



УКРАЇНА

(19) UA (11) 50036 (13) U
(51) МПК (2009)
G10L 15/00

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ ДЕПАРТАМЕНТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ

ОПИС ДО ПАТЕНТУ НА КОРИСНУ МОДЕЛЬ

видається під
відповідальність
власника
патенту

(54) СПОСІБ ОПИСУВАННЯ ТА РОЗПІЗНАВАННЯ МОВЛЕННЄВИХ СИГНАЛІВ

1

(21) u200911557

(22) 13.11.2009

(24) 25.05.2010

(46) 25.05.2010, Бюл.№ 10, 2010 р.

(72) ВІНЦЮК ТАРАС КЛИМОВИЧ, ГРИЦЕНКО ВОЛОДИМИР ІЛЛІЧ

(73) МІЖНАРОДНИЙ НАУКОВО-НАВЧАЛЬНИЙ ЦЕНТР ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА СИСТЕМ

(57) Спосіб описування та розпізнавання мовленнєвих сигналів, що представлені послідовностями елементів-векторів із значень поточних параметрів його аналізу, який включає знаходження значень елементарних мір схожості кожного спостережуваного елемента на всі елементи еталонних образів, визначення схожості мовленнєвих образів шляхом рекурентного накопичення методом динамічного програмування інтегральних мір схожості послідовності спостережуваних елементів на послідовності еталонних образів, що розпізнаються, який **відрізняється** тим, що кожен елемент-вектор як еталонних образів, так і мовленнєвого сигналу, що розпізнається, описують бінарним кодом, який в цілому характеризує форму поточного амплітудного спектра мовленнєвого сигналу, а кожен окремий біт цього коду визначають як знак різниці значень амплітудного спектра для двох різних частот, як елементарну міру схожості використовують хемінгову міру, значення якої для пари порівнюваних елементів обчислюють таблично, еталонний мовленнєвий образ формують шляхом потроєння ко-

2

жного еталонного елемента початкового еталонного образу з мінімально можливою тривалістю вимовляння, утворюючи послідовності трійок, інтегральну міру схожості для першого еталонного елемента кожної із трійок знаходять як суму елементарної міри схожості цього еталонного елемента на поточний спостережуваний елемент з найбільшою із трьох інтегральних мір схожості, накопичених для попередньої трійки еталонних елементів та попереднього спостережуваного елемента мовленнєвого сигналу, що розпізнається, а інтегральні міри схожості для другого та третього еталонних елементів цієї ж трійки знаходять як суми згадуваної елементарної міри схожості та інтегральних мір схожості, накопичених для першого та другого еталонних елементів цієї ж трійки відповідно для попереднього спостережуваного елемента, при цьому інтегральні міри схожості для першого, другого та третього еталонних елементів всіх трійок, крім першої, всіх еталонних мовленнєвих образів початково, до появи першого спостережуваного елемента, набувають мінімально можливих значень, а для першого, другого та третього еталонних елементів першої трійки всіх еталонних мовленнєвих образів - максимально можливих значень, а як кінцевий результат розпізнавання вибирають той мовленнєвий еталонний образ, одна з трьох інтегральних мір схожості якого, що накопичена для останньої трійки еталонних елементів та останнього розпізнаваного елемента, є абсолютно найбільшою.

Корисна модель належить до техніки оброблення мовленнєвої інформації з метою її стискання, кодування, автоматичного розпізнавання та відтворення. Може бути використана для голосового керування пристроями.

Відомий спосіб опису та розпізнавання мовленнєвих сигналів і пристрій для його реалізації [дивись патент України №48082].

Сутність відомого способу полягає в тому, що мовленнєві сигнали представляються послідовно-

стями елементів-векторів із значень поточних параметрів його аналізу, який включає знаходження значень елементарних мір схожості кожного спостережуваного елемента на всі елементи еталонних образів, визначення схожості мовленнєвих образів шляхом рекурентного накопичення методом динамічного програмування інтегральних мір схожості послідовності спостережуваних елементів на послідовності еталонних образів, що розпізнаються.

