



УКРАЇНА

(19) UA (11) 50041 (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 ЗЛИТОГО МОВЛЕННЯ

1

2

(21) u200911562

(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 злитого мовлення, що містить аналізатор, блок пам'яті табличних значень елементарних мір схожостей та блок пам'яті мовленнєвого сигналу, що розпізнається, обчислювач інтегральних мір схожостей 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 підключений 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 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 автоматичного розпізнавання. Може знайти використання для голосового управління пристроями.

Відомий пристрій для пофонемного розпізнавання (дивись патент України №48082) 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 також має недоліки, а саме, малу швидкість та низьку надійність розпізнавання.

В основу моделі покладена задача за рахунок введення конструктивних елементів у пристрій створити пристрій пофонемного розпізнавання усних команд та усталених словосполучень, що мають високу швидкість та надійність розпізнавання, забезпечує розпізнавання довільних наборів усних команд, словосполучень та злитого мовлення без необхідності перенастроювання системи розпізнавання на нові набори.

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, за вихідним сигналом якого, що вказує номер кластера, в який попадає поточний спостережуваний елемент, в блоці таблично визначаються значення елементарних мір належності цього спостереженого елемента до кожного із двох виходів із всіх трьох фаз фонем; обчислювач інтегральних мір схожостей містить сім вхідних регістрів, три компаратори та дев'ять суматорів і для кожного із двох виходів із всіх трьох

(13) U
(11) 50041
(19) UA

фаз кожної фонемі фонетичної транскрипції всіх усних команд та усталених словосполучень накопичує інтегральні міри схожості.

На Фіг.1 представлена структурна схема пристрою, Фіг.2-7 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; лінгвістичний блок 9 введення та зберігання орфографічних текстів та фонетичних транскрипцій всіх слів та допустимих у мові діалогу фраз; 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, поточні значення елементарним мір схожості для послідовності спостережуваних елементів-векторів для кожної із допустимих фраз 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 послідовність слів, для фонетичної транскрипції якої 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 $i\Delta T$ 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 $B_s = \sum_{n=0}^{M-1-s} f_n f_{n+s}$, $s=0:m$, які і утворюють поточний вектор автокореляції $B_i = (B_{i0}, B_{i1}, \dots, B_{is}, \dots, B_{im})$. Розв'язуючи систему рівнянь $\sum_{u=1}^m a_u B_{i[u-v]} = -B_v$, $v=1:m$, та обчислюючи потім $\sigma^2 = \frac{1}{M} \sum_{u=0}^{M-1} a_u B_u$, де $a_0 \equiv 1$, описуємо кожний поточний спостережуваний елемент B_i еквівалентним вектором-елементом (a_i, σ_i) - або (b_i, σ_i) - параметрів передбачення: $a=(a_1, a_2, \dots, a_s, \dots, a_n)$. $b=(b_0, b_1,$

$b_2, \dots, b_s, \dots, b_m)$; $b_0 = \sum_{u=0}^m a_u^2$, $b_s = 2 \sum_{u=0}^{m-s} a_u a_{u+s}$, $s=1:m$.

Обчислені в аналізаторі 1 елементи-вектори накопичуються в блоці 6 у вигляді послідовності як векторів автокореляцій B_i , так і векторів параметрів передбачення, утворюючи навчальну вибірку $X_{0i}=(x_1, x_2, \dots, x_i, \dots, x_l)$ із l спостережених елементів x_i .

В блоці 7 кластерного аналізу та обчислення параметрів фонем спершу за навчальною вибіркою X_{0i} з-посеред усіх l спостережених елементів x_i за допомогою деякої ітераційної процедури вибираємо задану кількість $J, J \ll l$, елементів (b_j^*, σ_j^*) , $j=1:J$, таких, які б найкращим чином апроксимували всі елементи навчальної вибірки:

$$(\sigma_j^*, b_j^*) = 1:J \left\{ \underset{(\sigma_j, b_j)}{\operatorname{argmin}} \min_{j=1:J} \sum_{i \in I_j} \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} \langle B_i, b_j \rangle \right) \right\}$$

, де через $\langle B, b \rangle$ 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

$$I_j = \left\{ i: j_i = \underset{j}{\operatorname{argmin}} \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} \langle B_i, b_j \rangle \right) = j \right\} \text{ позначено } j\text{-ий кластер. Всі елементи } j\text{-го кластера найкраще апроксимуються представником цього кластера}$$

через

$$(\sigma_j^*, b_j^*) = \underset{(\sigma_w, b_w)}{\operatorname{argmin}} \sum_{i \in I_j} \left(\frac{M}{2} \ln \sigma_w^2 + \frac{1}{2\sigma_w^2} \langle B_i, b_w \rangle \right).$$

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

$$I_j^0 = \left\{ i: j_i^0 = \underset{j}{\operatorname{argmin}} \left(\frac{M}{2} \ln (\sigma_j^0)^2 + \frac{1}{2(\sigma_j^0)^2} \langle B_i, b_j^0 \rangle \right) = j \right\},$$

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

$$(\sigma_j^{(g)}, b_j^{(g)}) = \underset{(\sigma_w, b_w)}{\operatorname{argmin}} \sum_{i \in I_j^{(g-1)}} \left(\frac{M}{2} \ln \sigma_w^2 + \frac{1}{2\sigma_w^2} \langle B_i, b_w \rangle \right),$$

$j=1:J$ представників кластерів. Потім на другому кроці g -тої ітерації, $g=1, 2, \dots$, знаходимо g -ту кластеризацію

$$I_j^{(g)} = \left\{ i: j_i^{(g)} = \underset{j}{\operatorname{argmin}} \left(\frac{M}{2} \ln (\sigma_j^{(g)})^2 + \frac{1}{2(\sigma_j^{(g)})^2} \langle B_i, b_j^{(g)} \rangle \right) = j \right\},$$

$j=1:J$. 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, σ_j) , $J=1:J$.

Вони далі використовуватимуться у векторно-матричному квантувачеві 10 при визначенні номера еталонного елемента j_i , що є найбільш схожим на спостережуваний елемент $B_i: j_i = \arg \min_j g(B_i, \sigma_j)$, де

$$g(B_i, \sigma_j) = \left(\frac{M}{2} \ln \left(\frac{1}{M} \sum_{k=1}^M b_{ik} \sigma_{jk} \right) + \frac{1}{2} \sum_{k=1}^M \frac{(b_{ik} - \sigma_{jk})^2}{\sigma_{jk}^2} \right) \text{ виступає як}$$

елементарна міра схожості елементів B_i та (b_i, σ_i) .

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 з 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, 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 спостережуваних елементів B в кожен із всіх кластерів $j = \arg \min_{u=1:J} g(B_u, \sigma_u)$ за умови перебу-

вання в першому, другому та третьому станах t кожної фонемі φ .

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

Представники кластерів або еталонні елементи-вектори (b_i, σ_i) , $j=1:J$ 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_{0i} 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 $\Psi_{0q}=(\varphi_1, \varphi_2, \dots, \varphi_s, \dots, \varphi_q)$, до φ_s - фонема з порядковим номером s у транскрипції, q - довжина транскрипції.

