



УКРАЇНА

(19) UA (11) 50039 (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) ПРИСТРІЙ ПОФОНЕМНОГО РОЗПІЗНАВАННЯ УСНИХ КОМАНД ТА УСТАЛЕНИХ СЛОВОСПОЛУЧЕНЬ

1

2

(21) u200911560

(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) Пристрій пофонемного розпізнавання усних команд та усталених словосполучень, що містить аналізатор, блок пам'яті табличних значень елементарних мір схожостей та блок пам'яті мовленнєвого сигналу, що розпізнається, обчислювач інтегральних мір схожостей та блок пам'яті проміжних результатів та контролер, який **відрізняється** тим, що в нього введені: блок пам'яті навчальної вибірки, процесор кластерного аналізу, блок пам'яті параметрів фонем; блок пам'яті орфографічного тексту та фонемної транскрипції, векторний квантувач, при цьому вихід аналізатора підк-

лючений через блок пам'яті мовленнєвих сигналів до входу векторного квантувача, а через блок пам'яті навчальної вибірки до входу процесора кластерного аналізу, виходи якого відповідно підключені до входу векторного квантувача та входу блока пам'яті параметрів фонем, на відповідні входи якого підключені виходи блока пам'яті орфографічного тексту та фонемної транскрипції, та вихід векторного квантувача, що також підключений до входу блока пам'яті табличних значень елементарних мір схожостей, виходи якого підключені до входів обчислювача інтегральних мір схожостей, відповідні виходи блока пам'яті орфографічного тексту та фонемної транскрипції підключені до відповідного входу блока пам'яті табличних значень елементарних мір схожостей та через блок пам'яті проміжних результатів до відповідного входу обчислювача інтегральних мір схожостей, а виходи контролера підключені до відповідних входів блоків пристрою.

Модель відноситься до техніки оброблення мовленнєвої інформації з метою її стиснення, кодування та автоматичного розпізнавання. Може бути використана для голосового управління пристроями.

Відомий пристрій пофонемного розпізнавання усних команд та усталених словосполучень [дивись патент України №48082].

Відомий пристрій для розпізнавання усних команд та усталених словосполучень містить аналізатор, блок пам'яті табличних значень елементарних мір схожостей та блок пам'яті мовленнєвого сигналу, що розпізнається, обчислювач інтегральних мір схожостей та блок пам'яті проміжних результатів та контролер.

Але він має недоліки, а саме малу швидкодію та низьку надійність розпізнавання.

В основу моделі покладена задача за рахунок введення конструктивних елементів створити пристрій для пофонемного розпізнавання усних команд та усталених словосполучень, що мають

високу швидкодію та надійність розпізнавання без необхідності перенастроювання системи.

Задача вирішується пристроєм для розпізнавання усних команд та усталених словосполучень, що містить аналізатор, блок пам'яті табличних значень елементарних мір схожостей та блок пам'яті мовленнєвого сигналу, що розпізнається, обчислювач інтегральних мір схожостей та блок пам'яті проміжних результатів та контролер, при цьому в нього введені: блок пам'яті для накопичення навчальної вибірки через аналізатор, процесор кластерного аналізу, вибору представників кластерів та оцінювання параметрів фонем; блок пам'яті для зберігання параметрів фонем; блок пам'яті для зберігання орфографічного тексту та фонемної транскрипції, якими супроводжується навчальна вибірка, та орфографічних текстів і фонемних транскрипцій всіх усних команд та усталених словосполучень, що входять до робочого словника; векторний квантувач 10, за вихідним сигналом якого, що вказує номер кластера, в який попадає поточний спостережуваний елемент, в

(13) U

(11) 50039

(19) UA

блоці пам'яті табличних значень визначаються значення елементарних мір належності цього спостереженого елемента до кожного із двох виходів із всіх трьох фаз фонем; обчислювач 4 інтегральних мір схожості містить сім вхідних регістрів, три компаратори та дев'ять суматорів і для кожного із двох виходів із всіх трьох фаз кожної фонемі фонетичної транскрипції всіх усних команд та усталених словосполучень накопичує інтегральні міри схожості.

На Фіг.1 представлена структурна схема пристрою, на Фіг.2 - схема одного з блоків; Фіг.3-8 пояснюють принцип роботи пристрою.

Пристрій містить аналізатор 1 мовленнєвого сигналу; блок 3 запам'ятовування мовленнєвого образу у вигляді послідовності елементів-векторів, що утворюються в результаті аналізу вхідного мовленнєвого сигналу; векторний квантувач 10, який кожному поточному спостереженому елементу-вектору ставить у відповідність номер кластеру, в який він попадає, або, що те саме, номер еталонного елемента, що представляє кластер та є найближчим, в певному розумінні, до спостереженого елемента; блок 9 введення та зберігання орфографічних текстів та фонетичних транскрипцій всіх усних команд та усталених словосполучень, що складають робочий набір та повинні розпізнаватися; блок 8 пам'яті значень параметрів моделей всіх фонем; блок 2 вибору значень елементарних мір приналежності спостережуваного елемента, за номером його кластеру, до фонем та їх фаз; обчислювач 4 інтегральних мір схожості, який накопичує, сумує, значення елементарних мір схожості для послідовності спостережуваних елементів-векторів для кожної із усних команд або кожного усталеного словосполучення відповідно до їх фонетичних транскрипцій; блок пам'яті 4а, який тимчасово зберігає накопичені інтегральні міри схожості; контролер 5, який синхронізує роботу всіх блоків, зокрема блоків 2, 4, 4а, 8, 9, 10; накопичувач 6 стандартизованої навчальної вибірки; блок 7 кластерного аналізу та обчислення параметрів фонем за навчальною вибіркою. Вихід пристрою з блоку 4 визначає номер усної команди або усталеного словосполучення, для фонетичної транскрипції якого накопичена найбільша інтегральна міра схожості.

Якщо пристрій не налаштований на голос користувача, тобто пам'ять 8 про значення параметрів моделей фонем є порожньою, він пропонує користувачеві наговорити навчальну вибірку - треба промовляти окремі слова або фрази, які головам називає пристрій.

В аналізаторі 1 мовленнєвий сигнал, що подається і мікрофона під час накопичення стандартизованої навчальної вибірки, піддається поточному автокореляційному та предиктивному аналізу в дискретному рівномірному часі і ΔT з кроком ΔT , наприклад $\Delta T=10\text{мс}$. Для поточного інтервалу аналізу і із M відліків $f_n, n=0:(M-1)$ мовленнєвого сигналу, які зважуються вікном Хемінга, обчислюються перші $m+1, m \ll M$ відліків автокореляційної функції $B_s = \sum_{n=0}^{M-1-s} f_n f_{n+s}, s=0$, які і утворюють поточний вектор автокореляції

$B_i = (B_{i0}, B_{i1}, \dots, B_{is}, \dots, B_{im})$. Розв'язуючи систему рівнянь $\sum_{u=1}^m \alpha_u B_{|u-v|} = -B_v, v=1:m$, та обчислюючи

потім $\sigma^2 = \frac{1}{M} \sum_{u=0}^{M-1} \alpha_u B_u$, де $\alpha_0=1$, описуємо кожний поточний спостережуваний елемент B_i , еквівалентним вектором-елементом (α_i, σ_i) - або (b_i, σ_i) - параметрів передбачення: $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s, \dots, \alpha_m)$. $b = (b_0, b_1, b_2, \dots, b_s, \dots, b_m)$;

$b_0 = \sum_{u=0}^m \alpha_u^2, b_s = 2 \sum_{u=0}^{m-s} \alpha_u \alpha_{u+s}, s=1:m$.