(13) U

(11) 50036

(19) UA

Але він має недоліки, які полягають в тому, що він має недостатню швидкість та низьку робастність процесів автоматичного розпізнавання мовленнєвих сигналів.

В основу моделі покладена задача пришвидшення та підвищення робастності процесів автоматичного розпізнавання мовленнєвих сигналів за рахунок введення нових операцій розпізнавання мовленнєвих сигналів.

Поставлена задача вирішується способом опису та розпізнавання мовленнєвих сигналів, що представляються послідовностями елементів-векторів із значень поточних параметрів його аналізу, який включає знаходження значень елементарних мір схожості кожного спостережуваного елемента на всі елементи еталонних образів, визначення схожості мовленнєвих образів шляхом рекурентного накопичення методом динамічного програмування інтегральних мір схожості послідовності спостережуваних елементів на послідовності еталонних образів, що розпізнаються, при цьому кожен елемент-вектор як еталонних образів, так і мовленнєвого сигналу, що розпізнається, описується бінарним кодом, який в цілому характеризує форму поточного амплітудного спектру мовленнєвого сигналу, а кожен окремий біт цього коду визначається як знак різниці значень амплітудного спектру для двох різних частот; як елементарна міра схожості використовується хемінгова міра, значення якої для пари порівнюваних елементів обчислюється таблично; еталонний мовленнєвий образ формують шляхом потроєння кожного еталонного елемента початкового еталонного образу з мінімально можливою тривалістю вимовляння, утворюючи послідовності трійок, інтегральну міру схожості для першого еталонного елемента кожної із трійок знаходять як суму елементарної міри схожості цього еталонного елемента на поточний спостережуваний елемент з найбільшою із трьох інтегральних мір схожості, накопичених для попередньої трійки еталонних елементів та попереднього спостережуваного елемента мовленнєвого сигналу, що розпізнається, а інтегральні міри схожості для другого та третього еталонних елементів цієї ж трійки знаходять як суми згадуваної елементарної міри схожості та інтегральних мір схожості, накопичених для першого та другого еталонних елементів цієї ж трійки відповідно для попереднього спостережуваного елемента, при цьому інтегральні міри схожості для першого, другого та третього еталонних елементів всіх трійок, окрім першої, всіх еталонних мовленнєвих образів початково, до появи першого спостережуваного елемента, набувають мінімально можливих значень, а для першого, другого та третього еталонних елементів першої трійки всіх еталонних мовленнєвих образів, - максимально можливих значень, а як кінцевий результат розпізнавання вибирають той мовленнєвий еталонний образ, одна з трьох інтегральних мір схожості якого, що накопичена для останньої трійки еталонних елементів та останнього розпізнаваного елемента, є абсолютно найбільшою.

На Фіг.1 представлена структурна схема пристрою, що реалізує спосіб; на Фіг.2-4 - схеми окре-

мих блоків; Фіг.5 пояснює принцип роботи пристрою.

Пристрій містить блок 1 пам'яті мовленнєвого образу, що поданий для аналізу та розпізнавання, блок 2 пам'яті еталонного образу, обчислювач 3 елементарної міри схожості, регістр 4, обчислювач 5 інтегральної міри схожості, блоки 6,6а та 7 пам'яті, контролер 8, аналізатор 9-10 мовленнєвого сигналу,

В аналізаторі 9 мовленнєвий сигнал, що подається з мікрофона, піддається поточному спектральному аналізу в n спектральних каналах. Отже, кожний поточний спектральний елемент y_i , що спостерігається в дискретному рівномірному часі $i\Delta T$ з кроком ΔT , наприклад $\Delta T = 10$ мс, є елементом-вектором: $y_i = (y_{i1}, y_{i2}, \dots, y_{im}, \dots, y_{in})$.

В аналізаторі 10 форми спектру кожен спостережуваний елемент y_i описується вектором-матрицею x_i з двійковими компонентами

$$x_{iuv} = \begin{cases} 1, & \text{якщо } ((y_{iu} - y_{iv}) \geq 0 \ \& \ (y_{iu} \geq \Theta_u)) \\ 0 & \text{в інших випадках,} \end{cases}$$

$$v < u; u, v = 1 : n.$$

Двійковий опис-елемент x містить

$$m = \frac{1}{2}n(n-1) \text{ компонент та несе інформацію про}$$

форму спектру, визначає відносні амплітуди спектральних компонентів та не залежить від гучності промовляння.

