УДК 681.3.019:621.39

М. М. Биков, к. т. н., доц.; В. В. Ковтун, к. т. н.; Н. Г. Савінова, студ.

НАДІЙНИЙ МЕТОД ВИДІЛЕННЯ СКЛАДОВИХ СЕГМЕНТІВ У МОВНОМУ СИГНАЛІ

Запропоновано новий метод, що підвищує надійність виділення складових сегментів у мовному сигналі, і розроблено алгоритм і пристрій, що реалізують запропонований метод.

Ключові слова: *розпізнавання мови, мовний сигнал, сегментація сигналу, ознаки образів, частотний діапазон, пристрій виділення складів, алгоритм виділення ознак складів.*

Вступ

Використання інформації про склади в ієрархічних системах розпізнавання мови дозволяє підвищити дикторонезалежність і контекстнонезалежність цих систем [1], і як наслідок, підвищити точність розпізнавання. У нашій запропоновано новий метод, який підвищує надійність виділення складових сегментів у мовному сигналі, і розроблено алгоритм і пристрій, що реалізують запропонований метод.

Постановка задачі

Незважаючи на те що на сьогодні існують комерційні системи автоматичного розпізнавання мови, актуальною залишається проблема розробки таких методів і засобів розпізнавання, що не потребують налагодження під індивідуальні голосові особливості диктора. Одним із підходів до побудови дикторонезалежних систем розпізнавання є використання на верхніх рівнях ієрархічної системи в якості елементів розпізнавання звукотипів, фонетичні характеристики яких мало залежать від диктора і від контексту, наприклад, складів, півскладів, дзвінких, фрикативних, пауз і таке інше. В роботі [2] досліджено, що використання, наприклад, тільки складової інформації дозволяє уже на верхньому рівні розпізнавання скоротити кількість кандидатів на класифікацію в 2 – 4 рази. Такою інформацією є тривалість складів і їх кількість у висловленні. Одним з основних параметрів, які використовуються для розмежування складів у мовному сигналі, є його енергія [3, 4]. При цьому ядро складу визначається в місці локалізації максимумів енергії, обмежених істотними (на 40 або 50 дБ) спадами енергії. Однак у ряді зазначених робіт, наприклад [5], відзначається часте виділення за цією ознакою помилкових складів, сформованих високоенергетичними фрикативними або сонорними звуками. Це підтверджують і експериментально зняті записи. У роботі [6] як параметр для виділення ознаки складу використвуєтьсся функція "гучності", одержувана як зважена сума амплітуд сигналів 22-х частотних каналів, розміщених у критичних смугах. Очевидний недолік такого методу формування ознаки складу – великі апаратурні або обчислювальні витрати та недостатня надійність.

Для виключення недоліків, властивих зазначеним методам, у нашій роботі ставиться за **мету** розробка нового методу, який би дозволив підвищити надійність виділення сегментів мовного сигналу, що відповідають складам мови, а також алгоритму і пристрою, що реалізують цей метод. У наступному розділі розглянута модель, яка дозволяє визначити інформативні ознаки складових сегментів, на основі чого розробляється метод, алгоритм і пристрій для виділення цих сегментів.

Математична модель і метод виділення ознак складів з мовного сигналу

Розкриття механізмів відображення спектральних параметрів мовного сигналу в просторі інваріантних ознак можливе шляхом доповнення процесів периферійної слухової системи

процесами в центральній слуховій системі структурною схемою моделі слуху так, як це показано на рис. 1.

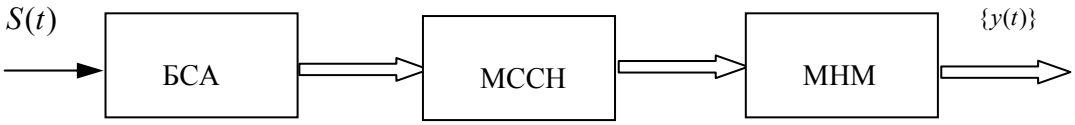

Рис. 1 Узагальнена модель слухової системи

У наведеній на рис. 1 узагальненій слуховій моделі блок спектрального аналізу (БСА) відображає частотно-вибіркові властивості моделі і представляє собою набір фільтрів, на вхід яких подається мовний сигнал $S(t)$. Зазвичай фільтри перекривають частотний діапазон від 250 до 6400 Гц, в якому зосереджена основна енергія мовного сигналу.