Еталонні мовленнєві образи усних команд та ustalених словосполучень формують шляхом об'єднання у послідовності ланцюгів породжувальних грамастик фонем з п'яти прихованих станів відповідно до фонетичних транскрипцій усних команд або словосполучень, причому так, щоб вихідний, четвертий, стан попередньої фонемі 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 фонемі φ подамо виразом

$$G(X_{\mu\nu}/\varphi) = \max_{w_1, w_2} \left\{ w_1 - \mu - 1 \ln p(\varphi_1/\varphi_2) \ln p(\varphi_2/\varphi_1) \sum_{i=\mu+1}^{w_1} \ln p(\varphi_i/\varphi_1) \right. \\ \left. + w_2 - w_1 - 1 \ln p(\varphi_2/\varphi_2) \ln p(\varphi_3/\varphi_2) \sum_{i=w_1+1}^{w_2} \ln p(\varphi_i/\varphi_2) \right. \\ \left. + w_3 - w_2 - 1 \ln p(\varphi_3/\varphi_3) \ln p(\varphi_4/\varphi_3) \sum_{i=w_2+1}^{\nu} \ln p(\varphi_i/\varphi_3) \right\},$$

де кожна складова в квадратних дужках виражає вклад кожної з трьох фаз фонемі. Сегмент $X_{\mu\nu}$ чи $J_{\mu\nu}$ відноситься до цієї фонемі, 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, словосполучення чи фрази з фоне-

тичною 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:

$$G(X_{\mu\nu}/\varphi_1, \varphi_2, \dots, \varphi_s, \dots, \varphi_q) = \sum_{s=1}^q G(X_{\mu_{s-1}\mu_s}/\varphi_s, w_{s1}, w_{s2}) = 1:q \sum_{s=1}^q G(X_{\mu_{s-1}\mu_s}/\varphi_s, w_{s1}, w_{s2}) \\ = \sum_{s=1}^q \left\{ w_{s1} - \mu_{s-1} - 1 \ln p(\varphi_{s1}/\varphi_{s1}) \ln p(\varphi_{s2}/\varphi_{s1}) \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(\varphi_i/\varphi_{s1}) \right. \\ \left. + w_{s2} - w_{s1} - 1 \ln p(\varphi_{s2}/\varphi_{s2}) \ln p(\varphi_{s3}/\varphi_{s2}) \sum_{i=w_{s1}+1}^{w_{s2}} \ln p(\varphi_i/\varphi_{s2}) \right. \\ \left. + w_{s3} - w_{s2} - 1 \ln p(\varphi_{s3}/\varphi_{s3}) \ln p(\varphi_{s4}/\varphi_{s3}) \sum_{i=w_{s2}+1}^{\mu_s} \ln p(\varphi_i/\varphi_{s3}) \right\}$$

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

позначено можливий s -тий підсегмент для фонемі φ_s з трьома його відповідними підпідсегментами, причому $\mu_0=\mu$, $\mu_{i-1}<\mu_i$, $\mu_q=\nu$. Оптимальне роз-

биття чи сегментація $(\varphi_s^*, w_{s1}^*, w_{s2}^*) = 1:q$, яка максимізує критерій правдоподібності, визначає

інтегральну міру схожості сегмента $X_{\mu\nu}$ чи $J_{\mu\nu}$ на слово, словосполучення чи фразу з фонетичною

транскрипцією Ψ_{0q} :

$$G_{\mu\nu} / \Psi_{0q} = \max_{\mu_s, w_{s1}, w_{s2}, s=1:q} G_{\mu\nu} / \langle \phi_1, \phi_2, \dots, \phi_s, \dots, \phi_q \rangle; \langle \mu_s^*, w_{s1}^*, w_{s2}^* \rangle, s=1:q$$

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

Вся ж навчальна вибірка $X_0 = (x_1, x_2, \dots, x_i, \dots, x_J)$ чи $J_0 = (j_1, j_2, \dots, j_i, \dots, j_J)$, 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. $\Psi_{0Q} = (\phi_1, \phi_2, \dots, \phi_s, \dots, \phi_Q)$ - фонетична транскрипція всієї навчальної вибірки з Q реалізацій фонем.

Максимально правдоподібні оцінки ймовірностей $P = \{p(\phi_t/\phi_t), p(j/\phi_t), j=1:J, t=1:3, \phi \in \Psi\}$, де Ψ - алфавіт фонем, обчислюються процесором 7 шляхом максимізації критерію правдоподібності для навчальної вибірки

$$G_{0Q} / \Psi_{0Q}; P; \langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:q = \sum_{s=1}^Q \left[\ln p(\mu_{s1} - \mu_{s-1} - 1) \ln p(\phi_{s1} / \phi_{s1}) \ln p(\phi_{s2} / \phi_{s1}) \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(\phi_i / \phi_{s1}) \right] + \left[\ln p(w_{s2} - w_{s1} - 1) \ln p(\phi_{s2} / \phi_{s2}) \ln p(\phi_{s3} / \phi_{s2}) \sum_{i=w_{s1}+1}^{w_{s2}} \ln p(\phi_i / \phi_{s2}) \right] + \left[\ln p(1 - w_{s2-1} - 1) \ln p(\phi_{s3} / \phi_{s3}) \ln p(\phi_{s4} / \phi_{s3}) \sum_{i=w_{s2}+1}^{\mu_s} \ln p(\phi_i / \phi_{s3}) \right]$$

Критерій правдоподібності можна переписати в децю іншій еквівалентній формі, згрупувавши окремо всі реалізації однієї й тієї ж фонем:

$$G_{0Q} / \Psi_{0Q}; P; \langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:q = \sum_{\phi \in \Psi} \sum_{s: \phi_s = \phi} \left[\ln p(\mu_{s1} - \mu_{s-1} - 1) \ln p(\phi_1 / \phi_1) \ln p(\phi_2 / \phi_1) \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(\phi_i / \phi_1) \right] + \left[\ln p(w_{s2} - w_{s1} - 1) \ln p(\phi_2 / \phi_2) \ln p(\phi_3 / \phi_2) \sum_{i=w_{s1}+1}^{w_{s2}} \ln p(\phi_i / \phi_2) \right] + \left[\ln p(1 - w_{s2-1} - 1) \ln p(\phi_3 / \phi_3) \ln p(\phi_4 / \phi_3) \sum_{i=w_{s2}+1}^{\mu_s} \ln p(\phi_i / \phi_3) \right]$$

З останнього запису випливає, що як до 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:

$$\begin{aligned} p(\phi_1 / \phi_1) &= \sum_{s: \phi_s = \phi} \langle w_{s1} - \mu_{s-1} - 1 \rangle \sum_{s: \phi_s = \phi} \langle w_{s1} - \mu_{s-1} \rangle, \\ p(\phi_2 / \phi_1) &= \sum_{s: \phi_s = \phi} 1 / \sum_{s: \phi_s = \phi} \langle w_{s1} - \mu_{s-1} \rangle, \\ p(\phi_2 / \phi_2) &= \sum_{s: \phi_s = \phi} \langle w_{s2} - w_{s1} - 1 \rangle \sum_{s: \phi_s = \phi} \langle w_{s2} - w_{s1} \rangle, \\ p(\phi_3 / \phi_2) &= \sum_{s: \phi_s = \phi} 1 / \sum_{s: \phi_s = \phi} \langle w_{s2} - w_{s1} \rangle, \\ p(\phi_3 / \phi_3) &= \sum_{s: \phi_s = \phi} \langle \mu_s - w_{s2} - 1 \rangle \sum_{s: \phi_s = \phi} \langle \mu_s - w_{s2} \rangle, \\ p(\phi_4 / \phi_3) &= \sum_{s: \phi_s = \phi} 1 / \sum_{s: \phi_s = \phi} \langle \mu_s - w_{s2} \rangle, \\ p(\phi / \phi_1) &= \sum_{s: \phi_s = \phi} \sum_{i=\mu_{s-1}+1}^{w_{s1}} 1 / \sum_{s: \phi_s = \phi} \langle w_{s2} - \mu_{s-1} \rangle, \\ p(\phi / \phi_2) &= \sum_{s: \phi_s = \phi} \sum_{i=w_{s1}+1}^{w_{s2}} 1 / \sum_{s: \phi_s = \phi} \langle w_{s2} - w_{s1} \rangle, \end{aligned}$$

$$p(\phi / \phi_3) = \sum_{s: \phi_s = \phi} \sum_{i=w_{s2}+1}^{\mu_s} 1 / \sum_{s: \phi_s = \phi} \langle \mu_s - w_{s2} \rangle, j=1:J, \phi \in \Psi.$$

