

Винахід стосується техніки обробки мовленнєвої інформації при розпізнаванні великих словників і може бути використаний у системах диктування тексту, голосового керування промисловими об'єктами.

Відомий спосіб розпізнавання ізольованих слів [Патент ЄПВ №0420825, МПК G10L5/06, 1991р.]. Цей спосіб використовується при розпізнаванні слів дуже великих словників. Він заснований на визначенні та зберіганні для кожного слова в словнику фонетичної моделі, що складена з послідовності фонетичних символів, які відповідають фонемам слова, визначенні й зберіганні характеристичних параметрів, що показують енергію та спектральний склад фонем, порівнянні значень характеристичних параметрів для всіх моделей словника і виборі за допомогою алгоритму динамічного програмування невеликої кількості моделей кандидатів.

Недоліком цього способу є значні затрати часу та обладнання при обробці моделей слів.

Найбільш близьким до способу, що заявляється, є спосіб розпізнавання суцільно вимовлених слів [а.с. СРСР 1159059, МПК G10L1/00, 1985р.]. Цей спосіб включає виділення поточних параметрів, формування масивів номерів і моментів початку слів, що потенційно закінчуються, рекурентне накопичення інтегрального ступеня подібності та прийняття рішення про послідовність вимовлених слів при формуванні масивів рекурентне накопичують інтегральні ступені подібності між послідовністю відліків поточних параметрів від першого відліку до поточного відліку та еталонними сигналами суцільного мовлення й закінчуються усіма можливими еталонними елементами всіх слів словника, при цьому зчитують інтегральні ступені подібності, що накопичені для декількох попередніх відліків поточних параметрів, а до масивів записують код номера слова, що потенційно закінчується в даний поточний момент, та код моменту його початку, що відповідає максимальній з отриманих інтегральних ступенів подібності.

Даний спосіб оперує еталонними елементами, які відповідають словам. Це є причиною його недостатньої швидкодії при роботі з великими словниками.

Основу винаходу складає задача створення способу розпізнавання фраз, що вимовляються суцільно й складаються із слів великого словника, який має високу швидкодію та і дозволяє здійснювати розпізнавання в режимі реального часу.

Указаний технічний результат досягається тим, що згідно зі способом розпізнавання, який містить виділення поточних параметрів, рекурентне зіставлення участку мовлення з усіма словами словника й обчислення інтегральних ступенів близькості, прийняття рішення щодо вимовленої послідовності слів, додатково після виділення поточних параметрів суцільно вимовлену фразу, що розпізнається, розбивають на ділянки, які є однорідними за значенням поточних параметрів, кодуєть кожен з однорідних ділянок, слова словника представляють як послідовність фонем, рекурентне зіставлення ділянки і мовлення суцільно вимовленої фрази з усіма словами словника й обчислення інтегральних ступенів близькості та меж слова виконують пофонемно для кожного слова й паралельно за всіма словами словника, формують списки послідовностей слів, для кожної сформованої послідовності слів підсумовують інтегральні ступені близькості слів, що її складають, прийняття рішення щодо вимовленої послідовності слів суцільно вимовленої фрази виконують за мінімальним значенням суми ступенів близькості.

При цьому пофонемно для кожного слова й паралельно за всіма словами словника зіставлення участку мовлення суцільно вимовленої фрази з усіма словами словника, обчислення інтегральних ступенів близькості та меж слова виконують шляхом рекурентного зіставлення поточних параметрів однорідних участків з еталонами поточної та наступної фонемі від першої до передостанньої фонемі всіх слів, що розглядаються, знаходять момент закінчення поточної фонемі та обчислюють ступінь близькості від поточної фонемі до участку мовлення, що їй відповідає, моментом початку наступної фонемі вважають момент, що є наступним за моментом закінчення поточної фонемі, для останньої фонемі слова наступними вважають перші фонемі всіх слів словника, підсумовують ступені близькості між фонемами слова та відповідними участками мовлення, запам'ятовують ступінь близькості та межі слова, при цьому лівою межею слова вважають порядковий номер першого поточного параметра, який віднесено до першої фонемі слова, правою межею слова вважають порядковий номер поточного параметра, що відповідає моменту закінчення останньої фонемі слова.