Модель сенсорних слухових нейронів (МССН) відображає дію слухових нейронів, які з'єднані з волосковими клітками базилярної мембрани вуха. Вона враховує такі слухові ефекти, як динамічне стискування вхідного сигналу, однонапівперіодне його випрямлення та регулювання підсилення сигналу. Механізми слухового сприйняття, які представлені даними моделями, досить добре вивчені, однак багаточисельні спроби їх застосування в розпізнаючих пристроях не дали бажаних результатів. Тому в роботі розглядається уточнена модель слухової системи з урахуванням роботи нейронної мережі, представленої її моделлю (МНМ).

Особливості формування мовних образів моделлю нейронної мережі з сигналами $\{y(t)\}$ на виходах нейронів проаналізовані в роботах [7, 8]. Проведений аналіз показав, що ознаки мовного сигналу необхідно шукати серед елементів автокореляційної матриці спектральних параметрів мовного сигналу:

$$\left\| y_x \right\| = x \cdot x^T = \begin{Vmatrix} x_1 \cdot x_1 & x_1 \cdot x_2 & \dots & x_1 \cdot x_n \\ x_2 \cdot x_1 & x_2 \cdot x_2 & \dots & x_2 \cdot x_n \\ \dots & \dots & x_i \cdot x_j & \dots \\ x_n \cdot x_1 & x_n \cdot x_2 & \dots & x_n \cdot x_n \end{Vmatrix} \qquad (1)$$

З огляду на отримані результати моделювання розроблено новий метод виділення ознак складових сегментів, для формування яких за первинні параметри використовуються огинаючі сигналів у частотних діапазонах Δ_1 = 800 – 2500 Гц і Δ_2= 250 – 540 Гц. Результуючий параметр, який в подальшому використовується для виділення ознак складів, отримується кореляційним методом і записується у вигляді:

$$U_c(t) = U_{\Delta 1}(t) \cdot U_{\Delta 2}(t), \qquad (2)$$

де $U_{\Delta 1}(t)$ – огинаюча енергії в смузі частот Δ_1, а $U_{\Delta 2}(t)$ – огинаюча енергії в смузі Δ_2.

Діапазон частот першого смугового фільтру 3, рівний 250 – 540 Гц, вибраний з огляду на те, що в ньому відсутня енергія високоенергетичних фрикативних звуків типу /ш/ і /ч/, які створюють помилкові складові ядра, а так само зосереджена значна частина енергії всіх дзвінких звуків, у тому числі і голосних. Проте в цьому діапазоні енергія сонорних звуків типу /л/, /м/, /н/ співвідносна з енергією голосних, тому визначення складових сегментів тільки за огинаючою мовного сигналу в цьому діапазоні супроводжуватиметься багатьма помилками. Тому діапазон частот другого смугового фільтру 4, вибраний в межах 800 – 2500 Гц, в якому енергія голосних звуків мінімум в два рази перевищує енергію сонорних звуків.

При виконанні операції множення огинаючих $U_{\Delta 1}(t)$ і $U_{\Delta 2}(t)$ в результуючій часовій функції відбувається посилення ділянок кривої у області голосних звуків через кореляцію їх енергій в обох діапазонах, а помилкові максимуми енергії, зумовлені наявністю в діапазоні 800 – 2500 Гц значної частини енергії фрикативних звуків, усуваються їх множенням на

практично нульове значення амплітуди фрикативних звуків в діапазоні 250 – 540 Гц.

Алгоритм і пристрій виділення ознак складів

Згідно з описом функціонування пристрою виділення складових сегментів алгоритм його роботи складається з таких кроків:

1. Введення сигналу мови.
2. Фільтрація сигналу двома смуговими фільтрами Баттерворта четвертого порядку в діапазонах 250 – 540 Гц і 800 – 2500Гц відповідно.
3. Детектування вихідних сигналів фільтрів для отримання огинаючих.
4. Перемноження огинаючих вихідних сигналів фільтрів.
5. Диференціювання результуючого сигналу.
6. Порівняння отриманого сигналу з додатньою і від'ємною пороговими напругами і виділення логічного сигналу.
7. Формування з отриманого сигналу логічних сигналів для додатніх і від'ємних півперіодів диференційованого сигналу.
8. Виділення сегментів складів і центрів складів шляхом логічного додавання і множення отриманих логічних сигналів відповідно.

Схема алгоритма зображена на рис. 2.

В алгоритмі використані такі позначення:

$ff_1 = filter\ (p_1,p_2,\ s)$ і $ff_2 = filter\ (p_3,p_4,\ s)$ – функції фільтрації в смугах частот Δ_1 = 800 – 2500 Гц і Δ_2 = 250 – 540 Гц відповідно; $U_{g1} = abs\ (ff_1)$, $U_{g2} = abs\ (ff_2)$ – огинаючих сигналів у вказаних діапазонах. Інші позначення розкриваються в описанні роботи пристрою виділення ознак.