Ітераційний процес обчислення параметрів P починаємо з того, що на першому кроці нульової ітерації знаходимо початкову сегментацію всієї навчальної вибірки $\langle \mu_s^*, w_{s1}^*, w_{s2}^* \rangle, s=1:Q$. Для цього виокремимо із навчальної вибірки сегменти, що відповідають окремо вимовленим словам, словосполученням чи фразам. Далі для кожного таким чином виділеного сегмента $X_0 = (x_1, x_2, \dots, x_i, \dots, x_J)$, відповідно до його фонетичної транскрипції $\Psi_{0q} = (\phi_1, \phi_2, \dots, \phi_s, \dots, \phi_q)$, виконаємо його оптимальне розбиття на q сегментів, виходячи із моделі однорідності сегментів та апелюючи до вже знайдених еталонних елементів - представників кластерів

$$\begin{aligned} & \left\{ \min_j \sum_{i=1}^{\mu_1} \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} \langle b_j \rangle \right) \right\} + \\ & + \sum_{s=2}^{q-1} \left\{ \left[\min_j \sum_{i=\mu_{s-1}+1}^{w_{s1}} \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} \langle b_j \rangle \right) \right] + \left[\min_j \sum_{i=w_{s1}+1}^{w_{s2}} \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} \langle b_j \rangle \right) \right] \right\} + \\ & + \left\{ \min_j \sum_{i=w_{s2}+1}^{\mu_s} \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} \langle b_j \rangle \right) \right\} + \min_j \sum_{i=\mu_{q-1}+1}^{w_q} \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} \langle b_j \rangle \right) \end{aligned}$$

Повторивши процедуру розбиття для всіх реалізацій слів та фраз, отримаємо початкову сегментацію $\langle w_{s1}, w_{s2} \rangle, s=1:Q$ всієї навчальної вибірки.

Потім на другому кроці нульової ітерації за відомою сегментацією $\langle w_{s1}, w_{s2} \rangle, s=1:Q$ 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

$$\begin{aligned} & \left\{ \langle w_{s1}, w_{s2} \rangle, s=1:Q \right\} \arg \max_{\langle w_{s1}, w_{s2} \rangle, s=1:Q} \sum_{\varphi \in \Psi} \sum_{s: \varphi_s = \varphi} \left[w_{s1} - \mu_{s-1} - 1 \right] \ln p^{(-1)} \langle 1/\varphi_1 \rangle + \\ & + \ln p^{(-1)} \langle 2/\varphi_1 \rangle \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p^{(-1)} \langle 1/\varphi_2 \rangle + \\ & + \left[w_{s2} - w_{s1} - 1 \right] \ln p^{(-1)} \langle 2/\varphi_2 \rangle \ln p^{(-1)} \langle 3/\varphi_2 \rangle \sum_{i=w_{s1}+1}^{w_{s2}} \ln p^{(-1)} \langle 1/\varphi_2 \rangle + \\ & + \left[\mu_s - w_{s2} - 1 \right] \ln p^{(-1)} \langle 3/\varphi_3 \rangle \ln p^{(-1)} \langle 4/\varphi_3 \rangle \sum_{i=w_{s2}+1}^{\mu_s} \ln p^{(-1)} \langle 1/\varphi_3 \rangle \end{aligned}$$

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

$P^{(r)} = \{p^{(r)}(\varphi_t/\varphi_t), p^{(r)}(j/\varphi_t), j=1:J, t=1:3, \varphi \in \Psi\}$:

$$\begin{aligned} p^{(r)} \langle 1/\varphi_1 \rangle &= \sum_{s: \varphi_s = \varphi} \left(\frac{1}{\mu_{s-1}} \right) \sum_{s: \varphi_s = \varphi} \left(\frac{1}{\mu_{s-1}} \right), \\ p^{(r)} \langle 2/\varphi_2 \rangle &= \sum_{s: \varphi_s = \varphi} \left(\frac{1}{w_{s1} - \mu_{s-1}} \right) \sum_{s: \varphi_s = \varphi} \left(\frac{1}{w_{s1} - \mu_{s-1}} \right), \\ p^{(r)} \langle 3/\varphi_3 \rangle &= \sum_{s: \varphi_s = \varphi} \left(\frac{1}{w_{s2} - w_{s1}} \right) \sum_{s: \varphi_s = \varphi} \left(\frac{1}{w_{s2} - w_{s1}} \right), \\ p^{(r)} \langle 4/\varphi_4 \rangle &= \left(\sum_{s: \varphi_s = \varphi} \sum_{i=\mu_{s-1}+1}^{w_{s1}} 1 \right) / \sum_{s: \varphi_s = \varphi} \left(\frac{1}{\mu_{s-1}} \right), \\ p^{(r)} \langle 1/\varphi_2 \rangle &= \left(\sum_{s: \varphi_s = \varphi} \sum_{i=w_{s1}+1}^{w_{s2}} 1 \right) / \sum_{s: \varphi_s = \varphi} \left(\frac{1}{w_{s1} - \mu_{s-1}} \right), \\ p^{(r)} \langle 1/\varphi_3 \rangle &= \left(\sum_{s: \varphi_s = \varphi} \sum_{i=w_{s2}+1}^{\mu_s} 1 \right) / \sum_{s: \varphi_s = \varphi} \left(\frac{1}{w_{s2} - w_{s1}} \right), \end{aligned}$$

$j=1:J, \varphi \in \Psi$.

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 - робочім словником усних команд, чи стандартних фраз з їхніми фонетичними транскрипціями, пристрій стає готовим до автоматичного розпізнавання: на кожну усну фразу з номером k в блоці пам'яті 4а проміжних результатів робляться початкові установки

$$\begin{aligned} F_{0,2}^k \langle 0,3 \rangle &= 0; \quad (F_{0,1}^k \langle 0,1 \rangle = -M, \quad F_{0,2}^k \langle 0,1 \rangle = -M, \\ F_{0,1}^k \langle 0,2 \rangle &= -M, \quad F_{0,2}^k \langle 0,2 \rangle = -M, \quad F_{0,1}^k \langle 0,3 \rangle = -M, \end{aligned}$$

$F_{0,2}^k \langle 0,3 \rangle = -M)$, $s=1:q^k$, $k=1:K$, де $(-M)$ - велике від'ємне число, q^k - довжина транскрипції k -го слова, K - обсяг словника або фразника.

Контролер 5, як тільки отримує інформацію від аналізатора 1 про появу чергового спостереженого елемента B_1 , забезпечує пересилку його на векторний квантувач 10, де визначається номер еталонного елемента $j_1 = \arg \min_{j=1:J} \left(\frac{M}{2} \ln \sigma_j^2 + \langle b_j \rangle \right)$, що

є найбільш схожим на спостережуваний. Під управлінням контролера 5 з блока 9 зчитується ім'я першої фонемі φ_1 першої фрази. 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 $F_{0,2}^k \langle 0,3 \rangle$, $F_{0,1}^k \langle 0,1 \rangle$,

$$F_{0,2}^k \langle 0,1 \rangle, \quad F_{0,1}^k \langle 0,2 \rangle, \quad F_{0,2}^k \langle 0,2 \rangle, \quad F_{0,1}^k \langle 0,3 \rangle, \quad F_{0,2}^k \langle 0,3 \rangle.$$