Обчислені в аналізаторі 1 елементи-вектори накопичуються в блоці 6 у вигляді послідовності як векторів автокореляцій B_i так векторів параметрів передбачення, утворюючи навчальну вибірку $X_{01} = (x_1, x_2, \dots, x_i, \dots, x_l)$ із l спостережених елементів x_i .

В блоці 7 кластерного аналізу та обчислення параметрів фонем спершу за навчальною вибіркою X_{01} з-посеред усіх l спостережених елементів x_i , за допомогою деякої ітераційної процедури вибираємо задану кількість $J, J \ll l$, елементів $(b_j^*, \sigma_j^*), j=1:J$, таких, які б найкращим чином апроксимували всі елементи навчальної вибірки:

$$(\phi_j^*, \sigma_j^*), j=1:J \rightarrow \arg \min_{(\phi_j, \sigma_j), j=1:J} \min_{j=1:J} \sum_{i=1}^J \sum_{i \in I_j} \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} (B_i, b_j) \right)$$

, де через (B, b) позначено скалярний добуток векторів-елементів B і b розмірності $m+1$, а через I_j - розбиття навчальної вибірки на J кластерів.

В останньому виразі через

$$I_j = \left\{ i : j_i = \arg \min_j \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} (B_i, b_j) \right) = j \right\} \text{ позначено } j\text{-ий кластер.}$$

Всі елементи j -го кластера найкраще апроксимуються представником цього кластера

$$(b_j^*, \sigma_j^*) = \arg \min_{(b_w, \sigma_w), w \in I_j} \left(\frac{M}{2} \ln \sigma_w^2 + \frac{1}{2\sigma_w^2} (B_i, b_w) \right) \dots$$

Ітераційний процес кластеризації починаємо з того, що на першому кроці нульової ітерації в якості представників кластерів вибирається кожний третій елемент $(b_j^{(0)}, \sigma_j^{(0)}) = (b_{3j}, \sigma_{3j}), j=1:J$ навчальної вибірки. На другому кроці нульової ітерації за знайденими представниками кластерів $(b_j^{(0)}, \sigma_j^{(0)}), j=1:J$ знаходимо самі кластери

$$I_j^{(0)} = \left\{ i : j_i^{(0)} = \arg \min_j \left(\frac{M}{2} \ln (\sigma_j^{(0)})^2 + \frac{1}{2(\sigma_j^{(0)})^2} (B_i, b_j^{(0)}) \right) = j, j=1:J \right\}$$

. Далі на першому кроці r - тої ітерації, $r=1,2,\dots$, за кластеризацією $I_j^{(r-1)}, j=1:J$, $(r-1)$ -ої ітерації вибираємо нових

$$(b_j^{(r)}, \sigma_j^{(r)}) = \arg \min_{(b_w, \sigma_w), w \in I_j^{(r-1)}} \left(\frac{M}{2} \ln \sigma_w^2 + \frac{1}{2\sigma_w^2} (B_i, b_w) \right), j=1:J$$

представників кластерів. Потім на другому кроці r -

тої ітерації, $r=1,2,\dots$, знаходимо r -ту кластеризацію

$$I_j^{(r)} = \left\{ i: j_i^{(r)} = \operatorname{argmin}_j \left(\frac{M}{2} \ln(\sigma_j^{(r)})^2 + \frac{1}{2(\sigma_j^{(r)})^2} (B_i, b_j^{(r)}) \right) = j, j = 1: J \right\}$$

І так далі. За скінчене число ітерацій досягнемо рівноваги, коли набори представників кластерів для двох сусідніх ітерацій збігатимуться. Отримані представники кластерів оголошуються еталонними елементами-векторами $(b_j, \sigma_j), j = 1: J$. Вони далі використовуватимуться у векторному квантувачеві 10 при визначенні номера еталонного елемента j_i , що є найбільш схожим на спостережуваний елемент $B_i: j_i = \operatorname{argmin}_j q(B_i, (b_j, \sigma_j))$, де

$$q(B_i, (b_j, \sigma_j)) = \left(\frac{M}{2} \ln(\sigma_j)^2 + \frac{1}{2(\sigma_j)^2} (B_i, b_j) \right) \text{ виступає як}$$

елементарна міра схожості елементів $B_i (b_j, \sigma_j)$.

В процесорі 7 також обчислюються значення параметрів моделей фонем за навчальною вибіркою. Кожна фонема φ у різному фонемному контексті подається породжувальною моделлю, що є ланцюгом 3 п'яти прихованих станів $\varphi_0, \varphi_1, \varphi_2, \varphi_3, \varphi_4$, що моделюють три стадії, три фази, розвитку процесу породження сигналів фонем. Параметрами моделей є: ймовірність $p(\varphi_1/\varphi_0)$ переходу з нульового стану в перший стан, і що дорівнює одиниці, ймовірність $p(\varphi_1/\varphi_1)$ переходу з першого стану в перший же стан та ймовірність $p(\varphi_2/\varphi_1)$ переходу з першого стану в другий стан, що доповнює до одиниці попередню ймовірність, ймовірність $p(\varphi_2/\varphi_2)$ переходу з другого стану в другий же стан та ймовірність $p(\varphi_3/\varphi_2)$ переходу з другого стану в третій стан, що доповнює до одиниці попередню ймовірність, ймовірність $p(\varphi_3/\varphi_3)$ переходу з третього стану в третій же стан та ймовірність $p(\varphi_4/\varphi_3)$ переходу з третього стану в четвертий стан, що доповнює до одиниці попередню ймовірність, та ймовірності $p(j/\varphi t), j = 1: J, t = 1, 2, 3$ попадань спо-

стережуваних елементів B в кожен із всіх кластерів $j = \operatorname{argmin}_{u=1:J} q(B, (b_u, \sigma_u))$ за умови перебування в

першому, другому та третьому станах t кожної фонем φ .

Окремо виділена фонема-пауза, що має один стан $t = 1$

Представники кластерів або еталонні елементи-вектори $(b_j, \sigma_j), j = 1: J$ та логарифми всіх ймовірностей $p(\varphi t / \varphi t), p(j / \varphi t), j = 1: J, t = 1: 3, \varphi \in \Psi$ - де Ψ - алфавіт фонем, складають індивідуальний усномовний паспорт людини. Згадані ймовірності також оцінюються за навчальною вибіркою X_{0q} за допомогою деякої іншої ітераційної процедури.

Навчальна вибірка складається з реалізацій окремо вимовлених слів, словосполучень та фраз. В свою чергу, кожне слово чи фраза описується фонетичною транскрипцією

$\Psi_{0q} = (\varphi_1, \varphi_2, \dots, \varphi_s, \dots, \varphi_q)$, де φ_s - фонема з порядковим номером s у транскрипції, q - довжина транскрипції.

Еталонні мовленнєві образи усних команд та усталених словосполучень формують шляхом об'єднання у послідовності ланцюгів породжувальних графіків фонем з п'яти прихованих станів відповідно до фонетичних транскрипцій усних команд або словосполучень, причому так, щоб вихідний, четвертий, стан попередньої фонем збігався з нульовим та першим станами наступної.