Пристрій для виділення складових сегментів у мовному сигналі працює так (рис. 3).

Мовний сигнал сприймається акустичним датчиком, перетворюється в електричний і надходить на вхід підсилювача. Електричний сигнал, посилений до величини, достатньої для роботи подальших каскадів, надходить на входи двох смугових фільтрів зі смугами Δ_1 = $[p_1;p_2]$ = 800 – 2500 Гц і Δ_2 = $[p_3;p_4]$ = 250 – 540 Гц. Амплітудний детектор, вхід якого підключений до виходу першого смугового фільтру, виділяє огинаючу U_{g2} мовного сигналу в діапазоні 250 – 540 Гц. Амплітудний детектор, вхід якого підключений до виходу другого смугового фільтру, виділяє огинаючу U_{g1} мовного сигналу в діапазоні 800 – 2500 Гц. Напруги U_{g1} іU_{g2}, що надходять на входи помножувача сигналів, перемножуються, внаслідок чого на його виході з'являється напруга U_n, рівна:

$$U_n = U_{g1}.U_{g2}.$$

Діапазони частот першого і другого смугових фільтрів підібрані таким чином, що в результаті виконання операції множення корелює тільки енергія голосних звуків, що приводить до усунення в огинаючій U_n максимумів, відповідних ділянкам високоенергетичних фрикативних звуків. Ця напруга надходить на вхід диференціатора 9, формувача ознак 8 (рис. 3), на виході якого утворюється напруга U_{ng}, пропорційна похідній від напруги U_n. Напруга надходить на перші входи порогових схем 10 і 11. На другий вхід порогової схеми 10 надходить додатна порогова напруга U_{n1}, а на другий вхід порогової схеми 11 надходить негативна порогова напруга U_{n2}, причому U_{n1} і U_{n2} вибрані так, що $|U_{n1}| = |U_{n2}| \approx 50$ мВ, при цьому усуваються помилкові спрацьовування порогових схем через фонові шуми під час переходу через U_{ng} нульові значення. З виходів порогових схем 10 і 11 одержують кліповані сигнали, приведені до стандартних рівнів цифрових сигналів. Ці сигнали надходять на входи схеми АБО 12, на виході якої утворюється цифровий сигнал $U_л$. Фронти сигналу $U_л$, що надходить на вхід лічильного тригера з прямим динамічним

управлінням 13, перемикають щоразу його в протилежний стан, внаслідок чого на його