Одночасно за номером еталонного елемента j_1 та ім'ям фонемі φ_1 в процесор 4 надходять логарифми ймовірностей $\ln p(j_1/\varphi_1)$, $\ln p(\varphi_1/\varphi_1)$, $\ln p(\varphi_2/\varphi_1)$, $\ln p(j_1/\varphi_2)$, $\ln p(\varphi_2/\varphi_2)$, $\ln p(\varphi_3/\varphi_2)$, $\ln p(j_1/\varphi_3)$, $\ln p(\varphi_3/\varphi_3)$, $\ln p(\varphi_4/\varphi_3)$, відповідно, та обчислюються нові значення інтегральних мір схожості на фразу з номером k :

$$\begin{aligned} F_{1,1}^k \langle 0,1 \rangle &= \max \{ F_{0,2}^k \langle 0,3 \rangle, F_{0,1}^k \langle 0,1 \rangle \} + \ln p \langle 1/\varphi_1 \rangle - \ln p \langle 0,1/\varphi_1 \rangle, \\ F_{1,2}^k \langle 0,1 \rangle &= \max \{ F_{0,2}^k \langle 0,3 \rangle, F_{0,1}^k \langle 0,1 \rangle \} + \ln p \langle 1/\varphi_1 \rangle - \ln p \langle 0,2/\varphi_1 \rangle, \\ F_{1,1}^k \langle 0,2 \rangle &= \max \{ F_{0,2}^k \langle 0,1 \rangle, F_{0,1}^k \langle 0,2 \rangle \} + \ln p \langle 1/\varphi_2 \rangle - \ln p \langle 0,2/\varphi_2 \rangle, \end{aligned}$$

$$F_{1,2}^k \phi_{1,2} = \max_{0,2}^k \phi_{1,1} \cdot F_{0,1}^k \phi_{1,2} \cdot \ln p(\phi_{1,2} / \phi_{1,1}) \cdot \ln p(\phi_{1,3} / \phi_{1,2}),$$

$$F_{1,1}^k \phi_{1,3} = \max_{0,2}^k \phi_{1,2} \cdot F_{0,1}^k \phi_{1,3} \cdot \ln p(\phi_{1,3} / \phi_{1,2}) \cdot \ln p(\phi_{1,4} / \phi_{1,3}),$$

$$F_{1,2}^k \phi_{1,3} = \max_{0,2}^k \phi_{1,2} \cdot F_{0,1}^k \phi_{1,3} \cdot \ln p(\phi_{1,3} / \phi_{1,2}) \cdot \ln p(\phi_{1,4} / \phi_{1,3}).$$

Далі за тим же номером еталонного елемента j_1 послідовно для всіх решти фонем $s=2:q^k$ фрази л діємо аналогічно. За іменем фонем ϕ_s з блока пам'яті 4а накопичуваних інтегральних мір схожості на обчислювач 4 інтегральних мір схожості зчитуються значення - $F_{0,1}^k \phi_{s-1,1}$, $F_{0,2}^k \phi_{s-1,1}$,

$F_{0,1}^k \phi_{s-1,2}$, $F_{0,2}^k \phi_{s-1,2}$, $F_{0,1}^k \phi_{s-1,3}$, $F_{0,2}^k \phi_{s-1,3}$. Од-
ночасно за номером еталонного елемента j_1 та ім'ям фонем ϕ_s в процесор 4 надходять логарифми ймовірностей $\ln p(j_1 / \phi_s 1)$, $\ln p(\phi_s 1 / \phi_s 1)$, $\ln p(\phi_s 2 / \phi_s 1)$, $\ln p(j_1 / \phi_s 2)$, $\ln p(\phi_s 2 / \phi_s 2)$, $\ln p(\phi_s 3 / \phi_s 2)$, $\ln p(j_1 / \phi_s 3)$, $\ln p(\phi_s 3 / \phi_s 3)$, $\ln p(\phi_s 4 / \phi_s 3)$, відповідно, та обчислюються нові значення інтегральних мір схожості на фразу з номером k :

$$F_{1,1}^k \phi_{s,1} = \max_{0,2}^k \phi_{s-2,3} \cdot F_{0,1}^k \phi_{s-1,1} \cdot \ln p(\phi_{s,1} / \phi_{s-1,1}) \cdot \ln p(\phi_{s,1} / \phi_{s,1}),$$

$$F_{1,2}^k \phi_{s,1} = \max_{0,2}^k \phi_{s-2,3} \cdot F_{0,1}^k \phi_{s-1,1} \cdot \ln p(\phi_{s,1} / \phi_{s-1,1}) \cdot \ln p(\phi_{s,2} / \phi_{s,1}),$$

$$F_{1,1}^k \phi_{s,2} = \max_{0,2}^k \phi_{s-1,1} \cdot F_{0,1}^k \phi_{s-1,2} \cdot \ln p(\phi_{s,2} / \phi_{s-1,2}) \cdot \ln p(\phi_{s,2} / \phi_{s,2}),$$

$$F_{1,2}^k \phi_{s,2} = \max_{0,2}^k \phi_{s-1,1} \cdot F_{0,1}^k \phi_{s-1,2} \cdot \ln p(\phi_{s,2} / \phi_{s-1,2}) \cdot \ln p(\phi_{s,3} / \phi_{s,2}),$$

$$F_{1,1}^k \phi_{s,3} = \max_{0,2}^k \phi_{s-1,2} \cdot F_{0,1}^k \phi_{s-1,3} \cdot \ln p(\phi_{s,3} / \phi_{s-1,3}) \cdot \ln p(\phi_{s,3} / \phi_{s,3}),$$

$$F_{1,2}^k \phi_{s,3} = \max_{0,2}^k \phi_{s-1,2} \cdot F_{0,1}^k \phi_{s-1,3} \cdot \ln p(\phi_{s,3} / \phi_{s-1,3}) \cdot \ln p(\phi_{s,4} / \phi_{s,3}).$$

Далі значення регістра 44а пересилаємо в регістр 41, затим обчислені вище значення інтегральних мір схожості $F_{1,1}^k \phi_{s,1}$, $F_{1,2}^k \phi_{s,1}$, $F_{1,1}^k \phi_{s,2}$, $F_{1,2}^k \phi_{s,2}$, $F_{1,1}^k \phi_{s,3}$, $F_{1,2}^k \phi_{s,3}$ пересилаємо на зберігання в блок пам'яті 4а за адресами значень $F_{0,1}^k \phi_{s-1,1}$, $F_{0,2}^k \phi_{s-1,1}$, $F_{0,1}^k \phi_{s-1,2}$, $F_{0,2}^k \phi_{s-1,2}$, $F_{0,1}^k \phi_{s-1,3}$, $F_{0,2}^k \phi_{s-1,3}$ відповідно.

Після закінчення оброблення фрази k за адресою $F_{0,2}^k \phi_{0,3}$ пересилаємо (-M), а всі обчислені нові значення інтегральних мір схожості $F_{1,1}^k \phi_{st}$, $F_{1,2}^k \phi_{st}$, $t=1:3$ переписуємо на місце старих $F_{0,1}^k \phi_{st}$, $F_{0,2}^k \phi_{st}$, $t=1:3$.

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

Далі процес повторюється в циклі по $i=2:I$.

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 слів, усталених словосполучень чи фраз інтегральні міри схожості $F_{1,1}^k \phi_{s,1}$, $F_{1,2}^k \phi_{s,1}$, $F_{1,1}^k \phi_{s,2}$, $F_{1,2}^k \phi_{s,2}$, $F_{1,1}^k \phi_{s,3}$, $F_{1,2}^k \phi_{s,3}$, що визначаються для кожного із двох виводів із першого, другого чи третього станів поточної фонемі їх транскрипцій, знаходять як суми значень відповідних елементарних мір схожості, обчислених для поточного спостереженого елемента y , для кожного із двох виходів із першого, другого чи третього станів поточної фонемі: $\ln p(j_i / \phi_s 1)$, $\ln p(\phi_s 1 / \phi_s 1)$, $\ln p(\phi_s 2 / \phi_s 1)$, $\ln p(j_i / \phi_s 2)$, $\ln p(\phi_s 2 / \phi_s 2)$, $\ln p(\phi_s 3 / \phi_s 2)$, $\ln p(j_i / \phi_s 3)$, $\ln p(\phi_s 3 / \phi_s 3)$, $\ln p(\phi_s 4 / \phi_s 3)$, з більшою із двох інтегральних мір схожості, нако-

