



УКРАЇНА

(19) UA (11) 50038 (13) U
(51) МПК (2009)
G10L 15/00МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИДЕРЖАВНИЙ ДЕПАРТАМЕНТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІОПИС
ДО ПАТЕНТУ
НА КОРИСНУ МОДЕЛЬвидається під
відповідальність
власника
патенту

(54) СПОСІБ ПОФОНЕМНОГО РОЗПІЗНАВАННЯ УСНИХ КОМАНД ТА УСТАЛЕНИХ СЛОВОСПОЛУЧЕНЬ

1

2

(21) u200911559

(22) 13.11.2009

(24) 25.05.2010

(46) 25.05.2010, Бюл.№ 10, 2010 р.

(72) ВІНЦЮК ТАРАС КЛИМОВИЧ, ГРИЦЕНКО ВО-
ЛОДИМИР ІЛЛІЧ(73) МІЖНАРОДНИЙ НАУКОВО-НАВЧАЛЬНИЙ
ЦЕНТР ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА СИС-
ТЕМ

(57) Спосіб пофонемного розпізнавання усних команд та усталених словосполучень, що ґрунтується на поданні мовленнєвого сигналу послідовностями елементів-векторів із значень поточних параметрів аналізу мовленнєвого сигналу, який включає знаходження значень елементарних мір схожості кожного спостережуваного елемента на всі еталонні елементи еталонних мовленнєвих 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 кожна особа характеризується своїм індивідуальним усномовним паспортом, що укладається при разовому зачитуванні вголос цією особою стандартизованого тексту - навчальної вибірки; індивідуальний усномовний паспорт людини складають: задана кількість еталонних елементів, які найкращим чином апроксимують всі спостережувані елементи навчальної вибірки та визначають розбиття мультимножини спостережуваних елементів навчальної вибірки на задану кількість кластерів й параметри моделей всіх фонем у різному фонемному контексті - попередньої та наступної фонем; цими моделями фонем є ланцюги породжувальних граматик з п'яти прихованих станів, що моделюють три стадії розвитку процесу породження реалізації фонем, а параметрами моделей є: ймовірність переходу з нульового стану в перший стан, що дорівнює одиниці, ймовірність переходу з першого стану в перший же стан та ймовірність переходу з першого стану в другий стан, що доповнює до одиниці попередню ймовірність, ймовірність переходу з другого стану в другий же стан та ймовірність переходу з другого стану в третій стан, що доповнює до одиниці попередню ймовірність, ймовірність переходу з третього стану в третій же стан та ймовір-

ність переходу з третього стану в четвертий стан, що доповнює до одиниці попередню ймовірність, та ймовірності попадань спостережуваних елементів в кожен із всіх кластерів за умови перебування в першому, другому та третьому станах кожної фонемі; при розпізнаванні для кожного поточного спостережуваного елемента визначають номер кластера, в який цей елемент попадає, а як елементарну міру належності спостережуваного елемента-вектора до першої, другої чи третьої фази фонемі використовують суму логарифмів ймовірності спостереженого кластера за умови першого, другого чи третього станів фонемі та ймовірності наступного переходу з першого в перший або другий, з другого в другий або третій, з третього в третій або четвертий стани фонемі відповідно до фази фонемі; еталонні мовленнєві образи усних команд та усталених словосполучень формують шляхом об'єднання у послідовності ланцюгів породжувальних граматик 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; інтегральні міри схожості початкових еталонних образів усних команд та усталених словосполучень, що визначаються для кожного із двох виходів із першого, другого чи третього станів поточної фонемі їх транскрипцій, знаходять як суми значень відповідної елементарної міри схожості, обчисленої для поточного спостереженого елемента для кожного із двох виходів із першого, другого чи третього станів поточної фонемі, з більшою із двох інтегральних мір схожості, накопичених для попереднього спостереженого вектора-елемента 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; значення інтегральної міри схожості, накопичене після оброблення останнього спостереженого елемента на другому виході із третього стану останньої фонемі, яка визначається фонетичною транскрипцією усної команди або словосполучення, визначають схожість пред'явленого

(13) U
(11) 50038
(19) UA

мовленнєвого сигналу на цю усну команду або словосполучення; пред'явлений мовленнєвий сигнал відносять до тієї усної команди або усталеного

словосполучення, для котрого накопичена схожість є абсолютно найбільшою.

