її застосування. – Львів, 2001. – 278 с. 3. Gilbert Strang, Truong Nguyen. Wavelets and Filter Banks. Wellesley-Cambridge Press, 1996. 4. Бабак В.П., Хандецький В.С., Шрюфер Е. Обробка сигналів. – К., 1996. 5. Bose N.K., Digital Filters Theory and Applications, Elsevier Science Publishing, New York, 1985. 6. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. М.: Мир, 1978 – 835 с. 7. Гольденберг Л.М., Матюшкин Б.Д., Поляк М.Н. Цифровая обработка сигналов. – М. 1990. 8. Макс Ж. Методы и техника обработки сигналов при физических измерениях. – М., 1983. Т. 1. – 312 с. 9. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. – Л., 1991. – 304 с.

УДК 681.3.019:621.39

ДИКТОРОНЕЗАЛЕЖНЕ ОПИСАННЯ ОБРАЗІВ В СИСТЕМАХ РОЗПІЗНАВАННЯ СИГНАЛІВ МОВИ

© Биков Микола[1], Раїмі Абдурахман[2], Биков Максим[3], 2006

[1]Вінницький національний технічний університет, м. Вінниця, Україна
[2]Université Cheikh Anta Diop, Dakar, Senegal
[3]Фірма "Альпарі", м. Вінниця

Запропонована модель сигналу мови на принципі "квазічастотної" модуляції голосового тракту, метод дикторонезалежного описання мовних образів, який ґрунтується на цій моделі, розроблено нейромережевий класифікатор для сегментації сигналу мови

Предложена модель речевого сигнала на принципе "квазичастотной" модуляции голосового тракта, метод дикторонезависимого описания речевых сигналов, основанный на этой модели, разработан нейросетевой классификатор для сегментации речевого сигнала.

The model of speech signal based on the principle of "kvazifrequency" modulation of vocal tract is offered, also the method of speaker independent description of speech signals, based on this model, the neuro network classifier for speech signal segmentation is developed.

Вступ. Однією з актуальних невирішених проблем у галузі інформаційно-вимірювальних систем є побудова систем автоматичного розпізнавання сигналів мови, інваріантних до диктора. Її вирішення дало б змогу б розширити коло користувачів таких систем і значно підвищити ефективність обміну інформацією в людино-машинних системах. Загалом задача побудови ефективної дикторонезалежної стратегії розпізнавання мови може бути сформульована як задача пошуку оптимального за загальносистемним критерієм дерева рішень, в якому на кожному кроці класифікації з апріорного алфавіту вибирають підмножину ознак, що максимально зменшує на досягнутому кроці ентропію про образ і збільшує швидкість класифікації [1]. Така стратегія передбачає використання множинного описання 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 роботі запропоновано модель мовоутворення [2], яка ґрунтується на принципі "квазічастотної" модуляції голосового тракту і метод дикторонезалежного опису мовних образів двійковими частотно-детектувальною і частотно-сегментувальною функціями, що ґрунтується на цій моделі, а також принцип ієрархічного структурування фонетичної інформації в акустичному сигналі, представленого двійковими значеннями вказаних функцій, за допомогою реалізації паралельного процесу сегментації і маркування мовного сигналу.

Аналіз стану досліджень та публікацій. Аналіз методів описання мовного сигналу, основаних на відомих моделях, показує їхню неінваріантність до диктора, оскільки їх використовують як ознаки енергетичних характеристик у вузьких діапазонах спектра [3], [4]. У статті розглядається модель мовоутворення, яка описує сигнал положенням частотних моментів енергії сигналу в широких формантних діапазонах, що дає змогу знизити варіацію ознак за рахунок спектральних варіацій. Двійкове кодування положення цих моментов на основі відношення енергій в частотних піддіапазонах допомагає уникнути впливу амплітудних варіацій.

Мета досліджень. Метою досліджень є підвищення швидкості і точності розпізнавання сигналів мови за рахунок використання дикторонезалежного описання мовних образів.