пичених для попереднього спостереженого елемента j_{i-1} на другому виході з третього стану попередньої фонемі $\phi_{s-1,3}$ та на першому виході з першого стану поточної фонемі $\phi_s 1$, на другому виході з першого стану поточної фонемі $\phi_s 1$ та на першому виході з другого стану поточної фонемі $\phi_s 2$ та на другому виході з другого стану поточної фонемі $\phi_s 2$ та на першому виході з третього стану поточної фонемі $\phi_s 3$, відповідно:

$$F_{1,1}^k \phi_{s,1} = \max_{1,2}^k \phi_{s-2,3} \cdot F_{1,1}^k \phi_{s-1,1} \cdot \ln p(\phi_{s,1} / \phi_{s-1,1}) \cdot \ln p(\phi_{s,1} / \phi_{s,1}),$$

$$F_{1,2}^k \phi_{s,1} = \max_{1,2}^k \phi_{s-2,3} \cdot F_{1,1}^k \phi_{s-1,1} \cdot \ln p(\phi_{s,1} / \phi_{s-1,1}) \cdot \ln p(\phi_{s,2} / \phi_{s,1}),$$

$$F_{1,1}^k \phi_{s,2} = \max_{1,2}^k \phi_{s-1,1} \cdot F_{1,1}^k \phi_{s-1,2} \cdot \ln p(\phi_{s,2} / \phi_{s-1,2}) \cdot \ln p(\phi_{s,2} / \phi_{s,2}),$$

$$F_{1,2}^k \phi_{s,2} = \max_{1,2}^k \phi_{s-1,1} \cdot F_{1,1}^k \phi_{s-1,2} \cdot \ln p(\phi_{s,2} / \phi_{s-1,2}) \cdot \ln p(\phi_{s,3} / \phi_{s,2}),$$

$$F_{1,1}^k \phi_{s,3} = \max_{1,2}^k \phi_{s-1,2} \cdot F_{1,1}^k \phi_{s-1,3} \cdot \ln p(\phi_{s,3} / \phi_{s-1,3}) \cdot \ln p(\phi_{s,3} / \phi_{s,3}),$$

$$F_{1,2}^k \phi_{s,3} = \max_{1,2}^k \phi_{s-1,2} \cdot F_{1,1}^k \phi_{s-1,3} \cdot \ln p(\phi_{s,3} / \phi_{s-1,3}) \cdot \ln p(\phi_{s,4} / \phi_{s,3}).$$

Значення інтегральної міри схожості $F_{1,2}^k \phi_{q,3}$,

накопичене після оброблення останнього спостереженого елемента на другому виході із третього стану останньої фонемі q^k , яка визначається фонетичною транскрипцією слова, словосполучення або фрази, визначає схожість пред'явленого мовленнєвого сигналу на це слово, словосполучення або фразу.

Пред'явлений мовленнєвий сигнал контролером 5 відноситься до тієї фрази із вибраного фразника, для котрої накопичена схожість є абсолютно найбільшою.

У випадку, коли задано робочий словник слів-сформ, а фрази укладаються зі слів цього словника за певними правилами, відповідь розпізнавання усних фраз потрібно шукати тільки серед допустимих в мові діалогу фраз.

Якщо у словнику всього K слів, то у випадку вільного порядку слідування слів кількість допустимих фраз, що складаються не більш як з 10 слів,

буде дорівнювати $\sum_{n=1}^{10} K^n$.

Граф породження всіх можливих еталонних мовленнєвих сигналів для цього випадку подано на рис. 6. Граф нагадує квітку, у центрі якої маємо головний стан $s=0$. Графи окремих слів $k=1:K$ моделюються пелюстками - ланцюгами послідовно з'єднаних станів, по три стани на кожну фонему, відповідно до фонетичної транскрипції слова. Причому граф кожного слова має початкову та кінцеву фонему-паузи.

Стартуючи зі стану $s_{поч.}$, за 0 тактів переходимо в головний стан $s=0$, а звідти, також за 0 тактів, переходимо на генерацію еталонних мовленнєвих сигналів того чи того слова. Через другий вихід останньої фонемі транскрипції слова за 0 тактів виходимо знову в стан $s=0$ та знову маємо можливість вільного вибору породження наступного слова.

Щоб відповісти на питання, яка послідовність слів передається поданою для аналізу послідовністю спостережених елементів $X_{0i}=(x_1, x_2, \dots, x_i, \dots, x_i)$ чи $J_{0i}=(j_1, j_2, \dots, j_i, \dots, j_i)$, треба знайти для цієї послідовності найбільш правдоподібну траєкторію на графі злитого мовлення та вказати, через які пелюстки та в якій послідовності пелюсток вона проходить.

З цієї метою спочатку зробимо такі початкові установки для моменту часу $i=0$: для головного стану $s=0$ покладемо $F_0(0)=0$, $v_0(0)=0$, а для всіх станів всіх слів відповідно $F_{0,2}^k \langle \phi_0 3 \rangle = F_0 \langle \rangle$, $v_{0,2}^k \langle \phi_0 3 \rangle = v_0 \langle \rangle$, $(F_{0,1}^k \langle \phi_s 1 \rangle = -M$, $F_{0,2}^k \langle \phi_s 1 \rangle = -M$, $F_{0,1}^k \langle \phi_s 2 \rangle = -M$, $F_{0,2}^k \langle \phi_s 2 \rangle = -M$, $F_{0,1}^k \langle \phi_s 3 \rangle = -M$, $F_{0,2}^k \langle \phi_s 3 \rangle = -M)$, $s=1:q^k$, $k=1:K$.

Далі в циклі по $i=1:l$, з появою чергового спостережуваного елемента x_i або j_i , тобто 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 всіх фраз, що можуть закінчуватись словом k , $s=1:q^k$, $k=1:K$, інтегральні міри схожості $F_{i,1}^k \langle \phi_s 1 \rangle$, $F_{i,2}^k \langle \phi_s 1 \rangle$, $F_{i,1}^k \langle \phi_s 2 \rangle$, $F_{i,2}^k \langle \phi_s 2 \rangle$, $F_{i,1}^k \langle \phi_s 3 \rangle$, $F_{i,2}^k \langle \phi_s 3 \rangle$, що визначаються для кожного із двох виходів із першого, другого чи третього станів поточної 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: $\ln p(j_i/\phi_s 1)$, $\ln p(\phi_s 1/\phi_s 1)$, $\ln p(\phi_s 2/\phi_s 1)$, $\ln p(j_i/\phi_s 2)$, $\ln p(\phi_s 2/\phi_s 2)$, $\ln p(\phi_s 3/\phi_s 2)$, $\ln p(j_i/\phi_s 3)$, $\ln p(\phi_s 3/\phi_s 3)$, $\ln p(\phi_s 4/\phi_s 3)$, з більшою із двох інтегральних мір схожості, накопичених для попереднього спостереженого елемента j_{i-1} на другому виході з третього стану попередньої фонемі та на першому виході з першого стану поточної фонемі на другому виході з першого стану поточної фонемі та на першому виході з другого стану поточної фонемі й на другому виході з другого стану поточної фонемі та на першому виході з третього стану поточної фонемі, відповідно:

$$F_{i,1}^k \langle \phi_s 1 \rangle = \max_{k=1:K} \{ F_{i-1,2}^k \langle \phi_{s-2} 3 \rangle + F_{i-1,1}^k \langle \phi_{s-1} 1 \rangle + \ln p(\phi_s 1/\phi_s 1) + \ln p(\phi_s 1/\phi_s 1) \},$$

$$F_{i,2}^k \langle \phi_s 1 \rangle = \max_{k=1:K} \{ F_{i-1,2}^k \langle \phi_{s-2} 3 \rangle + F_{i-1,1}^k \langle \phi_{s-1} 1 \rangle + \ln p(\phi_s 1/\phi_s 1) + \ln p(\phi_s 2/\phi_s 1) \},$$