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

Відомий спосіб та пристрій пофонемного розпізнавання усних команд та усталених словосполучень (дивись патент України №48082)

Сутність відомого способу полягає в тому, що розпізнавання окремо вимовлюваних усних команд та усталених словосполучень ґрунтується на поданні мовленнєвого сигналу послідовностями елементів-векторів із значень поточних параметрів аналізу мовленнєвого сигналу, який включає знаходження значень елементарних мір схожості кожного спостережуваного елемента на всі еталонні елементи еталонних мовленнєвих образів, визначення схожості мовленнєвих образів шляхом рекурентного накопичення методом динамічного програмування інтегральних мір схожості послідовності спостережуваних елементів на послідовності еталонних образів, що розпізнаються.

Але він має недоліки, які полягають в тому, що в нього мала швидкодія та низька надійність розпізнавання.

В основу моделі покладена задача за рахунок введення нових операцій обробки сигналів створити спосіб для пофонемного розпізнавання усних команд та усталених словосполучень, що мають високу швидкодію та надійність розпізнавання без необхідності перенастроювання системи.

Поставлена задача вирішується способом розпізнавання окремо вимовлюваних усних команд та усталених словосполучень, що ґрунтується на поданні мовленнєвого сигналу послідовностями елементів-векторів із значень поточних параметрів аналізу мовленнєвого сигналу, який включає знаходження значень елементарних мір схожості кожного спостережуваного елемента на всі еталонні елементи еталонних мовленнєвих 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 кожна особа характеризується своїм індивідуальним усномовним паспортом, що укладається при разовому зачитуванні вголос цією особою стандартизованого тексту - навчальної вибірки; індивідуальний усномовний паспорт людини складають: задана кількість еталонних елементів, які найкращим чином апроксимують всі спостережувані елементи навчальної вибірки та визначають розбиття мультимножини спостережуваних елементів навчальної вибірки на задану кількість кластерів й параметри моделей всіх фонем у різному фонемному контексті - попередньої та наступної фонем; цими моделями фонем є ланцюги породжувальних граматик з п'яти прихова-

них станів, що моделюють три стадії розвитку процесу породження реалізації фонем, а параметрами моделей є: ймовірність переходу з нульового стану в перший стан, що дорівнює одиниці, ймовірність переходу з першого стану в перший же стан та ймовірність переходу з першого стану в другий стан, що доповнює до одиниці попередню ймовірність, ймовірність переходу з другого стану в другий же стан та ймовірність переходу з другого стану в третій стан, що доповнює до одиниці попередню ймовірність, ймовірність переходу з третього стану в третій же стан та ймовірність переходу з третього стану в четвертий стан, що доповнює до одиниці попередню ймовірність, та ймовірності попадань спостережуваних елементів в кожен із всіх кластерів за умови перебування в першому, другому та третьому станах кожної фонемі; при розпізнаванні для кожного поточного спостережуваного елемента визначається номер кластера, в який цей елемент попадає, а в якості елементарної міри належності спостережуваного елемента-вектора до першої, другої чи третьої фази фонемі використовується сума логарифмів ймовірності спостереженого кластера за умови першого, другого чи третього станів фонемі та ймовірності наступного переходу з першого в перший або другий, з другого в другий або третій, з третього в третій або четвертий стани фонемі відповідно до фази фонемі; еталонні мовленнєві образи усних команд та усталених словосполучень формують шляхом об'єднання у послідовності ланцюгів породжувальних граматик 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; інтегральні міри схожості початкових еталонних образів усних команд та усталених словосполучень, що визначаються для кожного із двох виходів із першого, другого чи третього станів поточної фонемі їх транскрипцій, знаходять як суми значень відповідної елементарної міри схожості, обчисленої для поточного спостереженого елемента для кожного із двох виходів із першого, другого чи третього станів поточної фонемі, з більшою із двох інтегральних мір схожості, накопичених для попереднього спостереженого вектора-елемента 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; значення інтегральної міри схожості, накопичене після оброблення останнього спостереженого елемента на другому виході із третього стану останньої фонемі, яка визначається фонетичною транскрипцією усної команди або словос-

получення, визначає схожість пред'явленого мовленнєвого сигналу на цю усну команду або словосполучення; пред'явлений мовленнєвий сигнал відноситься до тієї усної команди або усталеного словосполучення, для котрого накопичена схожість є абсолютно найбільшою.

На фіг. 1 представлена структурна схема пристрою, що реалізує спосіб; 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 та навчальної 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 голо- сом називає пристрій.

В аналізаторі 1 мовленнєвий сигнал, що подається з 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 i 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 $B_i = \{B_0, B_1, \dots, B_m\}$. Розв'язуючи систему рівнянь $\sum_{u=1}^m a_u B_{i-u} = -B_i$, $v = 1:m$, 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: $a = (a_1, a_2, \dots, a_m)$ $b = (b_0, b_1, b_2, \dots, b_m)$;

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

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 $X_0 = (x_1, x_2, \dots, x_i, \dots, x_\ell)$ із ℓ спостережених елементів x_i .