Як правило, пороги $\Theta_u, u = 1 : n$ вибираються так, щоб в стаціонарних умовах, коли на вхід мікрофона надходять лише акустичні завади приміщення, формувались тільки нульові елементи-коди.

У випадку, коли $v = u - 1$, двійковий код x має $n - 1$ компонент, він стає дискретним аналогом знаку похідної спектру за частотою.

В блоці пам'яті 1 розпізнаваного мовленнєвого образу зберігається послідовність поточних бінарних елементів-векторів: $X_{0l} = (x_1, x_2, \dots, x_l, \dots, x_l)$, де l - довжина мовленнєвого образу.

В блоці 2 пам'яті еталонного образу зберігається еталонний образ слова, представлений аналогічною послідовністю бінарних еталонних елементів-кодів: $E_{0q} = (e_1, e_2, \dots, e_s, \dots, e_q)$, причому вона відповідає самому швидкому та все ще чіткому промовлянню слова; q - довжина початкового еталона слова.

При автоматичному порівнянні та розпізнаванні початковий еталон E_{0q} слова піддається нелінійним перетворенням в часі шляхом повторення кожного еталонного елемента початкового еталона нуль (нема повторення), один або два рази, причому так, щоб зберігався порядок слідування еталонних елементів, визначений початковим еталоном, й так, щоб в результаті у перетвореному еталоні слова було всього l еталонних елементів.

Перетворені еталонні послідовності довжини l порівнюються з мовленнєвим сигналом X_{0l} , що розпізнається, та обчислюються інтегральні міри

Г схожості сигналу X_{0l} на перетворені еталонні образи слова як суми відповідних значень елементарних мір схожості $g(x, e)$:

$$G(X_{0l}, E_{0l}) = \sum_{i=1}^l g(x_i, e_{s(i)}),$$

де функція $s(i) : s(1) = 1, s(l) = q$, - встановлює відповідність номерів еталонних та спостережуваних елементів.

Найбільше значення інтегральної міри схожості, що досягається на множині $s(i) : s(1) = 1, s(l) = q$ допустимих відповідностей визначає схожість мовленнєвого образу X_{0l} , що розпізнається, на еталонний мовленнєвий образ E_{0q} .

Елементи X_l зберігаються в блоці пам'яті 1 за адресами $0 \dots (l-1)$, що задаються сигналом з першого виходу контролера 8. Адресі 0 відповідає елемент x_1 , адресі $(l-1)$ - елемент x_l . За адресою l зберігається ознака $(l+1)$ - "Кінець образу, що розпізнається", формально відповідний елементові x_{l+1} . Ця ознака зчитується з першого виходу блоку 1 та подається на перший вхід контролера 8.

Елементи e_s зберігаються в блоці пам'яті 2 за адресами $0 \dots (q-1)$, що задаються сигналом з другого виходу контролера 8. Адресі 0 відповідає елемент e_1 , адресі $(q-1)$ - елемент e_q . За адресою q зберігається ознака $(q+1)$ - "Кінець еталонного образу". Ця ознака організована аналогічно ознаці $(l+1)$, вона зчитується з першого виходу блоку 2 й подається на другий та третій входи контролера 8 та через другий та третій входи обчислювача 5 інтегральної міри схожості - на входи установок регістрів 12, 12а і 13. За адресами $0 \dots q$ зберігаються також проміжні значення інтегральної міри схожості в блоках пам'яті 6, 6а та 7.

За адресою 0 в блоках пам'яті 6, 6а та 7 записуються відповідно початкові значення $G(0,1) = 0, G(0,2) = 0$ та $G(0,3) = 0$, а за адресами $1, 2, \dots, s, \dots, (q-1)$ цих же блоків запишемо відповідно початкові значення $G(s,1) = G(s,2) = G(s,3) = -M$, де M - велике ціле додатне число.

Перед початком порівняння та розпізнавання всі регістри, а також лічильник 21 обнуляються. В регістр 11 записується число $-M$.