$$F_{i,1}^k \langle \phi_s 2 \rangle = \max_{k=1:K} \{ F_{i-1,2}^k \langle \phi_{s-1} 1 \rangle + F_{i-1,1}^k \langle \phi_{s-2} 2 \rangle + \ln p(\phi_s 2/\phi_s 2) + \ln p(\phi_s 2/\phi_s 2) \},$$

$$F_{i,2}^k \langle \phi_s 2 \rangle = \max_{k=1:K} \{ F_{i-1,2}^k \langle \phi_{s-1} 1 \rangle + F_{i-1,1}^k \langle \phi_{s-2} 2 \rangle + \ln p(\phi_s 2/\phi_s 2) + \ln p(\phi_s 3/\phi_s 2) \},$$

$$F_{i,1}^k \langle \phi_s 3 \rangle = \max_{k=1:K} \{ F_{i-1,2}^k \langle \phi_{s-2} 2 \rangle + F_{i-1,1}^k \langle \phi_{s-1} 3 \rangle + \ln p(\phi_s 3/\phi_s 3) + \ln p(\phi_s 3/\phi_s 3) \},$$

$$F_{i,2}^k \langle \phi_s 3 \rangle = \max_{k=1:K} \{ F_{i-1,2}^k \langle \phi_{s-2} 2 \rangle + F_{i-1,1}^k \langle \phi_{s-1} 3 \rangle + \ln p(\phi_s 3/\phi_s 3) + \ln p(\phi_s 4/\phi_s 3) \},$$

причому кожна із величин $F_{i,1}^k \langle \phi_s 1 \rangle$, $F_{i,2}^k \langle \phi_s 1 \rangle$, $F_{i,1}^k \langle \phi_s 2 \rangle$, $F_{i,2}^k \langle \phi_s 2 \rangle$, $F_{i,1}^k \langle \phi_s 3 \rangle$, $F_{i,2}^k \langle \phi_s 3 \rangle$, супроводжується потенційно-оптимальним початком k -го слова $v_{i,1}^k \langle \phi_s 1 \rangle$, $v_{i,2}^k \langle \phi_s 1 \rangle$, $v_{i,1}^k \langle \phi_s 2 \rangle$, $v_{i,2}^k \langle \phi_s 2 \rangle$, $v_{i,1}^k \langle \phi_s 3 \rangle$, $v_{i,2}^k \langle \phi_s 3 \rangle$, що дорівнюють потенційно-оптимальним початкам k -го слова $v_{i-1,2}^k \langle \phi_{s-2} 3 \rangle$ чи $v_{i-1,1}^k \langle \phi_{s-1} 1 \rangle$, $v_{i-1,2}^k \langle \phi_{s-1} 1 \rangle$ чи $v_{i-1,1}^k \langle \phi_{s-2} 2 \rangle$, $v_{i-1,2}^k \langle \phi_{s-2} 2 \rangle$ чи $v_{i-1,1}^k \langle \phi_{s-1} 3 \rangle$, відповідно, в залежності від того, який член, перший чи другий є більшим у виразах для $F_{i,1}^k \langle \phi_s 1 \rangle$, $F_{i,2}^k \langle \phi_s 1 \rangle$, $F_{i,1}^k \langle \phi_s 2 \rangle$, $F_{i,2}^k \langle \phi_s 2 \rangle$, $F_{i,1}^k \langle \phi_s 3 \rangle$, $F_{i,2}^k \langle \phi_s 3 \rangle$.

Оброблення i -го елемента x_i або j_i завершуємо обчисленням

$$F_i \langle \rangle = \max_{k=1:K} F_{i,2}^k \langle \phi_q^k 3 \rangle,$$

$$k_i \langle \rangle = \operatorname{argmax}_{k=1:K} F_{i,2}^k \langle \phi_q^k 3 \rangle, \quad v_i \langle \rangle = v_{i,2}^k \langle \phi_q^k 3 \rangle, \quad \text{тобто}$$

виявляємо, з якою найбільшою інтегральною мірою схожості, яке слово та коли воно могло початись, закінчуючись в момент i . Щоб підготуватись до прийняття та оброблення чергового спостережуваного елемента x_{i+1} або j_{i+1} , робимо пересилання $F_{i,2}^k \langle \phi_0 3 \rangle = F_i \langle \rangle$, $v_{0,2}^k \langle \phi_0 3 \rangle = v_i \langle \rangle$, $k=1:K$ та переходимо до наступного циклу рекурентних обчислень.

Після їх завершення формуємо відповідь розпізнавання за масивом $(F_i(0), v_i(0), k_i(0))$, $i=1:l$.

Значення $F_i(0)$ визначає найбільшу інтегральну міру схожості, а відповідь розпізнавання у вигляді послідовності спів, що передана мовленнєвим сигналом $X_{0i}=(x_1, x_2, \dots, x_i)$ або $J_{0i}=(j_1, j_2, \dots, j_i)$, формуємо з таким алгоритмом виписування:

$k_1^* = k_1 \langle \rangle$ буде першим з кінця словом, що почалося в момент $v_1^* = v_1 \langle \rangle$, $k_2^* = k_{v_1} \langle \rangle$ буде другим з кінця словом, що почалося в момент $v_2^* = v_{v_1} \langle \rangle$, $k_3^* = k_{v_2} \langle \rangle$ буде третім з кінця словом, що почалося в момент $v_3^* = v_{v_2} \langle \rangle$. І так далі. Виписування завершуємо при досягненні $v^*=0$.

У випадку, коли задано робочий словник слів, відомий підсловник $V_{\text{поч.}}$, словами якого розпочинаються допустимі фрази, підсловники V_k , що містять слова, які можуть слідувати за словом k , та підсловник $V_{\text{кінц.}}$, словами котрого закінчуються допустимі фрази, структура допустимих фраз стає більш складною.

Граф породження всіх можливих еталонних мовленнєвих сигналів для цього випадку подано на рис. 7. Тепер заводимо багато головних станів $(k,0)$, $k=1:K$. Ці головні стани розташуємо на вході до слова k .

Стартуючи зі стану $s_{\text{поч.}}$, за 0 тактів переходимо в головний стан $(k,0)$, $k \in V_{\text{поч.}}$, а звідти, також за 0 тактів, переходимо на генерацію еталонних мовленнєвих сигналів початкових слів.

Щоб відповісти на питання, яка послідовність елі з передається поданою для аналізу послідовністю спостережених елементів $X_{0i}=(x_1, x_2, \dots, x_i, \dots, x_l)$ чи $J_{0i}=(j_1, j_2, \dots, j_i, \dots, j_l)$, треба знайти для цієї послідовності найбільш правдоподібну траєкторію із $s_{\text{поч.}}$ в $s_{\text{кінц.}}$ на графі злитого мовлення рис. 7 та вказати, через які слова та в якій послідовності слів вона проходить.

З цієї метою спочатку зробимо такі початкові установки для моменту часу $i=0$: для головного стану $(k,0)$, $k \in V_{\text{поч.}}$, покладемо $F_0(k,0)=0$, $v_0(k,0)=0$, а для всіх головних станів решти слів та всіх внутрішніх станів слів відповідно $F_{0,2}^k \langle \phi_0 3 \rangle = F_0 \langle \rangle$, $v_{0,2}^k \langle \phi_0 3 \rangle = v_0 \langle \rangle$, $(F_{0,1}^k \langle \phi_s 1 \rangle = -M$, $F_{0,2}^k \langle \phi_s 1 \rangle = -M$, $F_{0,1}^k \langle \phi_s 2 \rangle = -M$,

$F_{0,2}^k \langle \phi_s 2 \rangle = -M$, $F_{0,1}^k \langle \phi_s 3 \rangle = -M$, $F_{0,2}^k \langle \phi_s 3 \rangle = -M$,
 $s=1:q^k$, $k=1:K$.

