегмента сигналу довжиною, зважених вікном Хеммінга (4); w – кількість вікон, які кратні довжині сегмента, $w = n/l$; l – довжина вікна ШПФ у відліках.

Вирішальна статистика для кожного сегмента обчислюється на основі оцінок спектральної щільності потужності сегмента аналізованого слова та зразка зі словника:

$$\Gamma_{x,r}^{(i)} = \frac{1}{F} \sum_{f=1}^F \left(\frac{G_x^{(i)}(f)}{G_r^{(i)}(f)} + \ln \frac{G_r^{(i)}(f)}{G_x^{(i)}(f)} \right). \quad (6)$$

Сумарна вирішальна статистика для слова цілком обчислюється як середнє значення статистик, обчислених для кожного з сегментів окремо:

$$\Gamma_{x,r} = \frac{1}{L} \sum_{i=1}^L \Gamma_{x,r}^{(i)}, \quad (7)$$

де $\gamma_{x,r}$ – інформаційна неузгодженість між сегментом i сигналу x та сегментом i сигналу r зі словника; L – кількість сегментів.

Вирішальне правило на користь одного з сигналів словника визначається за критерієм мінімуму підсумкової інформаційної неузгодженості між аналізованим сигналом і сигналом зі словника.

$$\Gamma_{x,r} \rightarrow \min_{r=1, R}. \quad (8)$$

Таким чином, для обчислення остаточної статистики (8) проводиться обчислення інформаційної неузгодженості (7) між сегментами порівнюваних слів. Оцінка щільності потужності спектру для обчислення статистик проводиться методом ковзаючого вікна із застосуванням ШПФ (5). Результат розпізнавання визначається мінімумом (8) середнього значення статистик, обчислених за сегментами розпізнаваного слова.

3. Результати експериментальних досліджень

Словник складався з десяти слів – числівників від нуля до дев'яти включно [6]. Завдання такого типу в області розпізнавання мовлення, коли словник має мале число слів, носить назву завдання голосового управління автоматичними системами, і є актуальним для практичного застосування. Частота дискретизації 8кГц і квантування 8 біт були обрані з міркування мінімізації потоку звукових даних для

відповідності прийнятим в більшості стандартів параметрів для передачі та зберігання мови. Для максимального наближення експерименту до практичної ситуації були використані стандартний канал вбудованого звукового процесора побутового персонального комп'ютера і звичайний мікрофон. Для створення експериментальної бази даних були записані одним диктором з однією інтонацією по 100 слів кожного з 10 числівників.

Інтересом практичних досліджень було вимір ймовірності правильного розпізнавання слів незалежно від параметрів моделі. Параметрами моделі були довжина сегмента аналізу і довжина вікна ШПФ. Результати дослідження ілюструються графіком (рис. 1).

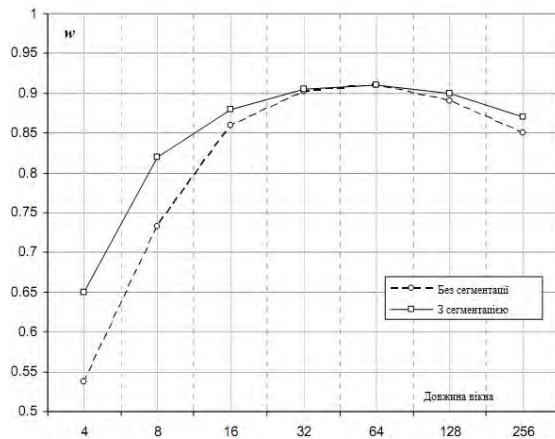


Рис. 1. Залежність ймовірності правильного розпізнавання слів в залежності від вікна ШПФ для обробки слова цілком та для пофонемної обробки

Ймовірність правильного розпізнавання обчислюється в результаті серії випробувань за такою формулою:

$$w = \frac{K_{\text{прав.}}}{K_{\text{заг.}}}, \quad (9)$$

де $K_{\text{прав.}}$ – кількість правильно розпізнаних слів в серії випробувань; $K_{\text{заг.}}$ – загальна кількість випробувань.

Висновки

Таким чином, запропонована реалізація спектрального підходу до задачі розпізнавання мовних сигналів на основі критерію мінімуму інформаційної неузгодженості та проведено її експериментальне дослідження. В результаті був реалізований алгоритм розпізнавання і отримані дані, які кількісно характеризують спосіб. Визначено, що ймовірність правильного розпізнавання в області коротких довжин вікон ШПФ зростає при врахуванні обліку фонемної структури слів методом сегментації. Порівняльний аналіз спектральних і кепстральних способів розпізнавання мови показав велику ймовірність правильного розпізнавання спектральними способами. Додатковим підтвердженням ефективності запропонованого алгоритму може служити хороший результат розпізнавання мови іншого диктора.

Список літератури

1. Oscar Mayora Ibarra, Francesco Curatelli. *A Brief Introduction to Speech Analysis and Recognition* Università degli Studi di Genova Dipartimento di Ingegneria Biofisica ed Elettronica, 2000.
2. Савченко В.В. Различение случайных сигналов в частотной области / В.В. Савченко // *Радиотехника и электроника*. – 1997. – Т.42, №4. – С. 426-429.
3. Joseph P. Campbell *Speaker Recognition: A Tutorial* / P. Joseph // *IEEE Invited Paper*, Cambridge press. – 2000.
4. Марпл С.Л. *Цифровой спектральный анализ и его приложения*. – М.: Мир, 1990.
5. Потапова Р.К. *Речь: коммуникация, информатика, кибернетика* / Р.К. Потапова. – М.: Радио и связь, 1997.
6. Greenberg S. *Understanding Speech Understanding: Towards a Unified Theory of Speech Perception*. - Department of Linguistics International Computer Science Institute University of California, Berkeley, CA 94720 USA.

Надійшла до редакції 25.10.2013

Рецензент: д-р техн. наук, проф. І.В. Рубан, Харківський університет Повітряних Сил, ім. І. Кожедуба, Харків.

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ РАСПОЗНАВАНИЯ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ СПОСОБА СПЕКТРАЛЬНОГО ОЦЕНИВАНИЯ

Ю.В. Данюк

Рассмотрена задача распознавания речевых сигналов на основе способа спектрального оценивания по критерию минимума информационной несогласованности с учетом фонемной структуры слов. Приведены практические результаты моделирования на реальных речевых сигналах. Проведено сравнение с кепстральным способом.

Ключевые слова: распознавание, речевой сигнал, спектральное оценивание.

RESEARCH THE EFFECTIVENESS OF RECOGNITION OF SPEECH SIGNALS BASED ON THE METHOD OF ESTIMATING THE SPECTRUM

Y.V. Danyuk

The problem of speech recognition based on the method of spectral estimation by the minimum of information inconsistent with the phonemic structure of words. The practical results of the simulation to the real speech signals. A comparison with the cepstral method.

Keywords: pattern recognition, speech signal spectral estimation.