Отже, сукупність наведених відмітних ознак підвищує швидкодію розпізнавання фраз, що вимовляються суцільно й складаються зі слів великого словника, і дозволяє здійснювати розпізнавання у режимі реального часу.

Особливості способу, що заявляється, полягають у наступному.

Для виділення поточних параметрів мовний сигнал, що є оцифрованим як 8-бітний із частотою дискретизації 22050Гц, підлягає перетворенню Хартлі з вікном 20мс та кроком 10мс. Поточні параметри мовленнєвого сигналу отримують логарифмуванням відношень енергій спектральних полос до загальної спектральної енергії вікна та є послідовністю векторів поточних параметрів R (або реалізацією R).

$$R = r_1, r_2, \dots, r_i, \dots, r_N \quad (1)$$

де R - послідовність векторів поточних параметрів,  $r_i$  - i-й поточний параметр.

Для кожного вікна обчислюється значення енергії мовленого сигналу.

Для розбиття на ділянки суцільно вимовленої фрази на ділянки, що є однорідними за значенням поточних параметрів, знаходять середнє значення відстані між сусідніми векторами поточних параметрів. Величину, що дорівнює середньому значенню, вважають пороговою. Послідовності поточних параметрів, таких, де відстань між сусідніми поточними параметрами не перевищує порогового значення, виділяють як однорідні за значенням поточних параметрів. Множина таких послідовностей поточних параметрів є множиною S однорідних за значенням поточних параметрів ділянок. Для сегментів S обчислюють середні значення поточних параметрів X.

$$S = s_1, s_2, \dots, s_i, \dots, s_y, \quad (2)$$

$$X = x_1, x_2, \dots, x_i, \dots, x_y$$

де S - множина однорідних ділянок,  $s_i$  - i-й однорідна ділянка, X - множина середніх значень поточних параметрів однорідних ділянок  $x_i$  - середнє значення поточних параметрів i-ї однорідної ділянки, y - кількість однорідних ділянок.

Кожну з однорідних ділянок кодуєть залежно до відносного значення середньої енергії однорідної ділянки в межах мовленого сигналу, розкиду енергії в межах однорідної ділянки, характеру зміни енергії всередині однорідної ділянки й значення цих же величин сусідніх однорідних ділянок. Перелік кодів та їх значення: (1 -

зростання, 2 - зростання з малим розкидом, 3 - зростання з максимальним значенням, 4 - зменшення, 5 - зменшення з малим розкидом, 6 - зменшення з мінімальним значенням, 7 - сталість, 8 - сталість із максимальним значенням, 9 - сталість з мінімальним значенням).

Слова словника представляють послідовністю фонем. Словник заданий користувачем у вигляді тексту. Він зберігає графічне представлення (написання) слів. За написанням слів будують їх фонетичну транскрипцію. Перелік транскрипційних символів, які позначають фонему, називають алфавітом фонем  $P$ . Кожній фонемі відповідає не порожня множина поточних параметрів, що зберігаються в кодовій книзі  $V$  та є еталонами цієї фонему.

$$\begin{aligned} P &= (C_1, C_2, \dots, C_i, \dots, C_K) \\ V &= \{b_1, b_2, \dots, b_i, \dots, b_K\}, \\ b_i &\subset V, \quad c_i \sim b_i. \end{aligned} \quad (3)$$

де  $P$  - алфавіт фонем;  $c_i$  -  $i$ -а фонема алфавіту фонем (або транскрипційний символ);  $V$  - кодова книга;  $b_i$  - множина поточних параметрів, що відповідають  $i$ -ій фонемі;  $V$  - множина поточних параметрів; знаком " $\sim$ " позначають відповідність символу алфавіту фонем множині елементів кодової книги;  $K$  - розмір алфавіту фонем.

Будь-яке слово задається своєю транскрипцією (послідовністю транскрипційних символів, що обзначають фонему)  $T$ .

$$\begin{aligned} T &= (c_1, c_2, \dots, c_k, \dots, c_L), \\ c_k &\in P, \end{aligned} \quad (4)$$

де  $k$  - номер фонему в слові;  $c_k$  -  $k$ -а фонема слова;  $L$  - довжина слова.