В блоці 7 кластерного аналізу та обчислення параметрів фонем спершу за навчальною вибіркою X_0 з-посеред усіх і спостережених елементів x_i за допомогою деякої ітераційної процедури вибираємо задану кількість $J, J \ll \ell$, елементів (b_j^*, σ_j^*) , $j = 1:J$, таких, які б найкращим чином апроксимували всі елементи навчальної вибірки:

$$(b_j^*, \sigma_j^*) = 1:J \left\{ \arg \min_{(b_j, \sigma_j)} \min_{j=1:J} \sum_{i=1}^J \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} \langle b_i, b_j \rangle \right) \right\}$$

де через (B, b) позначено скалярний добуток векторів-елементів B і b розмірності $m+1$, а через $\langle l_i \rangle$ - розбиття навчальної вибірки на J кластерів. В останньому виразі через

$$I_j = \left\{ i: i_j = \arg \min \left(\frac{M}{2} \ln \sigma_j^2 + \frac{1}{2\sigma_j^2} \langle b_i, b_j \rangle \right) = j \right\} \quad \text{позна-}$$

чено j -ий кластер. Всі елементи j -го кластера найкраще апроксимуються представником цього кластера

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

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

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

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

$$(b_j^{(r)}, \sigma_j^{(r)}) = \arg \min_{(b_w, \sigma_w), w \in I_j^{(r-1)}} \left(\frac{M}{2} \ln \sigma_w^2 + \frac{1}{2\sigma_w^2} \langle b_i, b_w \rangle \right), \quad j = 1-J \quad \text{нових}$$

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

$$I_j^* = \left\{ i: i_j^* = \arg \min_j \left(\frac{M}{2} \ln(\sigma_j^2) + \frac{1}{2(\sigma_j^2)} (B_i, b_j^{(r)}) \right) = j \right\}, j = 1: J, \cdot^l$$

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 векторному квантувачеві 10 при визначенні номера еталонного елемента j_i , що є найбільш схожим на спостережуваний елемент $B_i: j_i = \arg \min_j g(B_i, (b_j, \sigma_j))$, де $g(B_i, (b_j, \sigma_j))$

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

на міра схожості елементів B_i та (b_j, σ_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 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 фонемою. 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 $p(j/\varphi_t)$, $j =$

$1:J$, $i = 1, 2, 3$ попадань спостережуваних елементів B в кожен із всіх кластерів $j = \arg \min_{u=1:J} g(B, (b_u, \sigma_u))$ за

умови перебування в першому, другому та третьому станах t кожної фонемою φ .

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

Представники кластерів або еталонні елементи-вектори (b_j, σ_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_{01} 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 граматик фонем з п'яти прихованих станів відповідно до фонетичних транскрипцій усних команд або словосполучень, причому так, щоб вихідний, четвертий, стан попередньої фонемою збігався з нульовим та першим станами наступної.

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

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

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

Якщо ж сегмент $X_{\mu v} = (x_{\mu+1}, x_{\mu+2}, \dots, x_v)$ або $J_{\mu v} = (j_{\mu+1}, j_{\mu+2}, \dots, j_v)$ розглядати як реалізацію якоїсь усної команди чи словосполучення з фонетич-

ною 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:

$$\begin{aligned}
G_{\mu v} / \langle \phi_1, \phi_2, \dots, \phi_s, \dots, \phi_q; \langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:q \rangle &= \sum_{s=1}^q G_{\mu s-1 \mu s} / \langle \mu_s, w_{s1}, w_{s2} \rangle \\
&= \sum_{s=1}^q \left[(w_{s1} - \mu_{s-1} - 1) \ln p(\phi_s 1 / \phi_s 1) \ln p(\phi_s 2 / \phi_s 1) + \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(\phi_i / \phi_s 1) \right] \\
&+ \left[(w_{s2} - w_{s1} - 1) \ln p(\phi_s 2 / \phi_s 2) \ln p(\phi_s 3 / \phi_s 2) + \sum_{i=w_{s1}+1}^{w_{s2}} \ln p(\phi_i / \phi_s 2) \right] \\
&+ \left[(\mu_s - w_{s2} - 1) \ln p(\phi_s 3 / \phi_s 3) \ln p(\phi_s 4 / \phi_s 3) + \sum_{i=w_{s2}+1}^{\mu_s} \ln p(\phi_i / \phi_s 3) \right],
\end{aligned}$$