Виклад основного матеріалу. Аналіз залежності інформативних властивостей звуків мови від їхніх частотно-енергетичних параметрів показує, що основна інформація сигналу мови закодована у перших трьох формантних діапазонах, тому запропонована модель мовоутворення на основі "квазічастотної" модуляції голосового тракту. У цій моделі голосовий тракт вважають 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 у формантних діапазанах визначається положенням частотних моментів сигналу:

$$M_{kf} = \frac{\int_{F_{K-1}}^{F_K} A_f \cdot fdf}{\int_{F_{K-1}}^{F_K} fdf}, \qquad (1)$$

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 у формантних діапазонах здійснюють за допомогою смугових фільтрів, а вираз (1) для частотного моменту набуває вигляду:

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

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

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

Для двійкового кодування цих положень частотних моментів кожний частотний канал ділять на три піддіапазони; в одному каналі отримують два розряди частотно-детектувальної функції $\theta_{\partial i}$ у такий спосіб

$$\theta_{\partial i} = \bigcup_{i=1}^{2} \sigma\ (M_k^i\ \alpha\ M_k^{i+1}), \qquad (3)$$

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

За такого визначення частотно-детектувальної функції для першого розрізнюваного випадку девіації частоти $(\theta_{\partial 1}, \theta_{\partial 2}) = (0, 0)$, для другого – $(\theta_{\partial 1}, \theta_{\partial 2}) = (1, 1)$, для третього – $(\theta_{\partial 1}, \theta_{\partial 2}) = (0, 1)$.

Аналіз спектрів показує, що використання значень частотно-детектувальної функції у трьох каналах ефективне для описання голосних звуків мови і звучних приголосних (м, н, р, л). Для розрізняння шумних і проривних приголосних звуків істотними є глобальні характеристики спектра (спектр-форма) сигналу. Для їхнього визначення використовується співвідношення енергій сигналу в трьох сусідніх смугах ΔF_1, ΔF_2 і ΔF_3.

Отже, отримане початкове описання мовного сигналу за допомогою частотно-детектувальної функції має вигляд восьмибітового двійкового слова. Наприклад, для звука [а] двійкове описання має вигляд (а) = (0,1,1,0,0,0,1,1). Значення цієї функції обчислюють для кожного τ-го первинного сегмента сигналу мови, тривалість якого становить t_s = 20 мс. Для кожної пари суміжних в часі значень частотно-детектувальної функції обчислюються значення сегментувальної функції за формулою $\theta_s{}^{\tau} = \theta_d{}^{\tau} \oplus \theta_d{}^{\tau-1}$, де символ $\oplus$ означає логічну операцію "сума за модулем два". Значення сегментувальної функції використовуються для сегментації сигналу мови на окремі звукотипи. Отже, за цим підходом сегментація на звуки проходить паралельно з процесом їхньої класифікації.

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

$$f_{нк} = (2k-1)\frac{c}{4l'}$$

де c – швидкість звуку в [см/с], c= 35300; l' – середня довжина мовного тракту, l' = 17,5 см [4 – 5].

Тоді $f_{н1}$ =504 Гц, $f_{н2}$ =1512 Гц, $f_{н3}$ = 2524 Гц. У разі зміни положення артикуляційних органів форма голосового тракту і його довжина змінюються, що відповідає зміні параметрів модуляції, положення формант і форма спектра вихідного сигналу змінюються, і він відповідає тому чи іншому звукові мови.

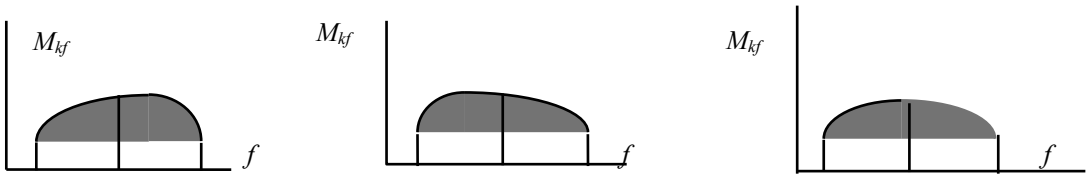

Рис. 1. Кодовані положення частотних моментів у частотних каналах:
а – девіація у бік високих частот;
б – девіація у бік низьких частот; в – нульова девіація

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] :

$$\Delta F_1 = 250 – 1000 \text{ Гц},$$
$$\Delta F_2 = 800 – 2200 \text{ Гц},$$
$$\Delta F_3 = 1780 – 3560 \text{ Гц}.$$

При спектральному аналізі 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, так розподілених по формантних каналах: ΔF_1 – (252-317), (317-400), 400-504), (504-635),(635-800),(800-1008); ΔF_2 -(1008-1270), (1270-1600), (1600-2016), (2016-2540); ΔF_3 -(2016-2540), (2540-3200), (3200-4032).

Навчання класифікатора на сегментацію та маркування мовного сигналу на фонотипи. Сегментувальна функція призначена для сегментації мовного сигналу на окремі звукотипи. Її значення являють собою комбінації з восьми бітів, що змінюються в часі. Логічно припустити, що у разі зміни значень розрядів у словах сегментувальної функціїї відбувається перехід енергії сигналу з одних частотних діапазонів в інші, що свідчить про перехід від одного звукотипу до іншого у певному часовому проміжку. Безумовно, переходи між різними комбінаціями фонем є різними. Це ускладнює завдання побудови автоматичного класифікатора, який повинен розпізнавати нестаціонарні (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 зображена на рис. 2.

