

ДИКТОРОНЕЗАЛЕЖНИЙ ОПИС ОБРАЗІВ В СИСТЕМАХ РОЗПІЗНАВАННЯ СИГНАЛІВ МОВИ

Вінницький національний технічний університет;

Анотація

Запропонована модель сигналу мови на принципі “квазічастотної” модуляції голосового тракту, метод дикторонезалежного опису мовних образів, оснований на даній моделі, розроблено нейронмережевий класифікатор для сегментації сигналу мови

Ключові слова: дикторонезалежний опис образів

Abstract

The model of speech signal based on the principle of “quazifrequency” 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.

Key words: non-narcognized description of images

Вступ. Однією з актуальних нерозв’язаних проблем в області інформаційно-вимірювальних систем є побудова систем автоматичного розпізнавання сигналів мови, інваріантних до диктора. Її розв’язання дозволило б розширити коло користувачів таких систем і значно підвищити ефективність обміну інформацією в людино-машинних системах. В загальному випадку задача побудови ефективної дикторонезалежної стратегії розпізнавання мови може бути сформульована як задача пошуку оптимального по загальносистемному критерію 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 [1]. 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. В даній роботі пропонується модель мовоутворення [2], основана на принципі “квазічастотної” модуляції голосового тракту і метод дикторонезалежного опису мовних образів двійковими частотндетектуючою і частотносегментуючою функціями, що базується на даній моделі, а також принцип ієрархічного структурування фонетичної інформації в акустичному сигналі, представленого двійковими значеннями вказаних функцій, шляхом реалізації паралельного процесу сегментації і маркування мовного сигналу.

Результат дослідження

Розглядається модель мовоутворення, яка описує сигнал положенням частотних моментів енергії сигналу в широких формантних діапазонах, що дозволяє знизити варіацію ознак за рахунок спектральних варіацій. Двійкове кодування положення цих моментів на основі відношення енергій в частотних піддіапазонах дозволяє уникнути впливу амплітудних варіацій.

Аналіз залежності інформативних властивостей звуків мови від їх частотно-енергетичних параметрів показує, що основна інформація сигналу мови закодована в перших трьох формантних діапазонах, тому в даній роботі запропонована модель мово утворення на основі “квазічастотної” модуляції голосового тракту. В цій моделі голосовий тракт вважається джерелом інформаційного (мовного) сигналу, кодування інформації в якому здійснюється шляхом модуляції трьох несучих частот - частоти 1-ої форманти, частоти 2-ої форманти та частоти 3-ої 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 полоси частот df ;

f - поточне значення частоти сигналу;

k - номер частотного каналу, $k = 1, 2, 3$.

Попередній аналіз сигналу мови в формантних діапазонах проводиться з допомогою смугових фільтрів, при цьому вираз (1) для частотного моменту приймає вигляд:

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

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 - число фільтрів в k -ому каналі.

Закодувавши декілька положень частотного моменту в кожній смузі частот, можна перейти від опису мовного сигналу в неперервному трьохвимірному просторі до дискретного опису в просторі двійкових значень частотодетектуючої функції. В кожному з вибраних частотних каналів можливо розглядати наступні три форми спектру (положення частотних моментів):

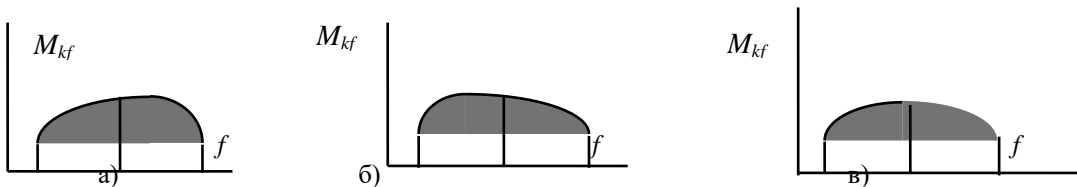


Рис. 1. Кодовані положення частотних моментів в частотних каналах: а- девіація в бік високих частот; б- девіація в бік низьких частот; в- нульова девіація

Для двійкового кодування цих положень частотних моментів кожний частотний канал розбивається на 3 піддіапазони, при цьому в одному каналі отримується 2 розряди частотодетектуючої функції θ_{di} наступним чином

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

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.

При такому визначенні частотодетектуючої функції для першого розрізнюваного випадку девіації частоти $(\theta_{d1}, \theta_{d2}) = (0, 0)$, для другого - $(\theta_{d1}, \theta_{d2}) = (1, 1)$, для третього - $(\theta_{d1}, \theta_{d2}) = (0, 1)$.

Аналіз спектрів показує, що використання значень частотодетектуючої функції в трьох каналах ефективно для опису голосних звуків мови і звучних приголосних (м,н,р,л). Для розрізнення шумних і проривних приголосних звуків суттєвими є глобальні характеристики спектру (спектр-форма) сигналу. Для їх визначення використовується співвідношення енергій сигналу в трьох сусідніх смугах ΔF_1 , ΔF_2 і ΔF_3 .

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

(а) = (0,1,1,0,0,0,1,1). Значення цієї функції обчислюється для кожного τ -го первинного сегмента сигналу мови, тривалість якого вибрана рівною $t_s = 20$ мсек. Для кожної пари суміжних в часі значень частотодетектуючої функції обчислюються значення сегментуючої функції по формулі $\theta_s^\tau = \theta_d^\tau \oplus \theta_d^{\tau-1}$

¹, де символ \oplus означає логічну операцію “сума по модулю два”. Значення сегментуючої функції використовуються для процесу сегментації сигналу мови на окремі звукотипи. Таким чином, в даному підході процес сегментації на звуки проходить паралельно з процесом їх 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 / сек], $c = 35300$;

l - середня довжина мовного тракту, $l' = 17,5$ см [4 - 5].

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

Частотні діапазони каналів 1-ої, 2-ої і 3-ої формант можуть бути визначені по статистичним даним 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 по логарифмічному закону впродовж частотної осі, в смугах ΔF_1 , ΔF_2 і ΔF_3 опиняться 12 частотних діапазонів, розподілених по формантних каналах наступним чином: 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).

Висновки. Запропонований метод дозволив підвищити інваріантність мовних образів до диктора і голосності вимовлення, понизити на порядок порівняно з відомими методами надлишковість представлення 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, pp. 76-82.

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. – м. Вінниця, ВДТУ, 1997. – 23 с.

Тарасівський Олександр Русланович — студент групи КІВ-166, факультет комп'ютерних систем і автоматики, Вінницький національний технічний університет, Вінниця, e-mail: tumeteorm@gmail.com

Науковий керівник: **Биков Микола Максимович** — кандидат технічних наук, доцент, професор кафедри КСУ, Вінницький національний технічний університет, м. Вінниця

Tarasivskiy Sasha R. — Faculty of Computer Systems and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: tumeteorm@gmail.com

Supervisor: Bykov N.M — Candidate of Technical Sciences, docent, Professor of the Department of Computer Control Systems, Vinnytsia National Technical University, Vinnytsia