Далі в циклі по $i=1:l$, з появою чергового спостережуваного елемента x_i або j_i , тобто 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 всіх допустимих фраз, що можуть закінчуватись словом k , $s=1:q^k$, $k=1:K$, інтегральні міри схожості $F_{i,1}^k \langle \phi_s 1 \rangle$, $F_{i,2}^k \langle \phi_s 1 \rangle$, $F_{i,1}^k \langle \phi_s 2 \rangle$, $F_{i,2}^k \langle \phi_s 2 \rangle$,

$F_{i,1}^k \langle \phi_s 3 \rangle$, $F_{i,2}^k \langle \phi_s 3 \rangle$, що визначаються для кожного із двох виходів із першого, другого чи третього станів поточної фонемі їх транскрипцій, знаходять як суми значень відповідних елементарних мір схожості, обчислених для поточного спостереженого елемента u , для кожного із двох виходів із першого, другого чи третього станів поточної фонемі $\ln p(j_i/\phi_s 1)$, $\ln p(\phi_s 1/\phi_s 1)$, $\ln p(\phi_s 2/\phi_s 1)$, $\ln p(j_i/\phi_s 2)$, $\ln p(\phi_s 2/\phi_s 2)$, $\ln p(\phi_s 3/\phi_s 2)$, $\ln p(j_i/\phi_s 3)$, $\ln p(\phi_s 3/\phi_s 3)$, з більшою із двох інтегральних мір схожості, накопичених для попереднього спостереженого елемента j_{i-1} на другому виході з третього стану попередньої фонемі та на першому виході з першого стану поточної фонемі на другому виході з першого стану поточної фонемі та на першому виході з другого стану поточної фонемі й на другому виході з другого стану поточної фонемі та на першому виході з третього стану поточної фонемі, відповідно:

$F_{i,1}^k \langle \phi_s 1 \rangle = \max_{k_{i-1,2}} \{ F_{i-1,2}^k \langle \phi_{s-2} 3 \rangle + \ln p(\phi_s 1/\phi_s 1) \}$,
 $F_{i,2}^k \langle \phi_s 1 \rangle = \max_{k_{i-1,2}} \{ F_{i-1,2}^k \langle \phi_{s-2} 3 \rangle + \ln p(\phi_s 2/\phi_s 1) \}$,
 $F_{i,1}^k \langle \phi_s 2 \rangle = \max_{k_{i-1,2}} \{ F_{i-1,1}^k \langle \phi_{s-1} 1 \rangle + \ln p(\phi_s 2/\phi_s 2) \}$,
 $F_{i,2}^k \langle \phi_s 2 \rangle = \max_{k_{i-1,2}} \{ F_{i-1,1}^k \langle \phi_{s-1} 2 \rangle + \ln p(\phi_s 3/\phi_s 2) \}$,
 $F_{i,1}^k \langle \phi_s 3 \rangle = \max_{k_{i-1,2}} \{ F_{i-1,2}^k \langle \phi_{s-2} 3 \rangle + \ln p(\phi_s 3/\phi_s 3) \}$,
 $F_{i,2}^k \langle \phi_s 3 \rangle = \max_{k_{i-1,2}} \{ F_{i-1,1}^k \langle \phi_{s-1} 3 \rangle + \ln p(\phi_s 4/\phi_s 3) \}$,
 причому кожна із величин $F_{i,1}^k \langle \phi_s 1 \rangle$, $F_{i,2}^k \langle \phi_s 1 \rangle$,
 $F_{i,1}^k \langle \phi_s 2 \rangle$, $F_{i,2}^k \langle \phi_s 2 \rangle$, $F_{i,1}^k \langle \phi_s 3 \rangle$, $F_{i,2}^k \langle \phi_s 3 \rangle$, супроводжується потенційно-оптимальним початком k -го слова $v_{i,1}^k \langle \phi_s 1 \rangle$, $v_{i,2}^k \langle \phi_s 1 \rangle$, $v_{i,1}^k \langle \phi_s 2 \rangle$, $v_{i,2}^k \langle \phi_s 2 \rangle$, $v_{i,1}^k \langle \phi_s 3 \rangle$, $v_{i,2}^k \langle \phi_s 3 \rangle$, що дорівнюють потенційно-оптимальним початкам k -го слова $v_{i-1,2}^k \langle \phi_{s-2} 3 \rangle$ чи $v_{i-1,1}^k \langle \phi_{s-1} 1 \rangle$,
 $v_{i-1,2}^k \langle \phi_{s-1} 1 \rangle$ чи $v_{i-1,1}^k \langle \phi_{s-1} 2 \rangle$, $v_{i-1,2}^k \langle \phi_{s-1} 2 \rangle$ чи

$v_{i-1,1}^k \langle \phi_{s-1} 3 \rangle$, відповідно, в залежності від того який член, перший чи другий є більшим у виразах для $F_{i,1}^k \langle \phi_s 1 \rangle$, $F_{i,2}^k \langle \phi_s 1 \rangle$, $F_{i,1}^k \langle \phi_s 2 \rangle$, $F_{i,2}^k \langle \phi_s 2 \rangle$, $F_{i,1}^k \langle \phi_s 3 \rangle$, $F_{i,2}^k \langle \phi_s 3 \rangle$.

Оброблення i -го елемента x_i або j_i завершуємо обчисленням

$$F_i \langle 0 \rangle = \max_{k \in V_k} F_{i,2}^k \langle \phi_q 3 \rangle,$$

$$k_i \langle 0 \rangle = \arg \max_{k \in V_k} F_{i,2}^k \langle \phi_q 3 \rangle, v_i \langle 0 \rangle = v_{i,2}^{k_i \langle 0 \rangle} \langle \phi_q 3 \rangle, \text{ тобто}$$

вияснюємо, з якою найбільшою інтегральною мірою схожості, яке слово та коли воно могло початись, закінчуючись в момент i . Одночасно для кінцевого стану $s_{\text{кінц.}}$ підраховуємо

$$F_i \langle s_{\text{кінц.}} \rangle = \max_{k \in V_{\text{кінц.}}} F_{i,2}^k \langle \phi_q 3 \rangle,$$

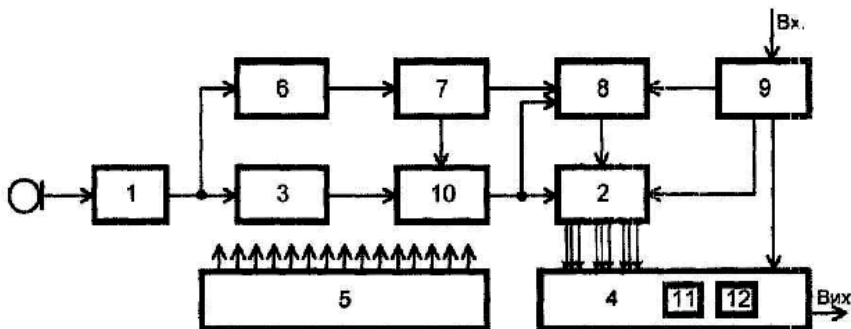
$$l_i \langle s_{\text{кінц.}} \rangle = \arg \max_{k \in V_{\text{кінц.}}} F_{i,2}^k \langle \phi_q 3 \rangle, v_i \langle s_{\text{кінц.}} \rangle = v_{i,2}^{l_i \langle s_{\text{кінц.}} \rangle} \langle \phi_q 3 \rangle,$$

тобто вияснюємо, яке слово може завершити допустиму фразу в момент i . Щоб підготуватись до прийняття та оброблення чергового спостережуваного елемента x_{i+1} або j_{i+1} , робимо пересилання $F_{i,2}^k \langle \phi_0 3 \rangle = F_i \langle 0 \rangle$, $v_{0,2}^k \langle \phi_0 3 \rangle = i$, $k=1:K$ та переходимо до наступного циклу рекурентних обчислень.