де через

$$J_{\mu s-1 \mu s} / \langle \mu_s, w_{s1}, w_{s2} \rangle = (J_{\mu s-1 w_{s1}}, J_{w_{s1} w_{s2}}, J_{w_{s2} \mu_s}) = (j_{\mu s-1+1}, \dots, j_{w_{s1}}, j_{w_{s1}+1}, \dots, j_{w_{s2}}, j_{w_{s2}+1}, \dots, j_{\mu})$$

позначено можливий s-тий підсегмент для фонеми ϕ_s з трьома його відповідними підпідсегментами, причому $\mu_0 = \mu$, $\mu_{s-1} < \mu_s$, $\mu_q = v$. Оптимальне розбиття чи сегментація $\langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:q$, яка максимізує критерій правдоподібності, визна-

чає інтегральну міру схожості сегмента $X_{\mu v}$ чи $J_{\mu v}$ на усну команду з фонетичною транскрипцією ψ_{0q} :

$$\begin{aligned}
G_{\mu v} / \psi_{0q} &= G_{J_{\mu v}} / \langle \phi_1, \phi_2, \dots, \phi_s, \dots, \phi_q; \langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:q \rangle \\
&= \max_{\langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:q} G_{\mu v} / \langle \phi_1, \phi_2, \dots, \phi_s, \dots, \phi_q; \langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:q \rangle
\end{aligned}$$

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

Вся ж навчальна вибірка $X_{0\ell} = \langle x_1, x_2, \dots, x_i, \dots, x_{\ell} \rangle$ чи $J_{0\ell} = \langle j_1, j_2, \dots, j_i, \dots, j_{\ell} \rangle$, 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

$$\begin{aligned}
G_{00Q} / \psi_{0Q}; P; \langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:q &= \sum_{s=1}^Q \left[(w_{s1} - \mu_{s-1} - 1) \ln p(\phi_s 1 / \phi_s 1) \ln p(\phi_s 2 / \phi_s 1) + \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(\phi_i / \phi_s 1) \right] \\
&+ \left[(w_{s2} - w_{s1} - 1) \ln p(\phi_s 2 / \phi_s 2) \ln p(\phi_s 3 / \phi_s 2) + \sum_{i=w_{s1}+1}^{w_{s2}} \ln p(\phi_i / \phi_s 2) \right] \\
&+ \left[(\mu_s - w_{s2} - 1) \ln p(\phi_s 3 / \phi_s 3) \ln p(\phi_s 4 / \phi_s 3) + \sum_{i=w_{s2}+1}^{\mu_s} \ln p(\phi_i / \phi_s 3) \right]
\end{aligned}$$

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

$$\begin{aligned}
G_{00Q} / \psi_{0Q}; P; \langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:q &= \sum_{\phi \in \psi} \sum_{s: \phi_s = \phi}^Q \left[(w_{s1} - \mu_{s-1} - 1) \ln p(\phi 1 / \phi 1) \ln p(\phi 2 / \phi 1) + \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(\phi_i / \phi 1) \right] \\
&+ \left[(w_{s2} - w_{s1} - 1) \ln p(\phi 2 / \phi 2) \ln p(\phi 3 / \phi 2) + \sum_{i=w_{s1}+1}^{w_{s2}} \ln p(\phi_i / \phi 2) \right] \\
&+ (\mu_s - w_{s2} - 1) \ln p(\phi 3 / \phi 3) \ln p(\phi 4 / \phi 3) + \sum_{i=w_{s2}+1}^{\mu_s} \ln p(\phi_i / \phi 3)
\end{aligned}$$

З останнього запису випливає, що якщо відомі границі $\langle \mu_s, w_{s1}, w_{s2} \rangle, s=1:Q$ сегментів всіх реалізацій всіх фонем із навчальної вибірки, то мак-

симально правдоподібні оцінки ймовірностей обчислюються як:

$$\begin{aligned}
p(\phi_1/\phi_1) &= \sum_{s:\phi} \sum_{S=\phi} \left(\frac{v_{s1} - \mu_{s-1} - 1}{v_{s1} - \mu_{s-1}} \right) \sum_{S=\phi} \left(\frac{v_{s1} - \mu_{s-1}}{v_{s1} - \mu_{s-1}} \right) \\
p(\phi_2/\phi_1) &= \sum_{s:\phi} \sum_{S=\phi} \left(\frac{1}{v_{s1} - \mu_{s-1}} \right) \sum_{S=\phi} \left(\frac{v_{s1} - \mu_{s-1}}{v_{s1} - \mu_{s-1}} \right) \\
p(\phi_2/\phi_2) &= \sum_{s:\phi} \sum_{S=\phi} \left(\frac{v_{s2} - w_{s1} - 1}{v_{s2} - w_{s1}} \right) \sum_{S=\phi} \left(\frac{v_{s2} - w_{s1}}{v_{s2} - w_{s1}} \right) \\
p(\phi_3/\phi_2) &= \sum_{s:\phi} \sum_{S=\phi} \left(\frac{1}{v_{s2} - w_{s1}} \right) \sum_{S=\phi} \left(\frac{v_{s2} - w_{s1}}{v_{s2} - w_{s1}} \right) \\
p(\phi_3/\phi_3) &= \sum_{s:\phi} \sum_{S=\phi} \left(\frac{v_{s3} - w_{s2} - 1}{v_{s3} - w_{s2}} \right) \sum_{S=\phi} \left(\frac{v_{s3} - w_{s2}}{v_{s3} - w_{s2}} \right) \\
p(\phi_4/\phi_3) &= \sum_{s:\phi} \sum_{S=\phi} \left(\frac{1}{v_{s3} - w_{s2}} \right) \sum_{S=\phi} \left(\frac{v_{s3} - w_{s2}}{v_{s3} - w_{s2}} \right) \\
p(\phi/\phi_1) &= \left(\sum_{s:\phi} \sum_{S=\phi} \sum_{i=\mu_{s-1}+1}^{w_{s1}} \mathbb{I}_{j=i} \right) / \sum_{S=\phi} \left(\frac{v_{s1} - \mu_{s-1}}{v_{s1} - \mu_{s-1}} \right) \\
p(\phi/\phi_2) &= \left(\sum_{s:\phi} \sum_{S=\phi} \sum_{i=w_{s1}+1}^{w_{s2}} \mathbb{I}_{j=i} \right) / \sum_{S=\phi} \left(\frac{v_{s2} - w_{s1}}{v_{s2} - w_{s1}} \right) \\
p(\phi/\phi_3) &= \left(\sum_{s:\phi} \sum_{S=\phi} \sum_{i=w_{s2}+1}^{\mu_s} \mathbb{I}_{j=i} \right) / \sum_{S=\phi} \left(\frac{v_{s3} - w_{s2}}{v_{s3} - w_{s2}} \right) \quad j = 1: J, \phi \in \psi.
\end{aligned}$$

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

$X_{0\ell} = (\phi_1, x_2, \dots, x_i, \dots, x_\ell)$ 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 (b_j, σ_j) , $j = 1: J$:

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

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

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

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

$$\begin{aligned}
& \left\{ \phi_s, w_{s1}, w_{s2}, s = 1: Q \right\} \arg \max_{\{s, w_{s1}, w_{s2}, s = 1: q\}} \sum_{\phi \in \psi} \sum_{S=\phi} \left[\left(\frac{v_{s1} - \mu_{s-1} - 1}{v_{s1} - \mu_{s-1}} \right) \ln p(\phi_1/\phi_1) \right. \\
& + \sum_{i=\mu_{s-1}+1}^{w_{s1}} \ln p(\phi_2/\phi_2) \left. \left[\left(\frac{v_{s2} - w_{s1} - 1}{v_{s2} - w_{s1}} \right) \ln p(\phi_2/\phi_2) \right. \right. \\
& + \sum_{i=w_{s2}+1}^{w_{s2}} \ln p(\phi_3/\phi_3) \left. \left. \left[\left(\frac{v_{s3} - w_{s2} - 1}{v_{s3} - w_{s2}} \right) \ln p(\phi_3/\phi_3) \right. \right. \right. \\
& \left. \left. \left. + \sum_{i=w_{s2}+1}^{\mu_s} \ln p(\phi_4/\phi_4) \right] \right] \right]
\end{aligned}$$

навчальної вибірки та на її основі - нові значення параметрів моделей фонем $P^{(t)} = \{p^{(t)}(\phi/\phi_t), p^{(t)}(j/\phi_t), j = 1: J, t = 1: 3, \phi \in \psi\}$.

вчальної вибірки знаходимо ймовірності $P^{(0)} = \{p^{(0)}(\phi/\phi_t), p^{(0)}(j/\phi_t), j = 1: J, t = 1: 3, \phi \in \psi\}$.