Сегмент мовленнєвого сигналу у вигляді послідовності спостережених елементів $X_{\mu\nu} = (x_{\mu+1}, x_{\mu+2}, \dots, x_\nu)$ або у вигляді послідовності спостережених символів $J_{\mu\nu} = (j_{\mu+1}, j_{\mu+2}, \dots, j_\nu)$ можна розглядати як такий, що утворений в результаті незалежних спостережень ланцюгів із трьох прихованих станів. Логарифм правдоподібності сегменту $X_{\mu\nu}$ чи $J_{\mu\nu}$, за умови фонем φ подамо виразом

$$\begin{aligned} G(J_{\mu\nu} / \varphi) = & \max_{w_1, w_2} \{ [(w_1 - \mu - 1) \ln p(\varphi_1 / \varphi_1) + \ln p(\varphi_2 / \varphi_1) + \sum_{i=\mu+1}^{w_1} \ln p(j_i / \varphi_1)] + \\ & + [(w_2 - w_1 - 1) \ln p(\varphi_2 / \varphi_2) + \ln p(\varphi_3 / \varphi_2) + \sum_{i=w_1+1}^{w_2} \ln p(j_i / \varphi_2)] + \\ & + [(v - w_2 - 1) \ln p(\varphi_3 / \varphi_3) + \ln p(\varphi_4 / \varphi_3) + \sum_{i=w_2+1}^v \ln p(j_i / \varphi_3)] \} \end{aligned}$$

де кожна складова в квадратних дужках виражає вклад кожної з трьох фаз фонем. Сегмент $X_{\mu\nu}$ чи $J_{\mu\nu}$ відноситься до тієї фонем, для котрої досягається найбільше значення виразу правдоподібності.

Якщо ж сегмент $X_{\mu\nu} = (x_{\mu+1}, x_{\mu+2}, \dots, x_\nu)$ або $J_{\mu\nu} = (j_{\mu+1}, j_{\mu+2}, \dots, j_\nu)$ розглядати як реалізацію якоїсь усної команди чи словосполучення з фонетичною транскрипцією $\Psi_{0q} = (\varphi_1, \varphi_2, \dots, \varphi_s, \dots, \varphi_q)$, то

логарифм правдоподібності цього сегмента за умови усної команди чи словосполучення Ψ_{0q} виразимо як суму логарифмів правдоподібностей за всіма s підсегментами, що відповідають окремим фонемам φ_s у послідовності Ψ_{0q} , причому кожен з цих логарифмів, в свою чергу, є сумою з трьох доданків відповідно до трьох фаз розвитку фонем:

$$\begin{aligned}
G(J_{\mu\nu} / (\varphi_1, \varphi_2, \dots, \varphi_s, \dots, \varphi_q); (\mu_s, w_{s1}, w_{s2}), s = 1: q) = \sum_{s=1}^q G(J_{\mu_{s-1}\mu_s} / (\varphi_s, w_{s1}, w_{s2})) = \\
\sum_{s=1}^q \{[(w_{s1} - \mu_{s-1} - 1) \ln p(\varphi_s 1 / \varphi_s 1) + \ln p(\varphi_s 2 / \varphi_s 1) + \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(j_i / \varphi_s 1)] + \\
+ [(w_{s2} - w_{s1} - 1) \ln p(\varphi_s 2 / \varphi_s 2) + \ln p(\varphi_s 3 / \varphi_s 2) + \sum_{i=w_{s1}+1}^{w_{s2}} \ln p(j_i / \varphi_s 2)] + \\
+ [(\mu_s - w_{s2} - 1) \ln p(\varphi_s 3 / \varphi_s 3) + \ln p(\varphi_s 4 / \varphi_s 3) + \sum_{i=w_{s2}+1}^{\mu_s} \ln p(j_i / \varphi_s 3)]\}
\end{aligned}$$

де через $J_{\mu_{s-1}\mu_s} / (\varphi_s, w_{s1}, w_{s2}) = (J_{\mu_{s-1}w_{s1}}, J_{w_{s1}w_{s2}}, J_{w_{s2}\mu_s}) = (j_{\mu_{s-1}+1}, \dots, j_{w_{s1}}; j_{w_{s1}+1}, \dots, j_{w_{s2}}; j_{w_{s2}+1}, \dots, j_{\mu_s})$

позначено можливий s-тий підсегмент для фонеми φ_s з трьома його відповідними підпідсегментами, причому $\mu_0 = \mu, \mu_{s-1} < \mu_s, \mu_q = \nu$. Оптимальне розбиття чи сегментація $((\mu_s^*, w_{s1}^*, w_{s2}^*), s = 1: q)$, яка

максимізує критерій правдоподібності, визначає інтегральну міру схожості сегмента $X_{\mu\nu}$ чи $J_{\mu\nu}$ на усну команду з фонетичною транскрипцією

$$\begin{aligned}
\psi_{0q} : G(J_{\mu\nu} / \psi_{0q}) = G(j_{\mu\nu} / (\varphi_1, \varphi_2, \dots, \varphi_s, \dots, \varphi_q); (\mu_s^*, w_{s1}^*, w_{s2}^*), s = 1: q) = \\
= \max_{((\mu_s, w_{s1}, w_{s2}), s = 1: q)} G(J_{\mu\nu} / (\varphi_1, \varphi_2, \dots, \varphi_s, \dots, \varphi_q); (\mu_s, w_{s1}, w_{s2}), s = 1: q)
\end{aligned}$$

Сегмент $X_{\mu\nu}$ чи $J_{\mu\nu}$ відноситиметься до тієї усної команди чи усталеного словосполучення, для котрої чи котрого досягатиметься абсолютно найбільше значення виразу правдоподібності.

Вся ж навчальна вибірка $X_{0l} = (x_1, x_2, \dots, x_i, \dots, x_l)$ чи $J_{0l} = (j_1, j_2, \dots, j_i, \dots, j_l)$, що відповідає стандартизованому тексту, складається з реалізацій окремо вимовлених слів, словосполучень та фраз, границі котрих -- початок та кінець початкової та кінцевої фонем-пауз - у навчальній

вибірці визначаються автоматично в процесі накопичення - введення - реалізацій. $\psi_{0Q} = (\varphi_1, \varphi_2, \dots, \varphi_s, \dots, \varphi_Q)$ - фонетична транскрипція всієї навчальної вибірки з Q реалізацій фонем.

