

побудований на основі недорогих серійних пристроїв з невисоким рівнем енергоспоживання як портативний автономний переносний прилад для розв'язання розглянутого в роботі класу задач в умовах виробництва.

1. Битюцкий О.И. и др. Оптико-электронная система бесконтактного контроля геометрических параметров полых цилиндров // *Автометрия*. – 1995.– №6. – С.69–74. 2. Non-contact dimensional gage for turned parts // *International Class: G01J 003/45 / Патент США № 4880991 від 14.11.1989 р.* 3. System for measuring shapes and dimensions of gaps and flushnesses on three dimensional surfaces of objects // *International Class: G06K 009/00 / Патент США № 5129010 від 07.07.1992 р.* 4. Комерційна інформація фірми Beta LaserMike. www.betalasermike.com. 5. Кринич-ний П.Я., Райтер П.М., Грицив А.Б., Вісков А.В. Контроль геометричних параметрів пластмасових труб у процесі їх виготовлення: Зб. наук. праць. Серія: Фізичні методи та засоби контролю середовищ, матеріалів та виробів. – Вип. 10 “Акустичні та електромагнітні методи неруйнівного контролю матеріалів та виробів”. – Львів., 2005. – С. 64–71. 6. Ильин В.А., Позняк Э.Г. Аналитическая геометрия. – М., 1999.

УДК 681.3.06

М.М. Биков, К.Ю. Смирнов, В.Ю. Марущак
Вінницький національний технічний університет

ДИКТОРОНЕЗАЛЕЖНИЙ МЕТОД КОДУВАННЯ ЗВУКІВ В СИСТЕМАХ РОЗПІЗНАВАННЯ МОВНИХ СИГНАЛІВ

© Биков М.М., Смирнов К.Ю., Марущак В.Ю., 2006

Аналізується модель мовоутворення, що ґрунтується на принципі “квазічастотної” модуляції голосового тракту, і метод дикторонезалежного кодування звуків мови двійковими частотодетектуючою і частотносегментуючою функціями, що ґрунтується на цій моделі. Запропоновано принцип ієрархічного структурування фонетичної інформації в акустичному сигналі, поданого двійковими значеннями вказаних функцій, реалізацією паралельного процесу сегментації і маркування мовного сигналу. Розглянуто процедури навчання класифікатора міток і сегментатора, наведено результати їх роботи.

In the given activity the model of speech production, founded on the principle of the vocal tract “quazifrequency” modulation is described. The method of the speaker independence encoding of the speech sounds with the binary frequency detecting and frequency segmenting functions based on this model is esteemed. The principle of hierarchic structuring of the phonetic information in an a speech signal, described by binary values of the refered functions is offered, by implementation of parallel process of segmentation and labelling. The procedures of the label classifier training and the speech segmentator are reviewed, the outcomes of their activity are adduced.

Вступ. Аналіз методів опису мовного сигналу, що ґрунтуються на відомих моделях, показує їх неінваріантність до диктора, оскільки вони використовують за ознаки енергетичні характеристики у вузьких діапазонах спектра. Крім того, подання звуків мови, що ґрунтуються на таких описах, є досить об'ємними, що вимагає великих обсягів пам'яті для їх зберігання, значної пропускну здатності каналу зв'язку. Тому виникає необхідність в розробці таких моделей мовоутворення [1], які дали можливість реалізувати економне і дикторонезалежне кодування звуків мовного сигналу.

Постановка проблеми і її зв'язок з важливими науковими завданнями. Однією з невіршених проблем в області інформаційно-вимірювальних систем є побудова дикторонезалежних систем автоматичного розпізнавання мовних сигналів. Її вирішення дало б змогу розширити коло користувачів таких систем і значно підвищити ефективність обміну інформацією в людино-машинних

системах. Загалом задача побудови ефективної системи розпізнавання може бути сформульована як задача пошуку оптимального за загальносистемним критерієм дерева рішень, в якому на кожному кроці класифікації з апіорного алфавіту вибирається підмножина ознак, що максимально зменшує на досягнутому кроці ентропію про образ і збільшує швидкість класифікації [2]. Така стратегія передбачає використання множинного опису слів у термінах різних фонетичних класів, що відповідають різним рівням дерева класифікації, а також вибору інформативних дикторонезалежних ознак для виділення фонетичних класів на кожному рівні. У цій роботі пропонується модель мовоутворення, що ґрунтується на принципі “квазічастотної” модуляції голосового тракту і метод дикторонезалежного кодування мовних звуків мови двійковими частотодетектуючою і частотносегментуючою функціями, що ґрунтується на цій моделі, а також принцип ієрархічного структурування фонетичної інформації в акустичному сигналі, поданого двійковими значеннями вказаних функцій, реалізацією паралельного процесу сегментації і маркування мовного сигналу.