При пофонемному для кожного слова  $T$  і паралельному по всіх словах словника зіставленні ділянки мовлення суцільно вимовленої фрази  $R$  з усіма словами словника, обчислюють інтегральні ступені близькості  $D(T, R)$  та межі слова  $w_b, w_e$ .

Однією зі складових цієї процедури є виділення учасків, що відповідають окремим фонемам (сегментація на ділянки, що відповідають фонемам). Сегментація невідомого мовного повідомлення на ділянки, що відповідають фонемам, являє собою досить складну задачу, яка на сьогоднішній день не вирішена та активно розробляється вченими всього світу. У даному випадку пропонується перейти від задачі сегментації "у загальному вигляді" до її окремих рішень. Тобто, пропонується сегментувати мовлене повідомлення, виходячи з припущення про його фонетичний склад. Ступень близькості між словом  $T$  і ділянкою реалізації  $R$  дорівнює сумі ступенів близькості між фонемами слова й ділянками реалізації, що їм відповідають.

$$D(T, R) = \sum_{i=1}^L D(c_i, R[m_{bi}, m_{ei}]) \quad (5)$$

де  $L$  - довжина слова  $T$ ;  $c_i$  -  $i$ -а фонема слова;  $m_{bi}, m_{ei}$  - порядкові номери векторів реалізації, що відповідають початку та закінченню  $i$ -ої фонему;  $R[m_{bi}, m_{ei}]$  - участок реалізації, що відповідає межах  $i$ -ої фонему.

Таким чином, задачі визначення меж слова та знаходження ступеня близькості між словом й ділянкою реалізації, що відповідає його межах, переростає в задачі визначення меж фонем та знаходження ступеня близькості між фонемами слова й ділянками реалізації, що їм відповідають.

Для знаходження меж фонему виконують зіставлення середніх значень поточних параметрів  $X_i$  однорідних учасків 5 з еталонами поточної ( $c_k$ ) та наступної ( $c_{k+1}$ ) фонему.

Ступенем близькості  $D(c, x_i)$  між фонемою  $c$  та середнім значенням поточних параметрів  $x_i$  вважають мінімум Евклідової відстані між значенням  $x_i$  та всіма еталонами фонему.

$$D(c, x_i) = \min_j D(b_j, x_i) \quad (6)$$

де  $b_j$  -  $j$ -ий еталон фонему  $c, b_j \in V, x_i \in V$ .

Кожній фонемі має відповідати не менше одної однорідної ділянки з множини 5'. Однорідна ділянка або послідовність однорідних ділянок, що відповідає голосній фонемі, повинні включати ділянку зростання енергії, максимум енергії та ділянку зменшення енергії. Однорідна ділянка або послідовність однорідних ділянок, що відповідає сонорній приголосній та є сусіднім до голосної, має у порівнянні з голосною меншу середню енергію й менший розкид енергії.

Доки значення  $D(c_k, x_i)$  ступеня близькості середнього значення поточних параметрів  $x_i$  однорідних ділянок, що розглядаються послідовно, і поточної фонему  $c_k$  не перевищує  $D(c_{k+1}, x_i)$  ступенів близькості середнього значення поточних параметрів  $x_i$  однорідних ділянок, що розглядаються послідовно, і наступної фонему  $c_{k+1}$ , сегменти відносять до поточної фонему. При досягненні однорідної ділянки, для якої виконується умова

$$D(c_{k+1}, x_i) < D(c_k, x_i) \quad (7)$$

фіксують момент закінчення поточної фонему  $m_{ek}$ , як номер останнього поточного параметра, що віднесено до однорідної ділянки  $i-1$ , і момент початку наступної фонему  $m_{bk+1}$ , як номер першого поточного параметра, що віднесено до сегменту  $i$ .