Максимально правдоподібні оцінка ймовірностей $P = \{p(\varphi_t / \varphi_t), p(j / \varphi_t), j = 1: J, t = 1: 3, \varphi \in \psi\}$, де ψ - алфавіт фонем, обчислюються процесором 7 шляхом максимізації критерію правдоподібності для навчальної вибірки

$$\begin{aligned}
G(J_{0Q} / \psi_{0Q}; P; \{(\mu_s, w_{s1}, w_{s2}), s = 1: q\}) = \\
= \sum_{s=1}^Q \{[(w_{s1} - \mu_{s-1} - 1) \ln p(\varphi_s 1 / \varphi_s 1) + \ln p(\varphi_s 2 / \varphi_s 1) + \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(j_i / \varphi_s 1)] + \\
+ [(w_{s2} - w_{s1} - 1) \ln p(\varphi_s 2 / \varphi_s 2) + \ln p(\varphi_s 3 / \varphi_s 2) + \sum_{i=w_{s1}+1}^{w_{s2}} \ln p(j_i / \varphi_s 2)] + \\
+ [(\mu_s - w_{s2} - 1) \ln p(\varphi_s 3 / \varphi_s 3) + \ln p(\varphi_s 4 / \varphi_s 3) + \sum_{i=w_{s2}+1}^{\mu_s} \ln p(j_i / \varphi_s 3)]\}.
\end{aligned}$$

Критерій правдоподібності можна переписати в де до іншій еквівалентній формі, згрупувавши окремо всі реалізації однієї й тієї ж фонемі:

$$\begin{aligned}
G(J_{0Q} / \psi_{0Q}; P; \{(\mu_s, w_{s1}, w_{s2}), s = 1: Q\}) = \\
= \sum_{\varphi \in \psi} \sum_{s: \varphi_s = \varphi} \{[(w_{s1} - \mu_{s-1} - 1) \ln p(\varphi 1 / \varphi 1) + \ln p(\varphi 2 / \varphi 1) + \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(j_i / \varphi 1)] + \\
+ [(w_{s2} - w_{s1} - 1) \ln p(\varphi_s 2 / \varphi_s 2) + \ln p(\varphi_s 3 / \varphi_s 2) + \sum_{i=w_{s1}+1}^{w_{s2}} \ln p(j_i / \varphi_s 2)] + \\
+ [(\mu_s - w_{s2} - 1) \ln p(\varphi_s 3 / \varphi_s 3) + \ln p(\varphi_s 4 / \varphi_s 3) + \sum_{i=w_{s2}+1}^{\mu_s} \ln p(j_i / \varphi_s 3)]\}.
\end{aligned}$$

З останнього запису випливає, що якщо відомі границі $\{(\mu_s, w_{s1}, w_{s2}), s = 1: Q\}$ сегментів всіх реалізацій всіх фонем із навчальної вибірки, то максимально правдоподібні оцінки ймовірностей обчислюються як:

$$\begin{aligned}
p(\varphi_1 / \varphi_1) &= (\sum_{s: \varphi_s = \varphi} (w_{s1} - \mu_{s-1} - 1)) / (\sum_{s: \varphi_s = \varphi} (w_{s1} - \mu_{s-1})), \\
p(\varphi_2 / \varphi_1) &= (\sum_{s: \varphi_s = \varphi} 1) / (\sum_{s: \varphi_s = \varphi} (w_{s1} - \mu_{s-1})), \\
p(\varphi_2 / \varphi_2) &= (\sum_{s: \varphi_s = \varphi} (w_{s2} - w_{s1} - 1)) / (\sum_{s: \varphi_s = \varphi} (w_{s2} - w_{s1})), \\
p(\varphi_3 / \varphi_2) &= (\sum_{s: \varphi_s = \varphi} 1) / (\sum_{s: \varphi_s = \varphi} (w_{s2} - w_{s1})), \\
p(\varphi_3 / \varphi_3) &= (\sum_{s: \varphi_s = \varphi} (\mu_s - w_{s2} - 1)) / (\sum_{s: \varphi_s = \varphi} (\mu_s - w_{s2})), \\
p(\varphi_4 / \varphi_3) &= (\sum_{s: \varphi_s = \varphi} 1) / (\sum_{s: \varphi_s = \varphi} (\mu_s - w_{s2})), \\
p(j / \varphi_1) &= (\sum_{s: \varphi_s = \varphi} \sum_{i=\mu_{s-1}+1}^{w_{s1}} j_i) / (\sum_{s: \varphi_s = \varphi} (w_{s1} - \mu_{s-1})), \\
p(j / \varphi_2) &= (\sum_{s: \varphi_s = \varphi} \sum_{i=w_{s1}+1}^{w_{s2}} j_i) / (\sum_{s: \varphi_s = \varphi} (w_{s2} - w_{s1})), \\
p(j / \varphi_3) &= (\sum_{s: \varphi_s = \varphi} \sum_{i=w_{s2}+1}^{\mu_s} j_i) / (\sum_{s: \varphi_s = \varphi} (\mu_s - w_{s2})), j = 1: J, \varphi \in \Psi.
\end{aligned}$$

Ітераційний процес обчислення параметрів Р починаємо з того, що на першому кроці нульової ітерації знаходимо початкову сегментацію всієї навчальної вибірки $\{(\mu_s^{(0)}, w_{s1}^{(0)}, w_{s2}^{(0)}), s = 1: Q\}$. Для цього виокремимо із навчальної вибірки сегменти, що відповідають окремо вимовленим словам, словосполученням чи фразам. Далі для кожного таким чином виділеного сегмен-

та $X_{0l} = (x_1, x_2, \dots, x_i, \dots, x_l)$, , відповідно до його фонетичної транскрипції $\Psi_{0q} = (\varphi_1, \varphi_2, \dots, \varphi_q)$, виконаємо його оптимальне розбиття на q сегментів, виходячи із моделі однорідності сегментів та апелюючи до вже знайдених еталонних елементів - представників кластерів $(b_j, \sigma_j), j = 1: J$:

$$\begin{aligned}
\{(\mu_s^{(0)}, w_{s1}^{(0)}, w_{s2}^{(0)}), s = 1: q\} &= \operatorname{argmin}_{\{(\mu_s, w_{s1}, w_{s2}), s = 1: q\}} \{ \min_j \sum_{i=1}^{\mu_1} (\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} (B_i, b_j)) + \\
&+ \sum_{s=2}^{q-1} \{ [\min_j \sum_{i=\mu_{s-1}+1}^{w_{s1}} (\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} (B_i, b_j))] + [\min_j \sum_{i=w_{s1}+1}^{w_{s2}} (\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} (B_i, b_j))] + \\
&+ [\min_j \sum_{i=w_{s2}+1}^{\mu_s} (\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} (B_i, b_j))] \} + \min_j \sum_{i=\mu_{q-1}+1}^{w_q} (\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} (B_i, b_j)) \}.
\end{aligned}$$

Оптимальне розбиття знаходимо як найкоротший шлях на графі із вершини $(0, 0)$ у вершину (l, q) . Це ілюструється Фіг.7.

Повторивши процедуру розбиття для всіх реалізацій слів та фраз, отримуємо початкову сегментацію $\{(\mu_s^{(0)}, w_{s1}^{(0)}, w_{s2}^{(0)}), s = 1: Q\}$ всієї навчальної вибірки.