Основна частина. Аналіз залежності інформативних властивостей звуків мови від їх частотно-енергетичних параметрів показує, що основна інформація сигналу мови закодована в перших трьох формантних діапазонах, тому у цій роботі запропоновано модель мовоутворення на основі “квазічастотної” модуляції голосового тракту. У цій моделі голосовий тракт вважається джерелом інформаційного (мовного) сигналу, кодування інформації в якому здійснюється модуляцією трьох несучих частот – частоти 1-ї форманти, частоти 2-ї форманти та частоти 3-ї форманти. Положення частоти в формантних діапазонах визначається положенням частотних моментів сигналу:

$$M_{kf} = \frac{\int_{F_{k-1}}^{F_k} A_f \cdot f df}{\int_{F_{k-1}}^{F_k} f df}, \quad (1)$$

де A_f – спектральна густина мовного сигналу для смуги частот df ; f – поточне значення частоти сигналу; k – номер частотного каналу, $k = 1, 2, 3$.

Попередній аналіз сигналу мови в формантних діапазонах проводився за допомогою смугових фільтрів, при цьому вираз (1) для частотного моменту набирає вигляду:

$$M_{kf} = \frac{\sum_{i=l}^{l+m} A_i \cdot f_i}{\sum_{i=l}^{l+m} f_i}, \quad (2)$$

де A_i – амплітуда сигналу на виході i -го фільтра; f_i – центральна частота смугового фільтра; l – номер першого смугового фільтра в k -ому частотному каналі; m – кількість фільтрів в k -ому каналі.

Закодувавши декілька положень частотного моменту в кожній смузі частот, можна перейти від опису мовного сигналу в неперервному тривимірному просторі до дискретного опису в просторі двійкових значень частотодетектуючої функції. У кожному з вибраних частотних каналів можливо розглядати три форми спектра (положення частотних моментів) (рис. 1).

f

f

f

а)

б)

в)

Рис. 1. Кодовані положення частотних моментів в частотних каналах:

а – девіація в бік високих частот; б – девіація в бік низьких частот; в – нульова девіація

Для двійкового кодування цих положень частотних моментів кожний частотний канал розбивається на три піддіапазони, при цьому в одному каналі отримується два розряди частотодетектуючої функції θ_{di} в такий спосіб:

$$\theta_{di} = \bigcup_{i=1}^2 \sigma(M_k^i + M_k^{i+1}) \quad (3)$$

де σ – одинична функція, $\sigma(M_k^i \alpha M_k^{i+1}) = 1$, якщо $M_k^i > M_k^{i+1}$, і дорівнює 0 в протилежному випадку; α – відношення домінування.

За такого визначення частотодетектуючої функції для першого розрізнюваного випадку девіації частоти $(\theta_{d1}, \theta_{d2}) = (0, 0)$, для другого – $(\theta_{d1}, \theta_{d2}) = (1, 1)$, для третього – $(\theta_{d1}, \theta_{d2}) = (0, 1)$.

Отже, отриманий початковий опис мовного сигналу за допомогою частотодетектуючої функції має вигляд восьмибітового двійкового слова. Наприклад, для звука [а] двійковий опис має вигляд

$$[a] = (0, 1, 1, 0, 0, 0, 1, 1).$$

Значення цієї функції обчислюється для кожного τ -го первинного сегмента сигналу мови, тривалість якого вибрана такою, що дорівнює $t_s = 20$ мсек. Для кожної пари суміжних в часі значень частотодетектуючої функції обчислюються значення сегментуючої функції за формулою $\theta_s^\tau = \theta_d^\tau \oplus \theta_d^{\tau-1}$, де символ \oplus означає логічну операцію “сума по модулю два”. Значення сегментуючої функції використовуються для процесу сегментації сигналу мови на окремі звукотипи. У такий спосіб в цьому підході процес сегментації на звуки проходить паралельно з процесом їх класифікації.

Значення середніх частот для кожного з трьох каналів $f_{ик}$ визначається значеннями формантних частот в нейтральному положенні голосового тракту.