Ступінь близькості  $D(c_k, R[m_{bk}, m_{ek}])$  між фонемою  $c_k$  та ділянкою реалізації, що відповідає її межах  $R[m_{bk}, m_{ek}]$ , визначають як суму ступенів близькості між фонемою  $c_k$  та всіма поточними параметрами  $r_i$  цієї ділянки.

$$D(c_k, R[m_{bk}, m_{ek}]) = \sum_{i=m_{bk}}^{m_{ek}} D(c_k, r_i) \quad (8)$$

Ступінь близькості  $D(c, r_i)$  між фонемою  $c$  та поточним параметром  $r_i$  визначають як мінімум Евклідової відстані між значенням  $r_i$  та всіма еталонами фонему.

$$D(c, r_i) = \min_j D(b_j, r_i) \quad (9)$$

де  $b_{j-j}$ -ий еталон фонем  $c, b_j \in V, r_i \in V$ .

Ці операції виконують рекурентне, у результаті чого отримують межі слова ( $w_b$  - ліва межа слова і  $w_e$  - права межа слова) та ступінь близькості  $D(T, R)$  від слова до реалізації.

$$W_b = m_{b1} \quad (10)$$

$$w_e = m_{eL}$$

де -  $w_b$  - ліва межа слова,  $w_e$  - права межа слова,  $m_{b1}$  - ліва межа першої фонемі слова,  $m_{eL}$  - права межа останньої фонемі слова.

$$D(T, R) = \sum_{k=1}^L \sum_{i=m_{bk}}^{m_{ek}} D(c_k, r_i) = \sum_{k=1}^L \sum_{i=m_{bk}}^{m_{ek}} \min D(b_{kj}, r_i) \quad (11)$$

де  $T$  - транскрипція слова,  $L$  - кількість фонем в слові  $T$ ,  $R$  - послідовність поточних параметрів або реалізація,  $r_i$  -  $i$ -й поточний параметр реалізації,  $k$  - номер фонемі,  $c_k$  -  $k$ -а фонема слова,  $m_{bk}$  - ліва межа  $k$ -ої фонемі,  $m_{eL}$  - права межа  $k$ -ої фонемі,  $b_{kj}$  -  $j$ -й еталонний вектор фонемі  $c_k$ .

Таким чином, лівою межею слова вважають порядковий номер першого поточного параметра, який віднесено до першої фонемі слова, правою межею слова вважають порядковий номер поточного параметра, що відповідає моменту закінчення останньої фонемі слова. Ступенем близькості між словом та ділянкою реалізації, що відповідає його межах, є сума ступенів близькості від фонем слова до ділянок реалізації, що відповідають межах цих фонем.

Для забезпечення паралельного (одночасного) доступу до перших, других, третіх і т.д. фонем усіх слів, словник розпізнавання організовано у вигляді дерева транскрипцій, такого, де вершини висоти 1 відповідають першим фонемам слів, вершини висоти 2 - другим фонемам слів. Наприклад, якщо слова словника починаються тільки з фонем: а, б, в, н, с, то дерево буде мати 5 вершин висоти 1. Вершини висоти 2, що є дочірніми до вершини, яка відповідає фонемі а, описують другі фонемі слів, що починаються з цієї фонемі і т.д. У випадку, якщо вершина відповідає останній фонемі слова, вона зберігає ненульовий номер слова. Ця вершина також може мати потемків (дочірні вершини).

Організація словника, що наведено, дозволяє виконувати пофонемне зіставлення ділянки мовлення суцільно вимовленої фрази з усіма словами словника паралельно. Спочатку знаходять межі усіх фонем всіх слів і ступені їх близькості до відповідної ділянки реалізації, далі знаходять межі других фонем, третіх і т.д. Ця операція виконується рекурентне для всіх фонем. На кожному кроці рекурсії фонемі, що є найбільш віддаленими від реалізації та слова, які містять їх як фонему, що розглядається на даному кроці рекурсії, із подальшого розглядання виключаються.

При досягненні вершини, що відповідає закінченню слова, номер цього слова, його межі та ступінь близькості цього слова до реалізації зберігають у списках послідовностей слів. Для кожної сформованої послідовності слів підсумовують інтегральні ступені близькості слів, що її складають, прийняття рішення щодо вимовленої послідовності слів суцільно вимовленої фрази виконують за мінімальним значенням суми.

Даний спосіб реалізовано за допомогою комп'ютерної системи, що складається з ПК Pentium III з процесором не нижче 1ГГц й оперативною пам'яттю 256Мб, що обладнано звуковою картою й мікрофоном і програми, розробленої автором.