Потім на другому кроці нульової ітерації за відомою сегментацією $\{(\mu_s^{(0)}, w_{s1}^{(0)}, w_{s2}^{(0)}), s = 1: Q\}$ на-

навчальної вибірки знаходимо ймовірності $P^{(0)} = \{p^{(0)}(\varphi_t / \varphi_t), p^{(0)}(j / \varphi_t), j = 1: J, t = 1: 3, \varphi \in \Psi\}$

Далі на першому кроці r -тої ітерації, $r = 1, 2, \dots$ за параметрами моделей фонем $P^{(r-1)} = \{p^{(r-1)}(\varphi_t / \varphi_t), p^{(r-1)}(j / \varphi_t), j = 1: J, t = 1: 3, \varphi \in \Psi\}$, знайденими на $(r-1)$ -шій ітерації, вибираємо нову сегментацію

$$\begin{aligned}
\{(\mu_s^{(r)}, w_{s1}^{(r)}, w_{s2}^{(r)}), s = 1: Q\} &= \\
&= \operatorname{argmax}_{\{(\mu_s, w_{s1}, w_{s2}), s = 1: Q\}} \sum_{\varphi \in \Psi} \sum_{s: \varphi_s = \varphi} \{ [(w_{s1} - \mu_{s-1} - 1) \ln p^{(r-1)}(\varphi_1 / \varphi_1) + \ln p^{(r-1)}(\varphi_2 / \varphi_1) + \\
&+ \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p^{(r-1)}(j_i / \varphi_2)] + \\
&+ [(w_{s2} - w_{s1} - 1) \ln p^{(r-1)}(\varphi_2 / \varphi_2) + \ln p^{(r-1)}(\varphi_3 / \varphi_2) + \sum_{i=w_{s1}+1}^{w_{s2}} \ln p^{(r-1)}(j_i / \varphi_2)] + \\
&+ [(\mu_s - w_{s2} - 1) \ln p^{(r-1)}(\varphi_3 / \varphi_3) + \ln p^{(r-1)}(\varphi_4 / \varphi_3) + \sum_{i=w_{s2}+1}^{\mu_s} \ln p^{(r-1)}(j_i / \varphi_3)] \}
\end{aligned}$$

навчальної вибірки та на її основі - нові значення параметрів моделей фонем

$$\begin{aligned}
p^{(r)} &= p^{(r)}_{\phi} \left(\frac{t}{\phi t} \right) p^{(r)}_{\phi} \left(\frac{t}{\phi t} \right) = 1: J, t = 1: 3, \phi \in \Psi; j \\
p^{(r)}(\phi 1 / \phi 1) &= (\sum_{s: \phi_s = \phi} (w_{s1}^{(r)} - \mu_{s-1}^{(r)} - 1)) / (\sum_{s: \phi_s = \phi} (w_{s1}^{(r)} - \mu_{s-1}^{(r)})), \\
p^{(r)}(\phi 2 / \phi 2) &= (\sum_{s: \phi_s = \phi} (w_{s2}^{(r)} - w_{s1}^{(r)} - 1)) / (\sum_{s: \phi_s = \phi} (w_{s2}^{(r)} - w_{s1}^{(r)})), \\
p^{(r)}(\phi 3 / \phi 3) &= (\sum_{s: \phi_s = \phi} (\mu_s^{(r)} - w_{s2}^{(r)} - 1)) / (\sum_{s: \phi_s = \phi} (\mu_s^{(r)} - w_{s2}^{(r)})), \\
p^{(r)}(j / \phi 1) &= (\sum_{s: \phi_s = \phi} \sum_{i=\mu_{s-1}^{(r)}+1}^{w_{s1}^{(r)}} 1) j_i = jD / (\sum_{s: \phi_s = \phi} (w_{s1}^{(r)} - \mu_{s-1}^{(r)})), \\
p^{(r)}(j / \phi 2) &= (\sum_{s: \phi_s = \phi} \sum_{i=w_{s1}^{(r)}+1}^{w_{s2}^{(r)}} 1) j_i = jD / (\sum_{s: \phi_s = \phi} (w_{s2}^{(r)} - w_{s1}^{(r)})), \\
p^{(r)}(j / \phi 3) &= (\sum_{s: \phi_s = \phi} \sum_{i=\mu_{s2}^{(r)}+1}^{\mu_s^{(r)}} 1) j_i = jD / (\sum_{s: \phi_s = \phi} (\mu_s^{(r)} - w_{s2}^{(r)})), j = 1: J, \phi \in \Psi.
\end{aligned}$$

За скінчене число ітерацій досягнемо рівноваги, коли значення параметрів моделей фонем для двох сусідніх ітерацій збігатимуться. Отримані значення параметрів моделей фонем пересилаються у блок 8, де і зберігаються.

В разі заповнення блоків 7 та 8 еталонними елементами та значеннями параметрів моделей фонем, відповідно, а блока 9 - робочим словником

усних команд та усталених словосполучень разом з їхніми фонетичними транскрипціями, пристрій стає готовим до автоматичного розпізнавання: на кожну усну команду з номером k в блоці пам'яті 4а проміжних результатів робляться початкові установки $F_{0,2}^k(\phi_0 3) = 0$;

$$(F_1^k(\phi_s 1) = -M, F_2^k(\phi_s 1) = -M, F_1^k(\phi_s 2) = -M, F_2^k(\phi_s 2) = -M, F_1^k(\phi_s 3) = -M, F_2^k(\phi_s 3) = -M), s = 1: q^k, k = 1: K,$$

де $(-M)$ - велике від'ємне число, q^k - довжина транскрипції k -го слова, K - обсяг словника.

Контролер 5, як тільки отримує інформацію від аналізатора 1 про появу першого спостереженого елемента B_1 забезпечує: пересилку його на векторний квантувач 10, де визначається номер еталонного елемента $j_i = \arg \min_{j=1:J} (\frac{M}{2} \ln \sigma_j^2 + (B_1, b_j))$, що є найбільш схожим на спостережуваний. Під управлінням контролера 5 з блока 9 зчитується ім'я першої фонемі ϕ_1 першого слова. За цим іменем фонемі та номером еталонного елемента j_1 з блока пам'яті 4а накопичуваних інтегральних мір схожості на обчислювач 4 інтегральних мір

схожості зчитуються значення $F_{0,2}^k(\phi_0 3)$, $F_{0,1}^k(\phi_1 1)$, $F_{0,2}^k(\phi_1 1)$, $F_{0,1}^k(\phi_1 2)$, $F_{0,2}^k(\phi_1 2)$, $F_{0,1}^k(\phi_1 3)$, $F_{0,2}^k(\phi_1 3)$ на регістри 41, 42, 42а, 43, 43а, 44, 44а відповідно. Одночасно за номером еталонного елемента j_i , та ім'ям фонемі ϕ_1 на другі входи суматорів 46, 47, 47а, 48, 49, 49а, 50, 51, 51а надходять логарифми ймовірностей $\ln p(j_1 / \phi_1 1)$, $\ln p(\phi_1 1 / \phi_1 1)$, $\ln p(\phi_1 2 / \phi_1 1)$, $\ln p(j_1 / \phi_1 2)$, $\ln p(\phi_1 2 / \phi_1 2)$, $\ln p(\phi_1 3 / \phi_1 2)$, $\ln p(j_1 / \phi_1 3)$, $\ln p(\phi_1 3 / \phi_1 3)$, $\ln p(\phi_1 4 / \phi_1 3)$ відповідно.