Для навчання була сформована вибірка із сорока слів української мови. Для кожного слова була виконана обробка, яка складалась з таких етапів:

– визначення меж слова;

– фільтрація сигналу у заданих частотних діапазонах формантних смуг;

– розрахунок частотно-детектувальної функції;

– розрахунок сегментувальної функції.

Ділення нестаціонарності мовного сигналу визначали за допомогою спектрограми мовного сигналу. На рис. 3 і рис. 4 наведено спектрограми командних слів "*менше*" і "*назад*".

Зміна енергії сигналу на різних частотах, як видно з рис. 3, показує проміжки нестаціонарності сигналу ($t_1, \ldots, t_4$).

Загалом навчання здійснювалось на вибірці з 1500 комбінації сегментувальної функції. Слова словника вимовлялись трьома дикторами по шість разів. Результат навчання показав похибку класифікації 0,3% для навчаючої вибірки.

Побудова автоматичного нейромережевого класифікатора з двома шарами нейронів (вхідним і вихідним) відповідає побудові розподільної поверхні між двома класами образів, що відомо в класичній теорії розпізнавання.

Знаючи значення ваг зв'язків між нейронами вхідного і вихідного шарів, можна скласти відповідне рівняння розподільної поверхні:

$$W_1x_1 + W_2x_2 + W_3x_3 + W_4x_4 + W_5x_5 + W_6x_6 + W_7x_7 + W_8x_8 + 1 = 0,$$

За допомогою навчання були встановлені числові значення вагових коефіцієнтів класифікатора:

$$0.37x_1 + 0.56x_2 + 0.72x_3 - 1.38x_4 + 1.179W_5x_5 - 1.06x_6 - 0.88x_7 - 0.14x_8 - 0.5 = 0.$$

На рис. 4 подано результати роботи класифікатора після навчання на слові "*менше*".

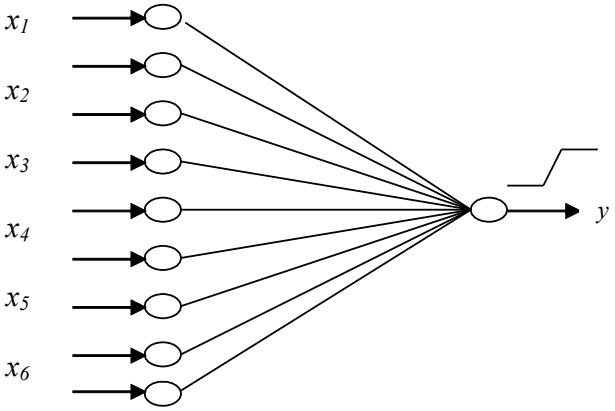

Рис. 2. Структура класифікатора

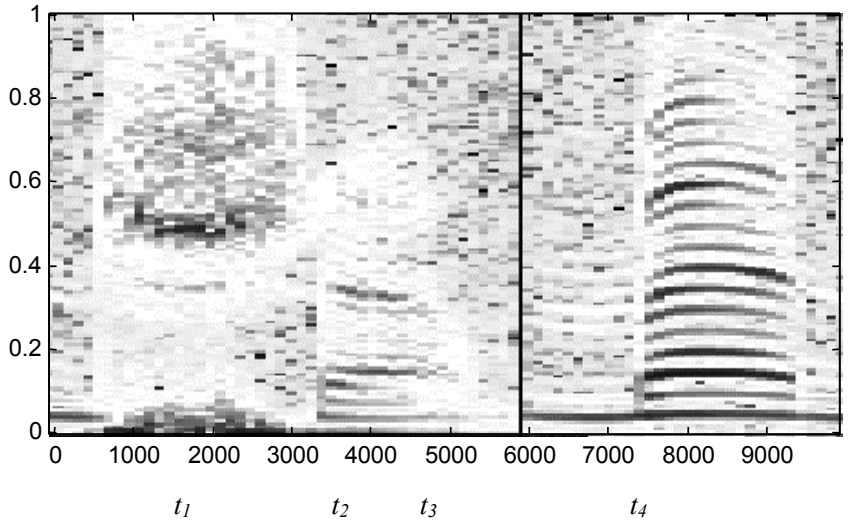

Рис. 3. Спектрограма слова "*менше*"

Рис. 3. Спектрограма слова "менше"