Частотні діапазони каналів 1-ї, 2-ї і 3-ї формант можуть бути визначені за статистичними даними про значення формантних смуг ΔF_1 , ΔF_2 і ΔF_3 [4 – 6]:

$$\Delta F_1 = 250 - 1000 \text{ Гц};$$

$$\Delta F_2 = 800 - 2200 \text{ Гц};$$

$$\Delta F_3 = 1780 - 3560 \text{ Гц}.$$

Під час спектрального аналізу сигналу мови смуговими фільтрами з центральними частотами, розміщеними за логарифмічним законом впродовж частотної осі, в смугах ΔF_1 , ΔF_2 і ΔF_3 опиняться 12 частотних діапазонів, розподілених по формантних каналах.

Сегментуюча функція призначена для сегментації мовного сигналу на окремі звукотипи. Її значення являють собою комбінації з восьми біт, що змінюються в часі. Логічно припустити, що в разі зміни значень розрядів в словах сегментуючої функції відбувається перехід енергії сигналу з одних частотних діапазонів в інші, що свідчить про перехід від одного звукотипу до іншого у цьому часовому проміжку. Безумовно, переходи між різними комбінаціями фонем є різними. Це ускладнює задачу побудови автоматичного класифікатора, який повинен розпізнавати нестационарні (перехідні) проміжки мовного сигналу, тобто ставити на цих місцях відповідні мітки переходів.

Побудову класифікатора виконано за допомогою математичного апарату штучних нейронних мереж. Структура автоматичного класифікатора для сегментації мовного сигналу має вигляд,

показаний на рис. 2.

Для проведення навчання було сформовано вибірку із сорока слів української мови. Для кожного слова була проведена обробка, яка складалась з таких етапів:

- 1 визначення меж слова;
- 2 фільтрація сигналу в заданих частотних діапазонах формантних смуг;
- 3 розрахунок частотно-детектуючої функції;
- 4 розрахунок сегментуючої функції.

Встановлення ділянок нестационарності мовного сигналу здійснювалось за допомогою спектрограми мовного сигналу. На рис. 3 показано спектрограму зліва “менше”.