Імпульсом "Старт" запускається генератор 20 контролера 8. Лічильник 21 контролера 8 формує адреси блоку пам'яті 1 та блоку пам'яті 2. Ці адреси надалі будуть позначатись як i та s відповідно. Окрім того, адреса s використовується в трьох блоках пам'яті 6, 6а та 7.

В регістр 11 записується число $-M$.

За адресою $i=0$ на виході блоку пам'яті 1 з'являється елемент x_1 у вигляді m -розрядного двійкового коду, який подається на перший вхід обчислювача елементарної міри схожості 3. Одночасно за адресою $s=0$ на виході блоку пам'яті 2 з'являється еталонний елемент e_1 , також у вигляді

ді m -розрядного двійкового коду, який подається на другий вхід обчислювача елементарної міри схожості 3. Останній обчислює значення елементарної міри схожості $g(x_1, e_1)$ - хемінгову відстань між кодами x_1 та e_1 . За адресою $s=0$ одночасно з блоків пам'яті 6, 6а та 7 зчитуються значення інтегральної міри схожості $G(0,1), G(0,2)$ і $G(0,3)$ та записуються у регістри 12, 12а та 13, відповідно. Затим зміст регістра 11 переписується в блок пам'яті 7 як $G(0,1)$. Змісти регістрів 12, 12а та 13 порівнюються в компараторі 15, і менше з посеред них через мультиплексор 16 подається на перший вхід суматора 17, на другий вхід якого надходить $g(x_1, e_1)$ з обчислювача 3. Значення $G(0,1)$ з виходу регістра 12 надходить також на перший вхід суматора 18, на другий вхід якого передається зміст регістра 4. Рівно ж значення $G(0,2)$ з виходу регістра 12а надходить також на перший вхід суматора 18а, на другий вхід якого передається зміст регістра 4а. Обчислені значення $G(1,1)$ з виходу суматора 17, $G(0,2)$ з виходу суматора 18 та $G(0,3)$ з виходу суматора 18а записуються відповідно в регістр 11 та блоки пам'яті 6 і 6а.

Далі для тієї ж самої адреси $i=0$ описаний процес повторюється для $j=1$. При цьому на виході блоку пам'яті 1 зберігається елемент x_1 , а на виході блоку пам'яті 2 - з'являється еталонний елемент e_2 . Обчислювач 3 елементарної міри схожості визначає значення $g(x_1, e_2)$ та зберігає його у регістрах 4 та 4а. З блоків пам'яті 6, 6а та 7 зчитуються значення інтегральної міри схожості $G(1,1), G(1,2)$ і $G(1,3)$ та записуються у регістри 12, 12а та 13 обчислювача 5, відповідно. Затим зміст регістра 11 переписується в блок пам'яті 7 як $G(1,1)$. Найменше із значень з виходів регістрів 12, 12а та 13, що визначається компаратором 15, через мультиплексор 16 подається на перший вхід суматора 17, на другий вхід якого надходить $g(x_1, e_2)$ з виходу обчислювача 3. Значення $G(1,1)$ з виходу регістра 12 надходить також на перший вхід суматора 18, на другий вхід якого передається зміст регістра 4. Рівно ж значення $G(1,2)$ з виходу регістра 12а надходить також на перший вхід суматора 18а, на другий вхід якого передається зміст регістра 4а. Обчислені значення $G(2,1)$ з виходу суматора 17, $G(1,2)$ з виходу суматора 18 та $G(1,3)$ з виходу суматора 18а записуються відповідно в регістр 11 та блоки пам'яті 6 і 6а.

Процес повторюється для $s = 2, 3, \dots, (q-1)$. При $s = q$ виробляється ознака $s = (q+1)$, яка блокує встановлювальні входи старших розрядів регістра 12.

Далі обчислення повторюються для $i=1$ та всіх $s = 0, 1, 2, 3, \dots, (q-1)$. Знову починаємо з того, що в регістр 11 записується число $-M$.

Далі процеси повторюються для всіх $i = 2, 3, 4, \dots, (l-1)$ та всіх $s = 0, 1, 2, 3, \dots, (q-1)$.