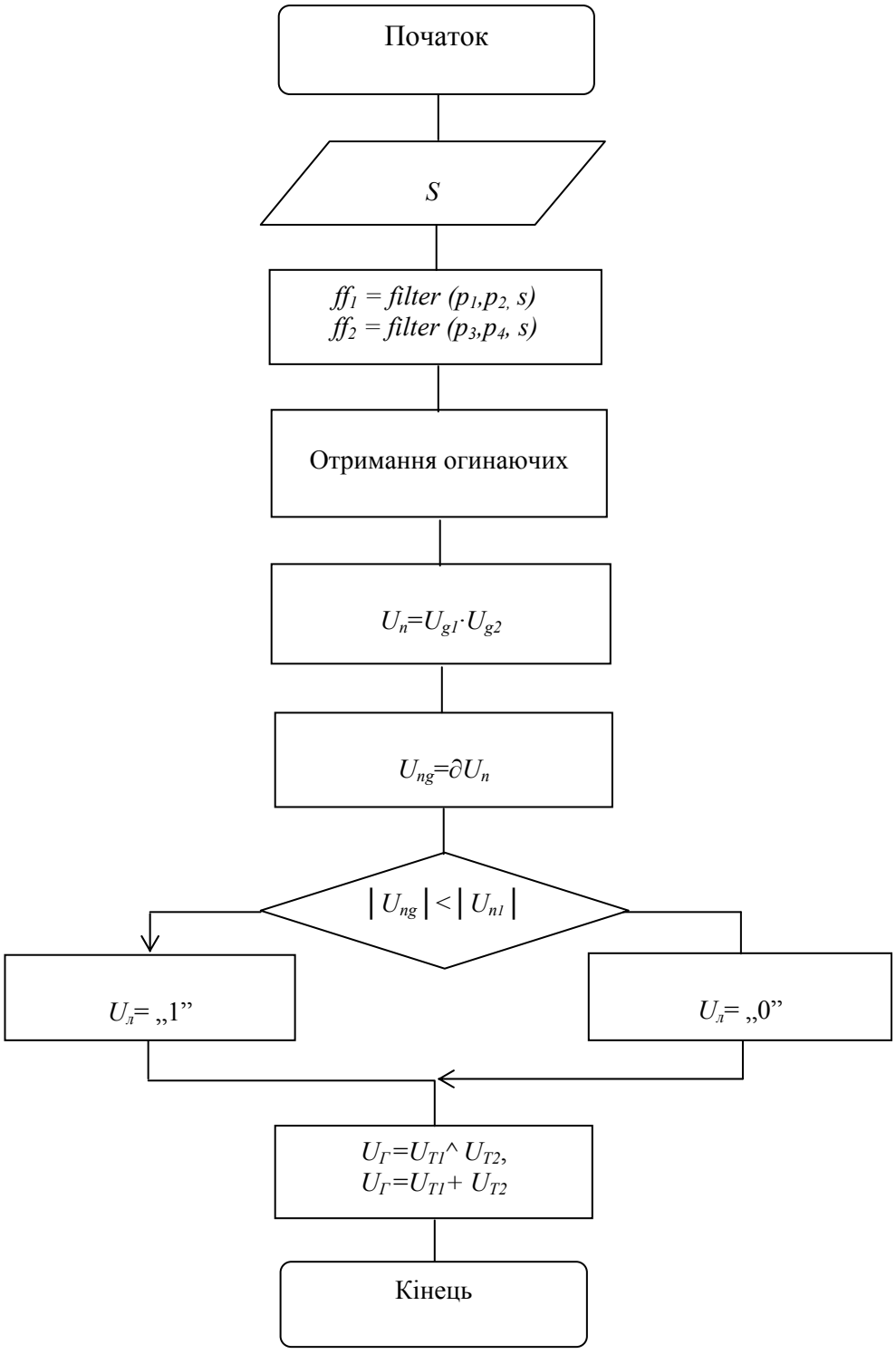

Рис. 2. Алгоритм виділення ознак складів

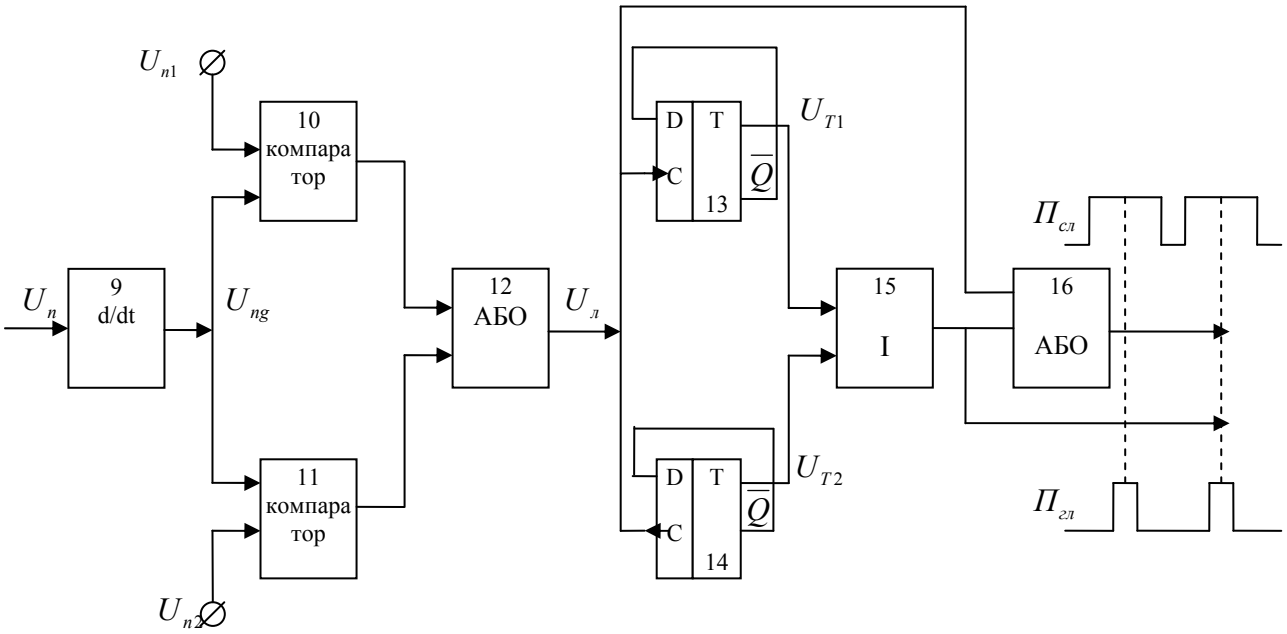

Рис. 3. Функціональна схема формувача сегментних границь складів

виході утворюється цифровий сигнал U_{T1}. Лічильний тригер із зворотним динамічним управлінням 14 перемикається спадами вхідного сигналу $U_л$ і видає на виході сигнал U_{T2}. Сигнали U_{T1} і U_{T2} надходять на входи схеми збігу 15, на виході якої утворюються короткі імпульси $U_г$, що локалізують центри складових ядер у слові. Ці імпульси поступають на перший вхід схеми АБО 16, а так само на вихід пристрою для виділення складових сегментів у слові, будучи ознакою $П_{гл}$ визначення місцеположення центру голосних звуків у складі. На другий вхід схеми АБО 16 поступає цифровий сигнал $U_л$, який схемою АБО об'єднується з сигналом $U_г$, внаслідок чого на виході схеми 16 виробляється сигнал $П_{сл}$, що виділяє в слові сегменти, відповідні положенню складів у слові. Тривалість сигналу $П_{сл}$ і кількість сигналів $П_{гл}$ використовуються системою розпізнавання для класифікації словника на підмножини, сформовані за даними ознаками.

Для тестування запропонованого методу було проведено експеримент, який полягав у сегментації 650 складів з використанням запропонованого методу. Статистична обробка експериментальних даних дозволила розрахувати надійність цьому методу, яка склала 96,4%. Тоді як надійність інших методів при еквівалентній тестуючій вибіці становить 76%.

Висновки

Запропоновані в роботі метод і алгоритм виділення ознак складів, що базуються на уточненій моделі слухової системи людини, дозволяють підвищити надійність сегментації мовного сигналу на складові сегменти і визначити такі їх ознаки, як тривалість, місцерозташування і кількість. Використання цих ознак дозволяє збільшити швидкість розпізнавання шляхом скорочення альтернатив для пошуку в 2 – 4 рази, а також підвищити надійність цього пристрою. Розроблений в роботі пристрій, що реалізує цей метод, має більшу надійність і меншу складність порівняно з відомими пристроями, а також може бути застосований при розробці автономних систем розпізнавання мови.