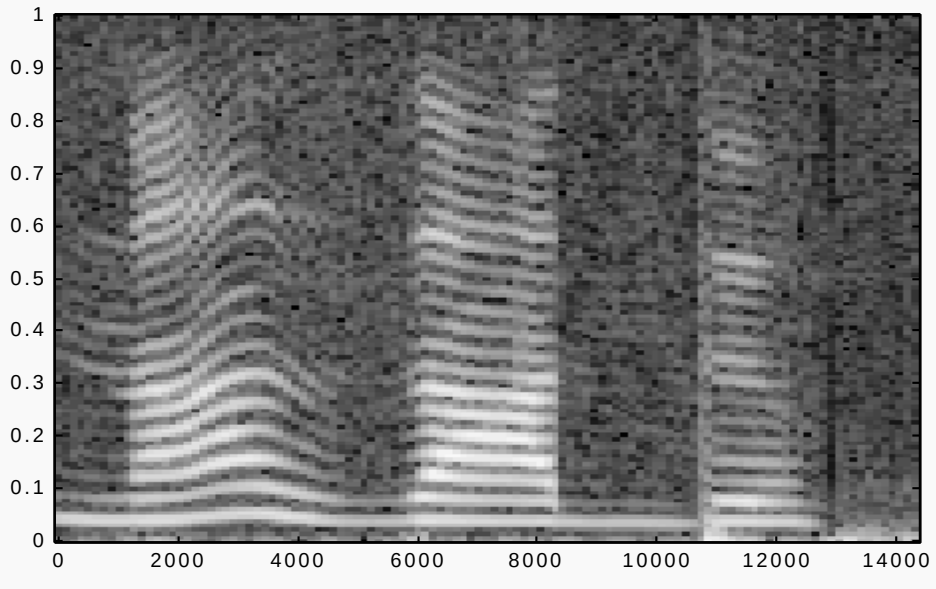

Рис. 4. Спектрограма слова "назад"

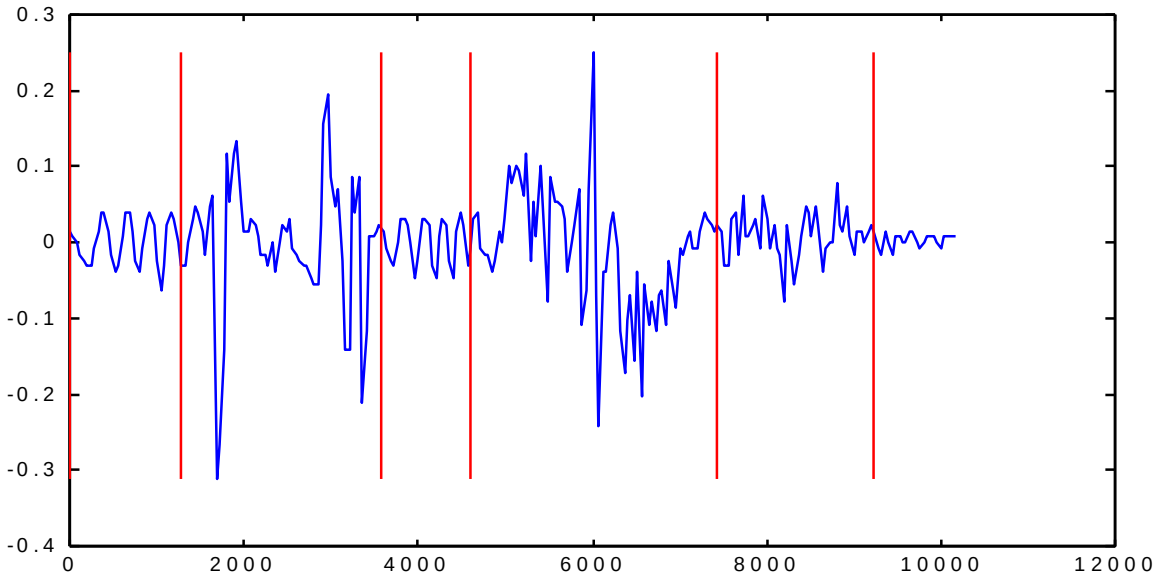

Рис. 5. Сегментація слова "менше"

Висновки. Запропонований метод описання мовного сигналу двійковими восьмирозрядними векторами частотно-детектувальної і частотно-сегментувальної функцій, одержаними кодуванням 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 відображення мовної інформації, здійснити процеси сегментації і маркування сигналу на звукотипи паралельно в часі і тим самим збільшити точність і швидкість розпізнавання.

1. 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. 2. Быков М.М. Методы и средства измерения и преобразования информации в системах машинного распознавания речи. – Дисс. на соискание уч. степени канд. техн. наук. – Винница, 1985. – С.67 – 73. 3. 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, ВДТУ, 1997. 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.