Оскільки порівнювані 45, 45а, 45б вибирають більше з двох чисел, то на виходах суматорів 47, 47а, 49, 49а, 51, 51а матимемо відповідно

$$\begin{aligned}
F_{1,1}^k(\phi_1 1) &= \max (F_{0,2}^k(\phi_0 3), F_{0,1}^k(\phi_1 1)) + \ln p(j_1 / \phi_1 1) + \ln p(\phi_1 1 / \phi_1 1), \\
F_{1,2}^k(\phi_1 1) &= \max (F_{0,2}^k(\phi_0 3), F_{0,1}^k(\phi_1 1)) + \ln p(j_1 / \phi_1 1) + \ln p(\phi_1 2 / \phi_1 1) \\
F_{1,1}^k(\phi_1 2) &= \max (F_{0,2}^k(\phi_1 1), F_{0,1}^k(\phi_1 2)) + \ln p(j_1 / \phi_1 2) + \ln p(\phi_1 2 / \phi_1 2) \\
F_{1,2}^k(\phi_1 2) &= \max (F_{0,2}^k(\phi_1 1), F_{0,1}^k(\phi_1 2)) + \ln p(j_1 / \phi_1 2) + \ln p(\phi_1 3 / \phi_1 2) \\
F_{1,1}^k(\phi_1 3) &= \max (F_{0,2}^k(\phi_1 2), F_{0,1}^k(\phi_1 3)) + \ln p(j_1 / \phi_1 3) + \ln p(\phi_1 3 / \phi_1 3) \\
F_{1,2}^k(\phi_1 3) &= \max (F_{0,2}^k(\phi_1 2), F_{0,1}^k(\phi_1 3)) + \ln p(j_1 / \phi_1 3) + \ln p(\phi_1 4 / \phi_1 3)
\end{aligned}$$

Далі значення регістра 44а пересилаємо в регістр 41, щоб підготуватись до оброблення наступної фонемі слова k , з тим обчислена вище значення інтегральних мір схожості

$$F_{1,1}^k(\phi_1 1), F_{1,2}^k(\phi_1 1), F_{1,1}^k(\phi_1 2), F_{1,2}^k(\phi_1 2), F_{1,1}^k(\phi_1 3),$$

$F_{1,2}^k(\phi_1 3)$ пересилаємо на зберігання в блок пам'яті

4а за адресами значень $F_{0,1}^k(\phi_1 1)$, $F_{0,2}^k(\phi_1 1)$,

$F_{0,1}^k(\phi_1 2)$, $F_{0,2}^k(\phi_1 2)$, $F_{0,1}^k(\phi_1 3)$, $F_{0,2}^k(\phi_1 3)$ відповідно.

Далі за тим же номером еталонного елемента j_i , послідовно для всіх решти фонем $s = 2: q^k$ слова k діємо аналогічно. За іменем фонемі ϕ_s з блока пам'яті 4а накопичуваних інтегральних мір схожості на обчислювач 4 інтегральних мір схожості зчитуються значення $F_{0,1}^k(\phi_{s-1} 1)$, $F_{0,2}^k(\phi_{s-1} 1)$, $F_{0,1}^k(\phi_{s-1} 2)$, $F_{0,2}^k(\phi_{s-1} 2)$, $F_{0,1}^k(\phi_{s-1} 3)$, $F_{0,2}^k(\phi_{s-1} 3)$ на ті ж регістри 41, 42, 42а, 43, 43а, 44, 44а відповідно. Одночасно за номером еталонного елемента j_i , та ім'ям

фонемі φ_s на другі входи суматорів 46, 47, 47а, 48, 49, 49а, 50, 51, 51а надходять логарифми ймовірностей

$$\ln p(j_1 / \varphi_s 1), \ln p(\varphi_s 1 / \varphi_s 1), \ln p(\varphi_s 2 / \varphi_s 1),$$

$\ln p(j_1 / \varphi_s 2), \ln p(\varphi_s 2 / \varphi_s 2), \ln p(\varphi_s 3 / \varphi_s 2), \ln p(j_1 / \varphi_s 3), \ln p(\varphi_s 3 / \varphi_s 3), \ln p(\varphi_s 4 / \varphi_s 3)$, відповідно. На виходах суматорів 47, 47а, 49, 49а, 51, 51а матимемо відповідно

$$\begin{aligned} F_{1,1}^k(\varphi_s 1) &= \max (F_{0,2}^k(\varphi_{s-2} 3), F_{0,1}^k(\varphi_{s-1} 1)) + \ln p(j_1 / \varphi_s 1) + \ln p(\varphi_s 1 / \varphi_s 1), \\ F_{1,2}^k(\varphi_s 1) &= \max (F_{0,2}^k(\varphi_{s-2} 3), F_{0,1}^k(\varphi_{s-1} 1)) + \ln p(j_1 / \varphi_s 1) + \ln p(\varphi_s 2 / \varphi_s 1) \\ F_{1,1}^k(\varphi_s 2) &= \max (F_{0,2}^k(\varphi_{s-1} 1), F_{0,1}^k(\varphi_{s-1} 2)) + \ln p(j_1 / \varphi_s 2) + \ln p(\varphi_s 2 / \varphi_s 2) \\ F_{1,2}^k(\varphi_s 2) &= \max (F_{0,2}^k(\varphi_{s-1} 1), F_{0,1}^k(\varphi_{s-1} 2)) + \ln p(j_1 / \varphi_s 2) + \ln p(\varphi_s 3 / \varphi_s 2) \\ F_{1,1}^k(\varphi_s 3) &= \max (F_{0,2}^k(\varphi_{s-1} 2), F_{0,1}^k(\varphi_{s-1} 3)) + \ln p(j_1 / \varphi_s 3) + \ln p(\varphi_s 3 / \varphi_s 3) \\ F_{1,2}^k(\varphi_s 3) &= \max (F_{0,2}^k(\varphi_{s-1} 2), F_{0,1}^k(\varphi_{s-1} 3)) + \ln p(j_1 / \varphi_s 3) + \ln p(\varphi_s 4 / \varphi_s 3) \end{aligned}$$

Далі значення регістра 44а пересилаємо в регістр 41, затим обчислені вище значення інтегральних мір схожості

$$F_{1,1}^k(\varphi_s 1), F_{1,2}^k(\varphi_s 1), F_{1,1}^k(\varphi_s 2), F_{1,2}^k(\varphi_s 2), F_{1,1}^k(\varphi_s 3),$$

$F_{1,2}^k(\varphi_s 3)$ пересилаємо на зберігання в блок пам'яті 4а за адресами значень $F_{0,1}^k(\varphi_{s-1} 1), F_{0,2}^k(\varphi_{s-1} 1), F_{0,1}^k(\varphi_{s-1} 2), F_{0,2}^k(\varphi_{s-1} 2), F_{0,1}^k(\varphi_{s-1} 3), F_{0,2}^k(\varphi_{s-1} 3)$ відповідно.

Після закінчення оброблення слова k : за адресом $F_{0,2}^k(\varphi_0 3)$ пересилаємо (-M).

Далі процес повторюємо в циклі по $k = 2 \dots K$.