СПИСОК ЛІТЕРАТУРИ

1. Быков Н.М. Методы и средства измерения и преобразования информации в системах машинного распознавания речи. – Дис. на соискание уч. ст. канд. техн. наук. – Винница, ВПИ, 1985. – 243 с.

2. N.M. Bykov, I.V. Kuzmin, A.I. Yakovenko. Development of effective strategy of pattern recognition. – Proceedings of SPIE, 2001, Vol. 4225, pp.76 – 83.

3. Джелинек Ф. Разработка экспериментального устройства, распознающего раздельно произносимые слова. // Тр. ин-та инженеров по электронике и радиоэлектронике.: Пер. с англ. 1985. – Т. 73. – № 11. – С. 91 – 100.

4. Биков М.М., Грищук Т.В. Методи підвищення дикторонезалежності опису і розпізнавання мовної інформації в мережі INTERNET // "Інтернет – Освіта – Наука – 2002", третя міжнародна конференція ІОН – 2002, 8 – 12 жовтня 2002 р. Збірник матеріалів конференції. – Вінниця: УНІВЕРСУМ – Вінниця, 2002. – Том 2.– С. 329 – 332.

5. Методы автоматического распознавания речи / Под ред. У. Ли.– М.: Мир, 1983. – Т.1. – 200 с.

6. Ruske C., Schotola F. An approach to speech recognition using syllabic decision units. – Proc. 1978, IEEE ICASSP, Tulsa, 1978. – N.Y., 1978, pp. 772 – 725.

7. Ковтун В.В. Вибір інформативних ознак в задачі ідентифікації диктора // МКІМ – 2002. Міжнародна конференція з індуктивного моделювання. Львів, 20 – 25 травня 2002: Праці в 4-х томах. – Львів,ДНДІ, 2002. – .Т.1, ч. 2 – С. 280 – 287

8. Биков М.М., Грищук Т.В. Розпізнавання мовних образів з використанням нейромережевого підходу // МКІМ – 2002. Міжнародна конференція з індуктивного моделювання, Львів, 20 – 25 травня 2002: Праці в 4-х томах. – Львів: Державний НДІ інформаційної інфраструктури, 2002. - Том 1., Ч. 2. – С. 203 – 207.

Биков Микола Максимович – професор кафедри;

Ковтун Вячеслав Васильович – асистент кафедри;

Савінова Наталія Геннадіївна – студентка.

Кафедра комп'ютерних систем управління, Вінницький національний технічний університет