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

$$\begin{aligned}
\rho^{\mathbb{C}_1/\varphi_1} &\rhd \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \left(\mathbb{V}_{\mathbf{s}_1 - \mu} \mathbb{C}_{\mathbf{s} - 1} \right) \right) / \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \left(\mathbb{V}_{\mathbf{s}_1 - \mu} \mathbb{C}_{\mathbf{s} - 1} \right) \right) \\
\rho^{\mathbb{C}_2/\varphi_2} &\rhd \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \left(\mathbb{V}_{\mathbf{s}_2 - w} \mathbb{C}_{\mathbf{s}_1 - 1} \right) \right) / \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \left(\mathbb{V}_{\mathbf{s}_2 - w} \mathbb{C}_{\mathbf{s}_1} \right) \right) \\
\rho^{\mathbb{C}_3/\varphi_3} &\rhd \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \left(\mathbb{V}_{\mathbf{s} - w} \mathbb{C}_{\mathbf{s}_2 - 1} \right) \right) / \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \left(\mathbb{V}_{\mathbf{s} - w} \mathbb{C}_{\mathbf{s}_2} \right) \right) \\
\rho^{\mathbb{C}_1/\varphi_1} &\rhd \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \sum_{i=1}^w \mathbb{S}_1^i \left(\mathbb{J}_i = j \right) / \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \left(\mathbb{V}_{\mathbf{s}_1 - \mu} \mathbb{C}_{\mathbf{s} - 1} \right) \right) \right) \\
\rho^{\mathbb{C}_1/\varphi_2} &\rhd \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \sum_{i=w}^w \mathbb{S}_2^i \left(\mathbb{J}_i = j \right) / \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \left(\mathbb{V}_{\mathbf{s}_2 - w} \mathbb{C}_{\mathbf{s}_1} \right) \right) \right) \\
\rho^{\mathbb{C}_1/\varphi_3} &\rhd \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \sum_{i=w}^w \mathbb{S}^i \left(\mathbb{J}_i = j \right) / \left(\sum_{\mathbf{s}: \varphi \mathbf{S} = \varphi} \left(\mathbb{V}_{\mathbf{s} - w} \mathbb{C}_{\mathbf{s}_2} \right) \right) \right) j = 1: J, \varphi \in \Psi.
\end{aligned}$$

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 уста-

$$F_{0,2}^k(\bullet_0 3) = 0; F_1^k(\bullet_s 1) = -M, F_2^k(\bullet_s 1) = -M, F_1^k(\bullet_s 2) = -M, F_2^k(\bullet_s 2) = -M, F_1^k(\bullet_s 3) = -M, s = 1: q^k, k = 1: K,$$

є найбільш схожим на спостережуваний. Під управлінням контролера 5 з блока 9 зчитується ім'я першої фонем φ_1 першого слова. За цим іменем фонем та номером еталонного елемента j_1 з блока пам'яті 4а накопичуються 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 елементу j_1 , та
ім'ям фонем 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 мати-
мемо відповідно

$$\begin{aligned} F_{1,1}^k \varphi_1 &\models \max \left(C_{0,2}^k \varphi_0, \bigwedge_{0,1}^k \varphi_1 \right) \upharpoonright \text{Inp } \varphi_1 / \varphi_1 \upharpoonright \text{Inp } \varphi_1 / \varphi_1, \\ F_{1,2}^k \varphi_1 &\models \max \left(C_{0,2}^k \varphi_0, \bigwedge_{0,1}^k \varphi_1 \right) \upharpoonright \text{Inp } \varphi_1 / \varphi_1 \upharpoonright \text{Inp } \varphi_2 / \varphi_1, \\ F_{1,1}^k \varphi_2 &\models \max \left(C_{0,2}^k \varphi_1, \bigwedge_{0,1}^k \varphi_2 \right) \upharpoonright \text{Inp } \varphi_1 / \varphi_2 \upharpoonright \text{Inp } \varphi_2 / \varphi_2, \\ F_{1,2}^k \varphi_2 &\models \max \left(C_{0,2}^k \varphi_1, \bigwedge_{0,1}^k \varphi_2 \right) \upharpoonright \text{Inp } \varphi_1 / \varphi_2 \upharpoonright \text{Inp } \varphi_3 / \varphi_2, \\ F_{1,1}^k \varphi_3 &\models \max \left(C_{0,2}^k \varphi_2, \bigwedge_{0,1}^k \varphi_3 \right) \upharpoonright \text{Inp } \varphi_1 / \varphi_3 \upharpoonright \text{Inp } \varphi_3 / \varphi_3, \\ F_{1,2}^k \varphi_3 &\models \max \left(C_{0,2}^k \varphi_2, \bigwedge_{0,1}^k \varphi_3 \right) \upharpoonright \text{Inp } \varphi_1 / \varphi_3 \upharpoonright \text{Inp } \varphi_4 / \varphi_3. \end{aligned}$$

Далі за тим же номером еталонного елемента j_1 , послідовно для всіх решти фонем $s = 2:q^k$ слова k діємо аналогічно. За іменем фонemi φ_s з блока пам'яті 4а накопичуваних інтегральних мір схожості на обчислювач 4 інтегральних мір схожості зчитуються значення