Далі процес повторюється в циклі по $i = 2 \dots I$

При порівнянні спостережуваного сигналу $X_{0i} = (x_1, x_2, \dots, x_i)$, відповідно $J_{0i} = (j_1, j_2, \dots, j_i)$, з фонетичними транскрипціями усних команд чи усталених словосполучень інтегральні міри схожості $F_{1,1}^k(\varphi_s 1), F_{1,2}^k(\varphi_s 1), F_{1,1}^k(\varphi_s 2), F_{1,2}^k(\varphi_s 2), F_{1,1}^k(\varphi_s 3), F_{1,2}^k(\varphi_s 3)$ ЩО визначаються для кожного із двох виходів із першого, другого чи третього станів по-

точної фонемі їх транскрипцій, знаходять як суми значень відповідних елементарних мір схожості, обчислених для поточного спостереженого елемента j_i , для кожного із двох виходів із першого, другого чи третього станів поточної фонемі

$$\ln p(j_i / \varphi_s 1), \ln p(\varphi_s 1 / \varphi_s 1), \ln p(\varphi_s 2 / \varphi_s 1),$$

$$\ln p(j_i / \varphi_s 2), \ln p(\varphi_s 2 / \varphi_s 2), \ln p(\varphi_s 3 / \varphi_s 2),$$

$$\ln p(j_i / \varphi_s 3), \ln p(\varphi_s 3 / \varphi_s 3), \ln p(\varphi_s 4 / \varphi_s 3),$$

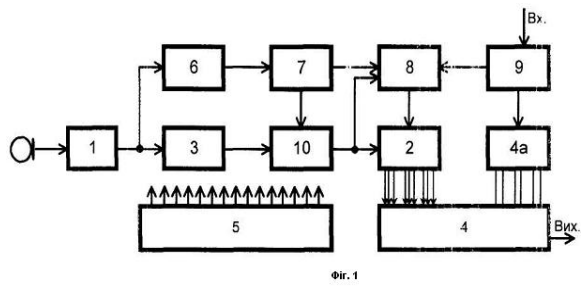
з більшою із двох інтегральних мір схожості, накопичених для попереднього спостереженого елемента j_{i-1} на другому виході з третього стану попередньої фонемі $\varphi_{s-1} 3$ та на першому виході з першого стану поточної фонемі $\varphi_s 1$, на другому виході з першого стану поточної фонемі $\varphi_s 1$ та на першому виході з другого стану поточної фонемі $\varphi_s 2$ й на другому виході з другого стану поточної фонемі $\varphi_s 2$ та на першому виході з третього стану поточної фонемі $\varphi_s 3$, відповідно:

$$\begin{aligned} F_{1,1}^k(\varphi_s 1) &= \max (F_{i-1,2}^k(\varphi_{s-2} 3), F_{i-1,1}^k(\varphi_{s-1} 1)) + \ln p(j_i / \varphi_s 1) + \ln p(\varphi_s 1 / \varphi_s 1), \\ F_{1,2}^k(\varphi_s 1) &= \max (F_{i-1,2}^k(\varphi_{s-2} 3), F_{i-1,1}^k(\varphi_{s-1} 1)) + \ln p(j_i / \varphi_s 1) + \ln p(\varphi_s 2 / \varphi_s 1) \\ F_{1,1}^k(\varphi_s 2) &= \max (F_{i-1,2}^k(\varphi_{s-1} 1), F_{i-1,1}^k(\varphi_{s-1} 2)) + \ln p(j_i / \varphi_s 2) + \ln p(\varphi_s 2 / \varphi_s 2) \\ F_{1,2}^k(\varphi_s 2) &= \max (F_{i-1,2}^k(\varphi_{s-1} 1), F_{i-1,1}^k(\varphi_{s-1} 2)) + \ln p(j_i / \varphi_s 2) + \ln p(\varphi_s 3 / \varphi_s 2) \\ F_{1,1}^k(\varphi_s 3) &= \max (F_{i-1,2}^k(\varphi_{s-1} 2), F_{i-1,1}^k(\varphi_{s-1} 3)) + \ln p(j_i / \varphi_s 3) + \ln p(\varphi_s 3 / \varphi_s 3) \\ F_{1,2}^k(\varphi_s 3) &= \max (F_{i-1,2}^k(\varphi_{s-1} 2), F_{i-1,1}^k(\varphi_{s-1} 3)) + \ln p(j_i / \varphi_s 3) + \ln p(\varphi_s 4 / \varphi_s 3) \end{aligned}$$

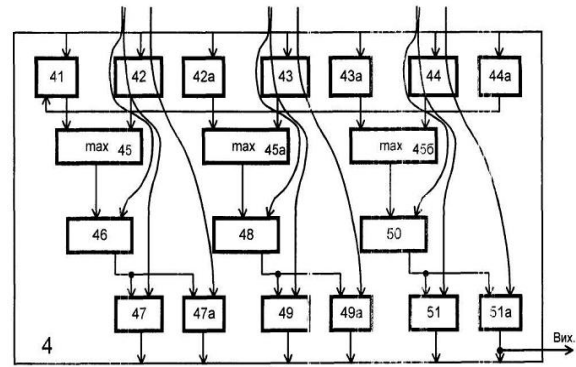
Значення інтегральної міри схожості $F_{1,2}^k(\varphi_q 3)$, накопичене після оброблення останнього спостереженого елемента на другому виході із третього стану останньої фонемі q^k , яка визначається фонетичною транскрипцією усної команди або словосполучення, визначає схожість пред'явленого мо-

вленнєвого сигналу на цю усну команду або словосполучення.

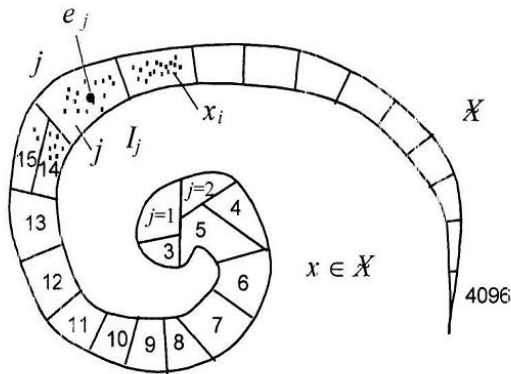
Пред'явлений мовленнєвий сигнал контролером 5 відноситься до тієї усної команди або усталеного словосполучення, для якого накопичена схожість є абсолютно найбільшою.