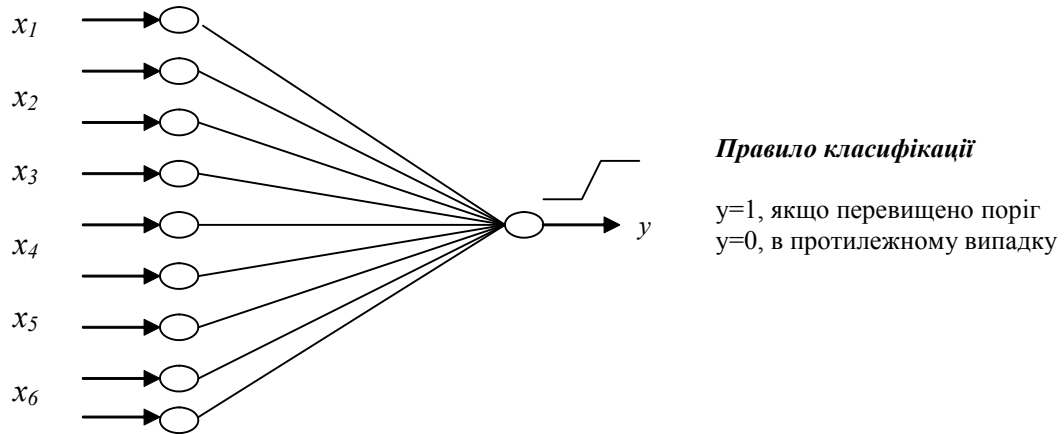


Рис. 2. Структура класифікатора

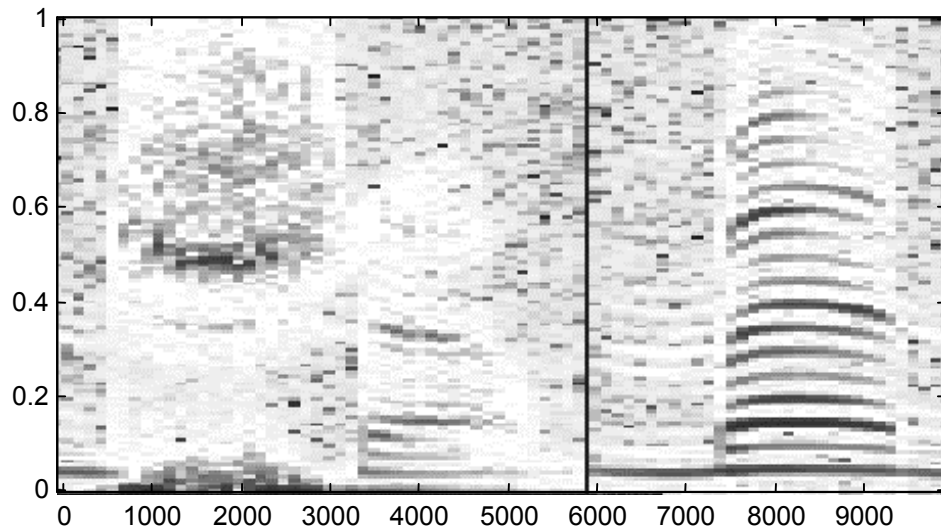


Рис. 3. Спектрограма слова “менше”

Зміна енергії сигналу на різних частотах, як бачимо з рис. 3, показує проміжки нестационарності сигналу (t_1, \dots, t_4) .

Загалом навчання проводилось на вибірці з 1500 комбінацій сегментуючої функції. Слова словника вимовлялись трьома дикторами по шість разів. Результат навчання показав похибку класифікації 0,3 % для навчальної вибірки.

Побудова автоматичного нейромережевого класифікатора з двома шарами нейронів (вхідним і вихідним) відповідає побудові розподільної поверхні між двома класами образів, що відомо з класичної теорії розпізнавання. Знаючи значення ваг зв'язків між нейронами вхідного і вихідного шару, можна скласти відповідне рівняння розподільної поверхні:

$$w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5 + w_6x_6 + w_7x_7 + w_8x_8 + 1 = 0 \quad (4)$$

Шляхом навчання були встановлені числові значення вагових коефіцієнтів класифікатора:

$$0.37x_1 + 0.56x_2 + 0.72x_3 + 1.38x_4 + 1.179x_5 - 1.06x_6 - 0.88x_7 - 0.14x_8 - 0.5 = 0 \quad (5)$$

На рис. 4 показано результати роботи класифікатора після навчання на слові “менше”.

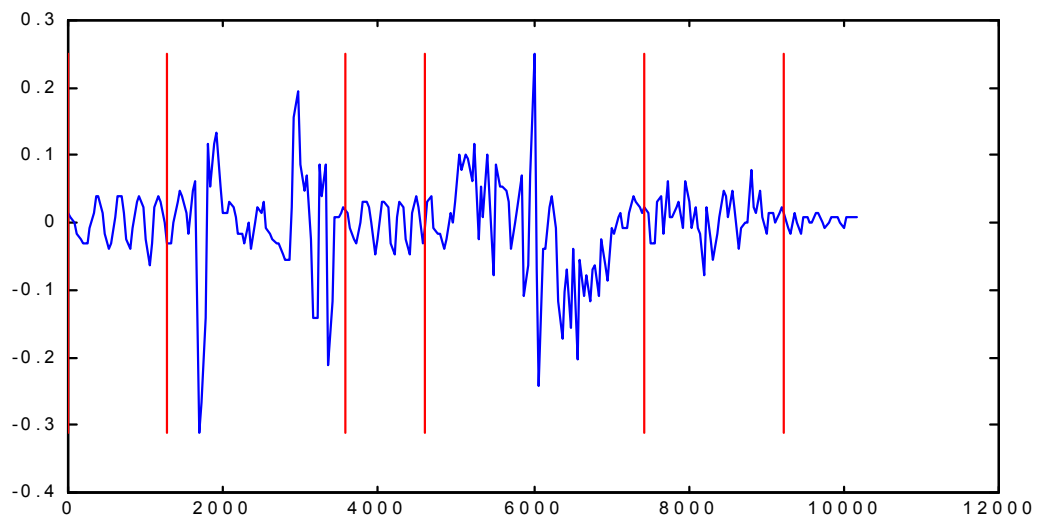


Рис. 4. Сегментація слова “менше”

Структурування фонетичної інформації у вигляді ієрархічного дерева звукотипів здійснювалось в двійковому просторі ознак шляхом кластерного аналізу навчальної вибірки векторів, отриманих шляхом опису відомих звуків у вказаний спосіб. За координати двійкового простору вибрано компоненти значень частотодетектуючої функції. Для отриманих кластерів було проведено навчання класифікатора з метою отримання вагових коефіцієнтів вирішувальних функцій, при цьому за ортогональний базис вибирались функції Уолша. Результати класифікації вирішувальними функціями тестової вибірки звукотипів на кожному рівні ієрархічного дерева показали надійну роздільність фонетичних класів аж до рівня, що передуює опису звуків мови в термінах фонем. Це дає можливість зробити висновок, що для фонемної транскрипції мовного сигналу потрібно залучити додатковий набір ознак.