$F_{0,1}^k \phi_{s-1}^1, F_{0,2}^k \phi_{s-1}^1, F_{0,1}^k \phi_{s-2}^1, F_{0,2}^k \phi_{s-2}^1, F_{0,1}^k \phi_{s-3}^1, F_{0,2}^k \phi_{s-3}^1$ на ті ж регістри 41, 42, 42а, 43, 43а, 44, 44а відповідно. Одночасно за номером еталонного елемента j_1 та ім'ям фонему ϕ_s на другі входи суматорів 46, 47, 47а, 48, 49, 49а, 50, 51, 51а надходять логарифми ймовірностей $\ln p(j_1 / \phi_s 1), \ln p(\phi_s 1 / \phi_s$

1), $\ln p(\phi_s 2 / \phi_s 1), \ln p(j_1 / \phi_s 2), \ln p(\phi_s 2 / \phi_s 2), \ln p(\phi_s 3 / \phi_s 1 2), \ln p(j_1 / \phi_s 3), \ln p(\phi_s 3 / \phi_s 3), \ln p(\phi_s 4 / \phi_s 3)$, відповідно. На виходах суматорів 47, 47а, 49, 49а, 51, 51а матимемо відповідно

$$\begin{aligned} F_{1,1}^k \phi_s 1 &= \max \left(F_{0,2}^k \phi_{s-2} 3, F_{0,1}^k \phi_{s-1} 1 \right) \ln p(\phi_s 1 / \phi_s 1) \ln p(\phi_s 1 / \phi_s 1), \\ F_{1,2}^k \phi_s 1 &= \max \left(F_{0,2}^k \phi_{s-2} 3, F_{0,1}^k \phi_{s-1} 1 \right) \ln p(\phi_s 1 / \phi_s 1) \ln p(\phi_s 2 / \phi_s 1), \\ F_{1,1}^k \phi_s 2 &= \max \left(F_{0,2}^k \phi_{s-1} 1, F_{0,1}^k \phi_{s-2} 2 \right) \ln p(\phi_s 2 / \phi_s 2) \ln p(\phi_s 2 / \phi_s 2), \\ F_{1,2}^k \phi_s 2 &= \max \left(F_{0,2}^k \phi_{s-1} 1, F_{0,1}^k \phi_{s-2} 2 \right) \ln p(\phi_s 2 / \phi_s 2) \ln p(\phi_s 3 / \phi_s 2), \\ F_{1,1}^k \phi_s 3 &= \max \left(F_{0,2}^k \phi_{s-2} 2, F_{0,1}^k \phi_{s-3} 3 \right) \ln p(\phi_s 3 / \phi_s 3) \ln p(\phi_s 3 / \phi_s 3), \\ F_{1,2}^k \phi_s 3 &= \max \left(F_{0,2}^k \phi_{s-2} 2, F_{0,1}^k \phi_{s-3} 3 \right) \ln p(\phi_s 3 / \phi_s 3) \ln p(\phi_s 4 / \phi_s 3). \end{aligned}$$

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 $F_{0,1}^k \phi_{s-1}^1, F_{0,2}^k \phi_{s-1}^1, F_{0,1}^k \phi_{s-2}^1, F_{0,2}^k \phi_{s-2}^1, F_{0,1}^k \phi_{s-3}^1, F_{0,2}^k \phi_{s-3}^1$ відповідно.

Після закінчення оброблення слова k на адресу $F_{0,2}^k \phi_0 3$ пересилаємо (-M).

Далі процес повторюємо в циклі по $k = 2 : K$. Далі процес повторюється в циклі по $i = 2 : \ell$.

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

$F_{1,1}^k \phi_s 1, F_{1,2}^k \phi_s 1, F_{1,1}^k \phi_s 2, F_{1,2}^k \phi_s 2, F_{1,1}^k \phi_s 3, F_{1,2}^k \phi_s 3$, що визначаються для кожного із двох виходів із першого, другого чи третього станів поточної фонемі їх транскрипцій, знаходять як суми значень