Після їх завершення формуємо відповідь розпізнавання за масивом $(F_i(k,0), v_i(k,0), k_i(k,0))$, $F_i(s_{\text{кінц.}}), k_i(s_{\text{кінц.}}), v_i(s_{\text{кінц.}})$, $i=1:l$, $k=1:K$.

Значення $F_i(s_{\text{кінц.}})$ визначає найбільш інтегральну міру схожості, а відповідь розпізнавання у вигляді послідовності слів, що передана мовленнєвим сигналом $X_{0i}=(x_1, x_2, \dots, x_i)$ або $J_{0i}=(j_1, j_2, \dots, j_i)$, формуємо за таким алгоритмом виписування:

$k_1^* = k_1 \langle s_{\text{кінц.}} \rangle$ буде першим з кінця словом, що почалося в момент $v_1^* = v_1 \langle s_{\text{кінц.}} \rangle$, $k_2^* = k_{v_1} \langle s_{\text{кінц.}} \rangle$ буде другим з кінця словом, що почалося в момент $v_2^* = v_{v_1} \langle s_{\text{кінц.}} \rangle$, $k_3^* = k_{v_2} \langle s_{\text{кінц.}} \rangle$ буде третім з кінця словом, що почалося в момент $v_3^* = v_{v_2} \langle s_{\text{кінц.}} \rangle$. І так далі. Виписування завершуємо при досягненні $v^*=0$.



Фиг. 1

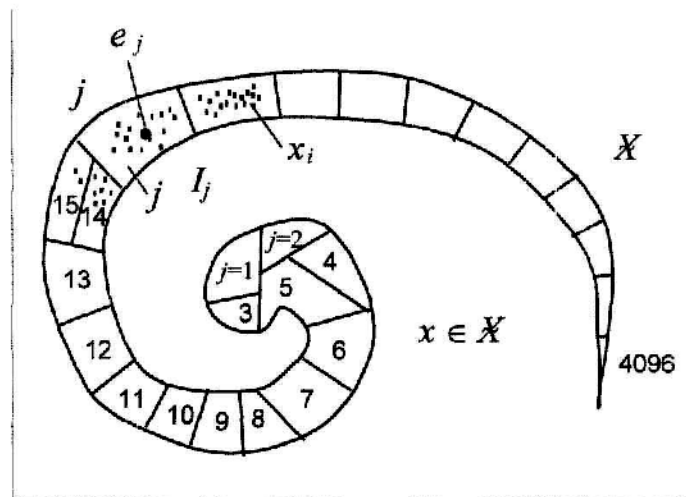


Fig. 2

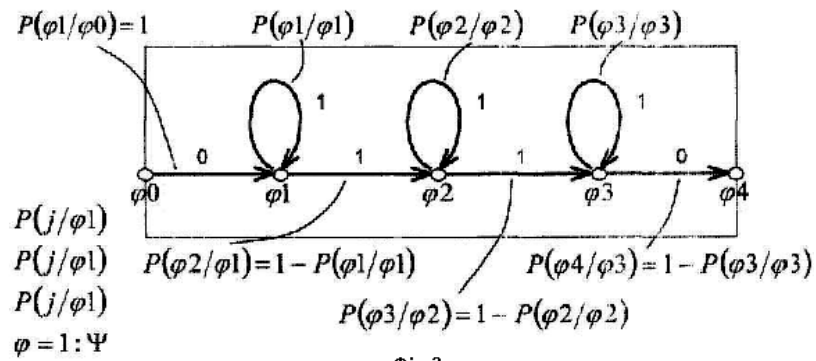


Fig. 3

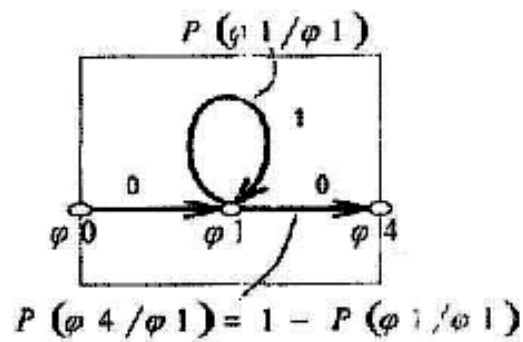


Fig. 4

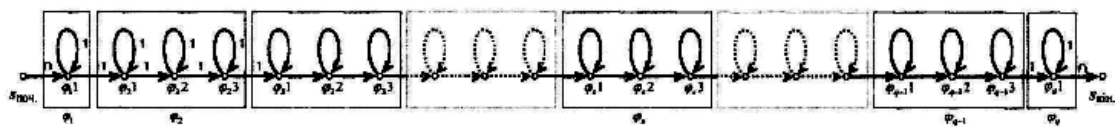
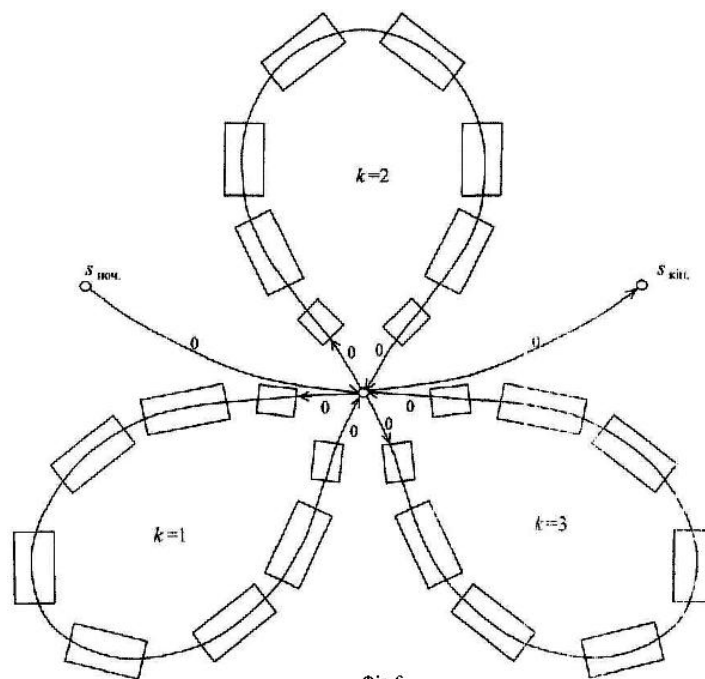
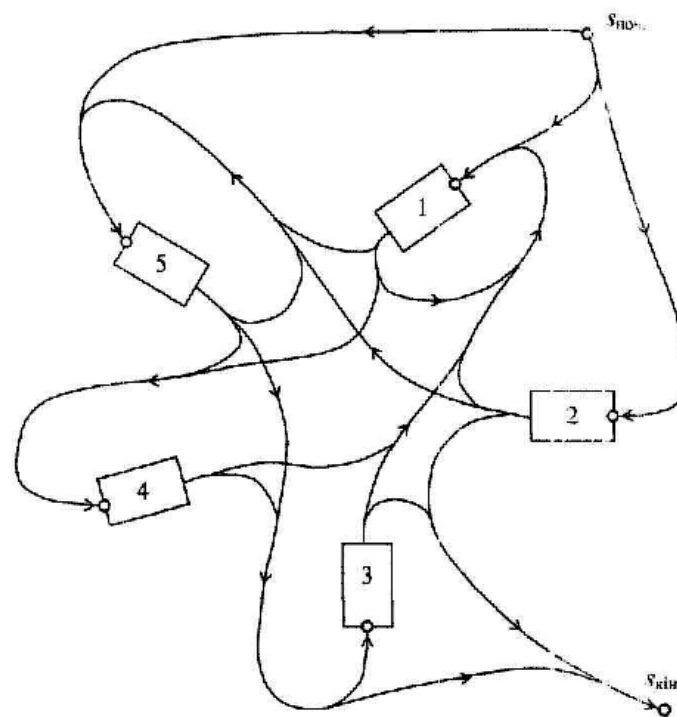


Fig. 5



Фиг. 6



Фиг. 7