Висновки. Запропонований метод опису мовного сигналу двійковими восьмирозрядними векторами частотодетектуючої і частотосегментуючої функцій, одержаними шляхом кодування положення частотних моментів у визначених діапазонах частот дав змогу підвищити інваріантність мовних образів до диктора і голосності вимовлення, понизити на порядок порівняно з відомими

методами надлишковість подання мовної інформації, здійснити процеси сегментації і маркування сигналу на звукотипи паралельно до часу і тим самим збільшити точність і швидкість розпізнавання. Запропонований метод ієрархічного структурування фонетичної інформації у вигляді дерева еталонів слів за допомогою навчених на вибіркового матеріалі класифікаторів дав можливість використати ієрархічний принцип розпізнавання мови з використанням додаткових модулів за збільшення складності задачі розпізнавання.

1. Биков М.М., Кузьмін І.В., Коберський О.Г., Пастушенко О.В. Звіт про науково-дослідну роботу "Розробка моделей, методів і алгоритмів для опису, кодування та розпізнавання сигналів мови" (шифр 46-Д-170), № держреєстрації 0197UO12877. – Вінниця, 1997. 2. Быков М.М. Методы и средства измерения и преобразования информации в системах машинного распознавания речи. – Дисс. ... канд. техн. наук. – Винница, 1985. – С.67–73. 3. Bykov N.M., Kuzmin I.V., Yakovenko A.I. Development of effective strategy of pattern recognition // Proceedings of SPIE. – 2000. – Vol.4425. – P.75–82. 4. Фант Г. Акустическая теория речеобразования. – М., 1964. 5. Сапожков М.А. Речевой сигнал в кибернетике и связи. – М., 1963. 6. Фланаган Дж. Анализ, синтез и восприятие речи. – М., 1968.

УДК 681.213

В.О. Поджаренко, В.Ю. Кучерук, В.М. Севастьянов
Вінницький національний технічний університет

МЕТРОЛОГІЧНА АТЕСТАЦІЯ АЛГОРИТМУ ВИМІРЮВАННЯ КУТОВОЇ ШВИДКОСТІ

© Поджаренко В.О., Кучерук В.Ю., Севастьянов В.М., 2006

Проведено аналіз метрологічних та технічних характеристик, необхідних для метрологічної атестації алгоритму вимірювання кутової швидкості.

The analysis metrological and characteristics necessary for metrological certification of algorithm of measurement of angular speed is carried out.

Вступ. Існує велика кількість алгоритмів функціонування цифрових тахометрів, проте їх структурна схема фактично однакова. Вона складається із послідовного з'єднання вала двигуна з валом тахометричного перетворювача, аналого-цифрового перетворювача та числового вимірювального перетворювача. Наведена послідовність становить вимірювальний канал кутової швидкості, для оцінки якості функціонування якого необхідно провести атестацію алгоритмічного забезпечення. Проте більшість вимірювань кутової швидкості проводяться у динамічному режимі, тобто за неусталеної кутової швидкості, тому пропонується атестувати алгоритм вимірювання кутової швидкості у динамічному режимі.

Аналіз стану досліджень та формування мети. Атестація алгоритмічного забезпечення – це визначення основних характеристик алгоритмів, які дають можливість зіставляти їх і вибирати для конкретних задач, а також оцінювати похибки одержаних результатів, що ґрунтується, з одного боку, на статистичних методах (насамперед "аналізі даних"), і з іншого, – на загальній метрологічній ідеології атестації різних об'єктів [1].

Процедура атестації алгоритмів має загальніший характер, ніж їх метрологічна атестація в рамках конкретних методик виконання вимірювань. Відмінність полягає у тому, що сукупність моделей початкових даних у першому випадку повинна бути максимально широкою, а в другому –