При досягненні $i=1$ та $s=q$ виробляється ознака $(l+1)$, яка зупиняє генератор 20.

Результат порівняння знаходиться в регістрі 11 та блоці пам'яті 6 як $G(q,1)$.

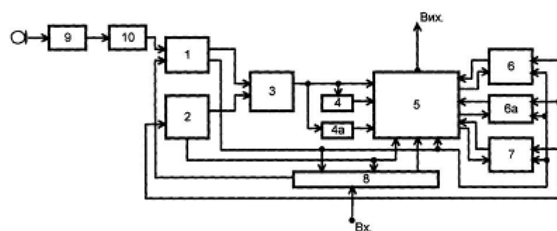


Fig. 1

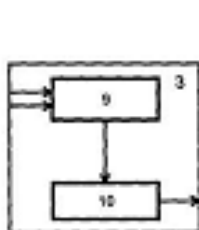


Fig. 2

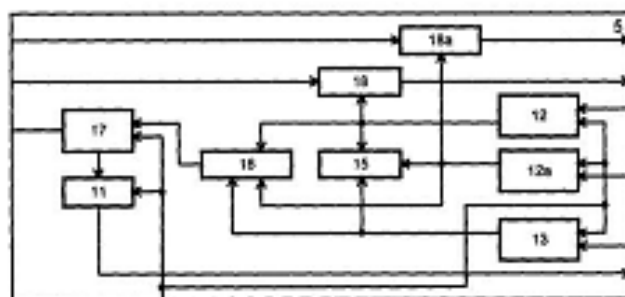


Fig. 3

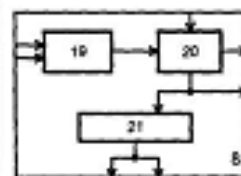


Fig. 4

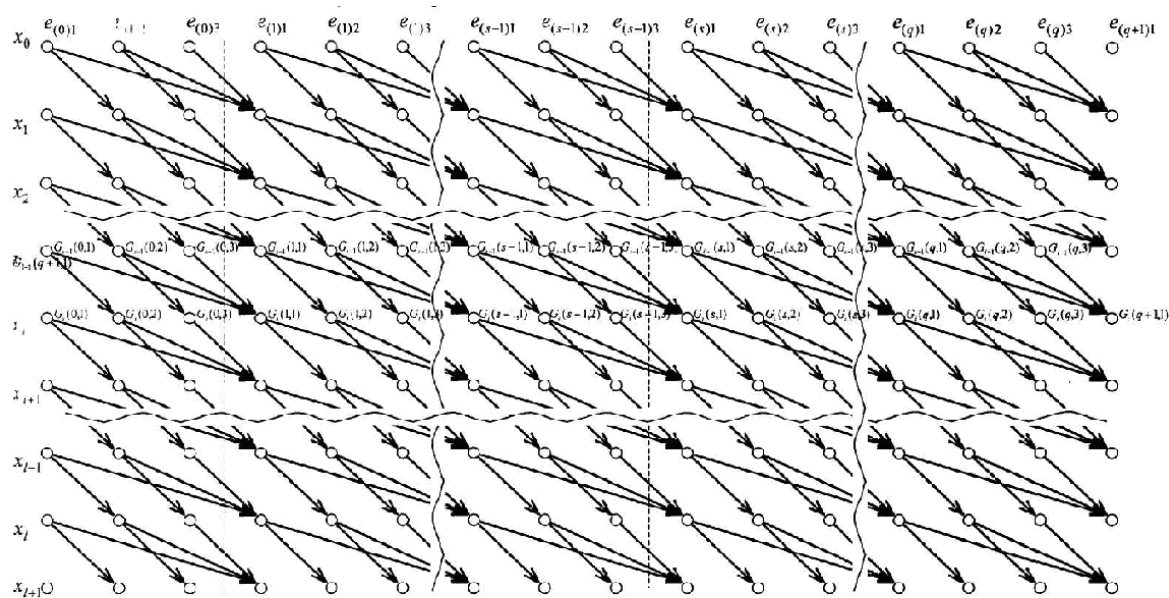


Fig. 5