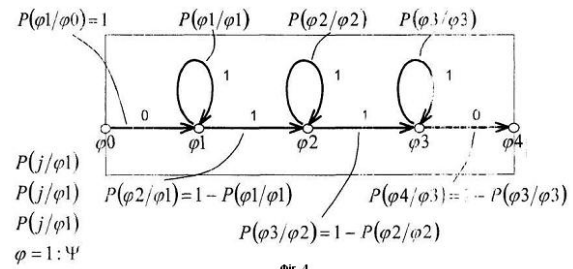
Фиг. 1



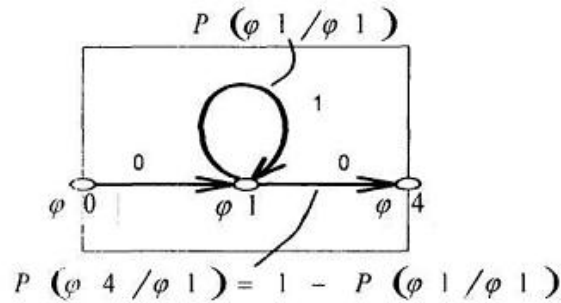
Фиг. 2



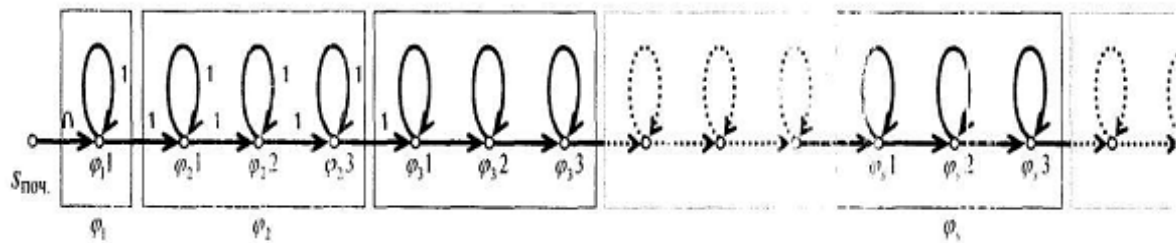
Фиг. 3



Фиг. 4



Фиг. 5



Фиг. 6

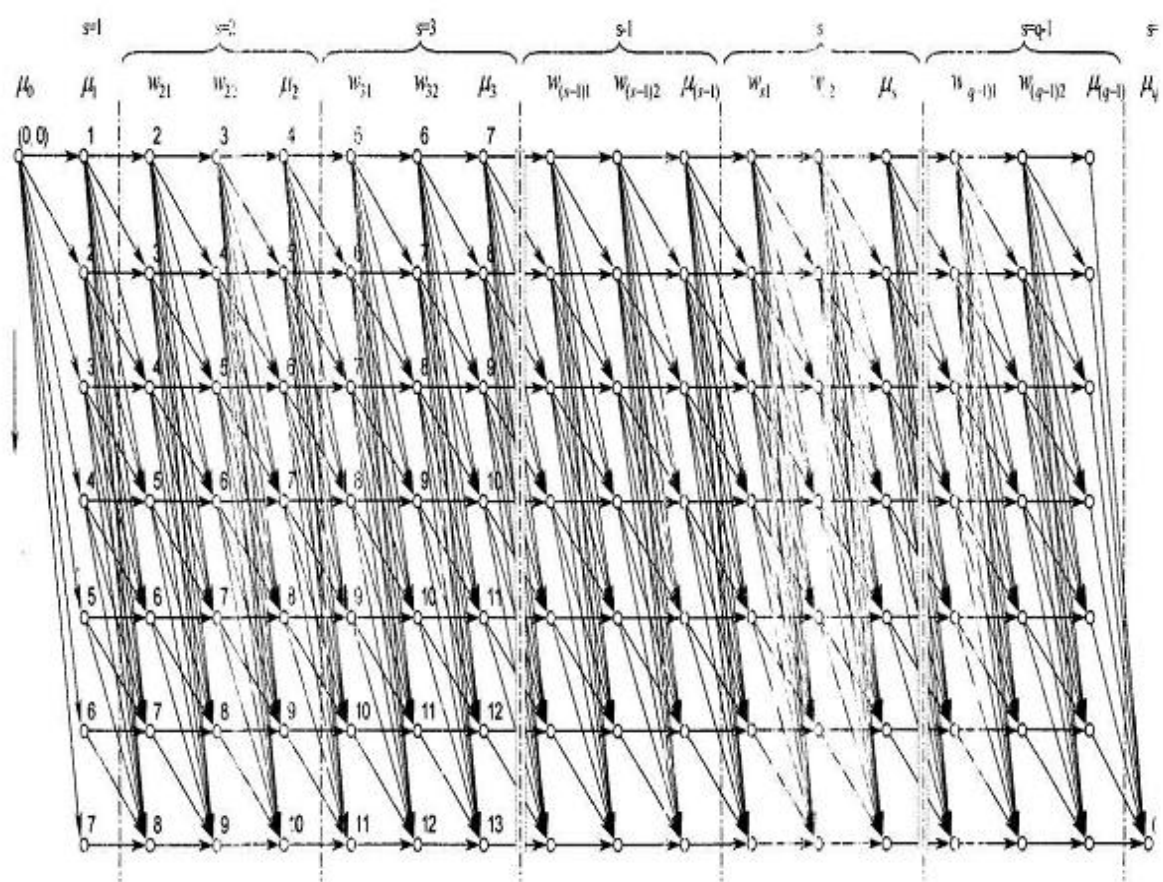
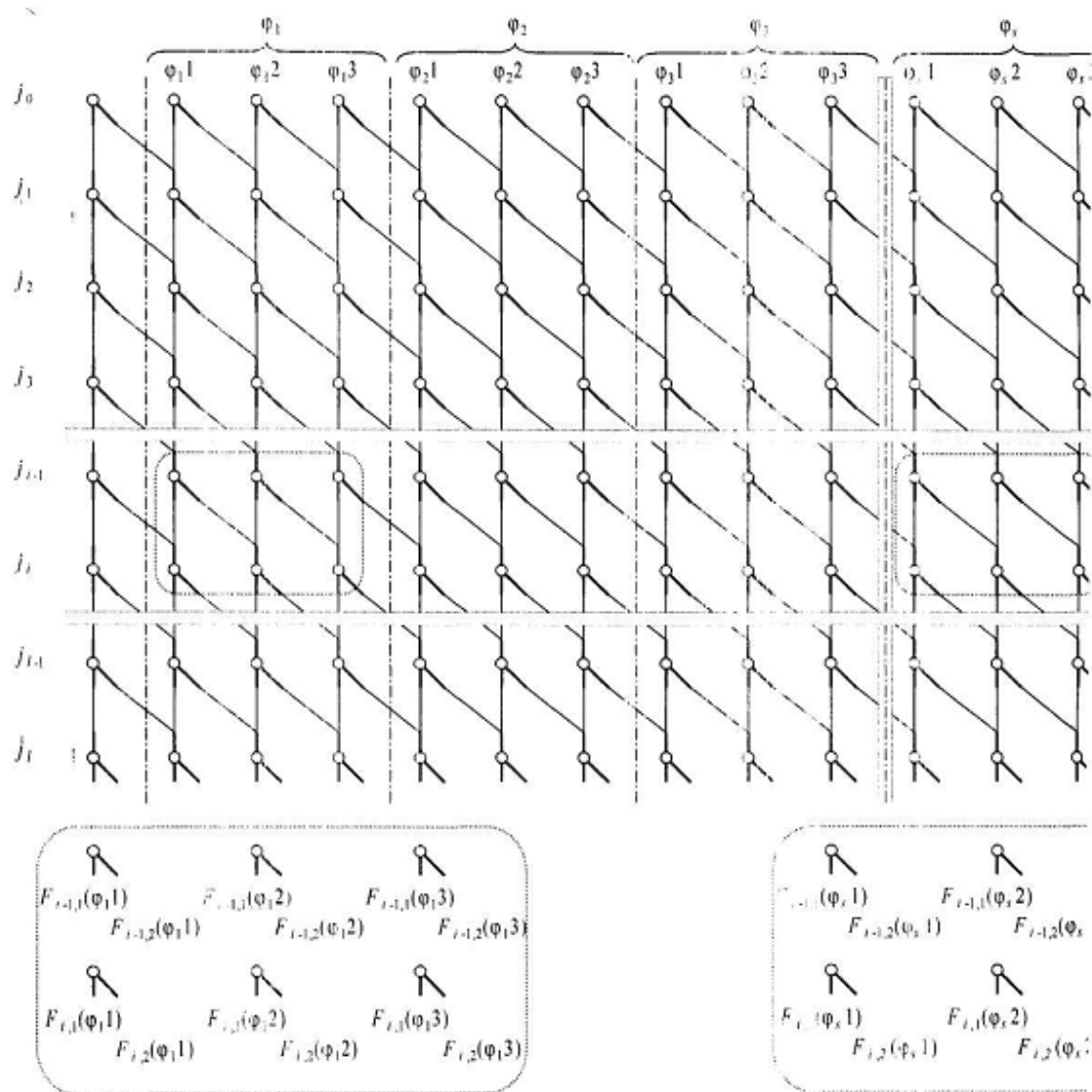


Fig. 7



Фиг. 8