$$\begin{aligned} F_{1,1}^k \phi_s 1 &= \max \left(F_{1,2}^k \phi_{s-2} 3, F_{1,1}^k \phi_{s-1} 1 \right) \ln p(\phi_s 1 / \phi_s 1) \ln p(\phi_s 1 / \phi_s 1), \\ F_{1,2}^k \phi_s 1 &= \max \left(F_{1,2}^k \phi_{s-2} 3, F_{1,1}^k \phi_{s-1} 1 \right) \ln p(\phi_s 1 / \phi_s 1) \ln p(\phi_s 2 / \phi_s 1), \\ F_{1,1}^k \phi_s 2 &= \max \left(F_{1,2}^k \phi_{s-1} 1, F_{1,1}^k \phi_{s-2} 2 \right) \ln p(\phi_s 2 / \phi_s 2) \ln p(\phi_s 2 / \phi_s 2), \\ F_{1,2}^k \phi_s 2 &= \max \left(F_{1,2}^k \phi_{s-1} 1, F_{1,1}^k \phi_{s-2} 2 \right) \ln p(\phi_s 2 / \phi_s 2) \ln p(\phi_s 3 / \phi_s 2), \\ F_{1,1}^k \phi_s 3 &= \max \left(F_{1,2}^k \phi_{s-2} 2, F_{1,1}^k \phi_{s-3} 3 \right) \ln p(\phi_s 3 / \phi_s 3) \ln p(\phi_s 3 / \phi_s 3), \\ F_{1,2}^k \phi_s 3 &= \max \left(F_{1,2}^k \phi_{s-2} 2, F_{1,1}^k \phi_{s-3} 3 \right) \ln p(\phi_s 3 / \phi_s 3) \ln p(\phi_s 4 / \phi_s 3). \end{aligned}$$

Значення інтегральної міри схожості

$F_{1,2}^k \left(\phi_q k 3 \right)$, накопичене після оброблення остан-

нього спостереженого елемента на другому виході із третього стану останньої фонемі q^k , яка визначається фонетичною транскрипцією усної команди або словосполучення, визначає схожість пред'яв-

відповідних елементарних мір схожості, обчислених для поточного спостереженого елемента j_i для кожного із двох виходів із першого, другого чи третього станів поточної фонемі $p(j_i / \phi_s 1), \ln p(\phi_s 1 / \phi_s 1), \ln p(\phi_s 2 / \phi_s 1), \ln p(j_i / \phi_s 2), \ln p(\phi_s 2 / \phi_s 2), \ln p(\phi_s 3 / \phi_s 1 2), \ln p(j_i / \phi_s 3), \ln p(\phi_s 3 / \phi_s 3), \ln p(\phi_s 4 / \phi_s 3)$, з більшою із двох інтегральних мір схожості, накопичених для попереднього спостереженого елемента j_{i-1} на другому виході з третього стану попередньої фонемі $\phi_{s-1} 3$ 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 $\phi_s 2$ й на другому виході з другого стану поточної фонемі $\phi_s 2$ та на першому виході з третього стану поточної фонемі $\phi_s 3$, відповідно:

леного мовленнєвого 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.

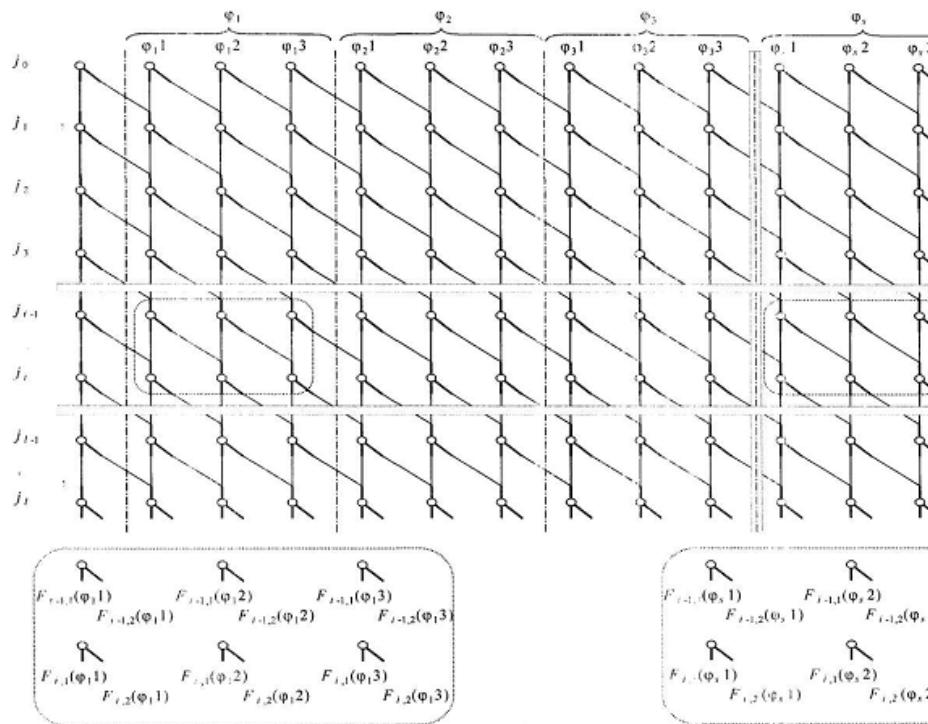


Fig. 8