



УКРАЇНА

(19) **UA** (11) **145601** (13) **U**
(51) МПК (2021.01)
G10L 17/00
G10L 17/02 (2013.01)

НАЦІОНАЛЬНИЙ ОРГАН
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ
ДЕРЖАВНЕ ПІДПРИЄМСТВО
"УКРАЇНСЬКИЙ ІНСТИТУТ
ІНТЕЛЕКТУАЛЬНОЇ
ВЛАСНОСТІ"

(12) ОПИС ДО ПАТЕНТУ НА КОРИСНУ МОДЕЛЬ

<p>(21) Номер заявки: u 2020 04344</p> <p>(22) Дата подання заявки: 13.07.2020</p> <p>(24) Дата, з якої є чинними права інтелектуальної власності: 29.12.2020</p> <p>(46) Публікація відомостей про державну реєстрацію: 28.12.2020, Бюл.№ 24</p>	<p>(72) Винахідник(и): Роботягов Андрій Валентинович (UA), Копоть Михайло Андрійович (UA), Колесников Дмитро Олегович (UA), Груздо Ірина Володимирівна (UA)</p> <p>(73) Володілець (володільці): ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ, пр. Науки, 14, м. Харків, 61166 (UA)</p>
---	--

(54) СПОСІБ ІДЕНТИФІКАЦІЇ ЛЮДИНИ НА ОСНОВІ АНАЛІЗУ ЕЛЕМЕНТАРНИХ СЕГМЕНТІВ ЗВУКІВ МОВИ

(57) Реферат:

Спосіб ідентифікації людини на основі аналізу елементарних сегментів звуків мови включає створення банку еталонних мовних об'єктів, причому кожний еталонний мовний об'єкт утворюється множиною елементарних сегментів (ЕС) мови людини, перетворення мовного сигналу у цифрову форму, ідентифікацію невідомого об'єкта з множиною еталонних об'єктів виконують за класичною схемою ідентифікації - "один з множиною" на основі евклідової відстані. При цьому у мовному сигналі в автоматичному режимі на ділянках з інтервалом в 1500 семплів за оптимальним критерієм виділяють три елементарних сегменти ЕС, кожен сегмент є структурою, яку утворюють три суміжні півхвилі мовного сигналу, структуру кожного ЕС з трьох суміжних півхвиль мовного сигналу відображають чотири суміжні точки, інформативними ознаками є значення тривалостей ЕС, мовний сигнал в автоматичному режимі поділяють на ділянки з інтервалом в 1500 семплів за правилом так, що початок наступної ділянки є кінець останньої півхвилі трьох ЕС, які знайдені за оптимальним критерієм, формується множина значень тривалостей ЕС порядку 100-2000 значень відповідно до тривалості мовного об'єкта, тривалість мовних об'єктів, що порівнюється, може бути різною.

UA 145601 U

Корисна модель призначена для проведення фоноскопичної ідентифікації людини за параметрами мовного сигналу - на основі аналізу тривалостей елементарних сегментів звуків мови.

Є відомим спосіб "Система і метод ідентифікації мовника за фонограмою спонтанної усної мови на основі вирівнювання (корекції) формант однакових голосних фонем" (Патент США № US 9047866 B2, МПК G10L 15/00, 17/14, опубл. 02.06.2015). У способі подібність між першою фонограмою мовника і другою (еталонною) або вибірковою фонограмою оцінюється шляхом зіставлення формантних частот в референтних висловлюваннях мовного сигналу, де висловлювання для порівняння вибираються з першої фонограми і другої (еталонної) фонограми. Вибрані для порівняння референтні висловлювання включають формантні шляхи з щонайменш трьох формантних частот, де перші дві форманти знаходяться в типових межах варіативності для одного типу фонем. Вибрані висловлювання мовника, які включають щонайменше дві ідентичні частоти формант, порівнюються між собою. Оцінка подібності порівнюваних референтних висловлювань проводиться за зіставленням формантних частот; подібність фонограм визначається за оцінкою подібності всіх порівнюваних референтних висловлювань.

До недоліків способу, що приводять до обмеження його застосування, належать: спектральні ознаки чуттєві до змін емоційного стану людини та умов запису і до мовного матеріалу пред'являють вимогу фонетичної порівнянності.

Є відомим "Спосіб аутентифікації диктора за парольною фразою" (Патент РФ № 2422921, МПК G10L 15/00, 17/00, опубл. 27.06.2011). Спосіб аутентифікації диктора по вхідному мовному сигналу диктора у вигляді парольної фрази включає порівняння параметрів цього сигналу із заданою точністю зі заздалегідь збереженими зразками параметрів вхідних мовних сигналів у вигляді тієї ж парольної фрази, яка промовлена заздалегідь відомими дикторами, з подальшою аутентифікацією, як інформативні параметри є низькочастотна частина вейвлет перетворення від нормованої функції розподілу особливих точок уздовж звукового файлу, що відповідає вхідному мовному сигналу диктора у вигляді парольної фрази, виділених шляхом порівняння відліку в цій точці у звуковому файлі з попередніми та наступними відліками за допомогою узагальнених коефіцієнтів лінійного передбачення і порога T , при цьому нормування функції розподілу зводиться до приведення її до стандартної довжини L_{ep} , прийнятої під час підрахунку еталонних параметрів вхідних мовних сигналів у вигляді парольної фрази, яка вимовлена заздалегідь відомими дикторами.

До недоліків способу, що приводять до обмеження його застосування, належать: високе значення ймовірності похибки другого роду (невірна аутентифікація диктора), яка дорівнює 0,26, до мовного матеріалу пред'являють вимогу фонетичної порівнянності.

Є відомим "Спосіб контактнорізничевої акустичної ідентифікації особи" (Патент РФ № 2451346, МПК G10L 15/00, 17/00, опубл. 20.05.2012). Спосіб акустичної ідентифікації особи, в якому акустичний сигнал, що вимірюється на тілі людини при проголошенні нею звуків, вводять в електронно-обчислювальний пристрій, визначають значення параметрів акустичного сигналу, визначають значення статистичних характеристик параметрів акустичного сигналу і формують на їх основі еталон (еталони), визначають ступінь відмінності між акустичним сигналом і еталоном, на підставі ступеня відмінності приймають рішення про приналежність акустичного сигналу людині, чиї значення статистичних характеристик були використані при формуванні еталона, як інформативні акустичні параметри використовується амплітудно-частотна характеристика тіла людини, при цьому для її розрахунку крім вхідного акустичного сигналу, що вимірюється на тілі людини, вводять вхідний мовний сигнал, а міру відмінності для порівняння біометричних параметрів ідентифікованої особи і еталона визначають як зважену Евклідову нев'язку параметрів амплітудно-частотної характеристики тіла ідентифікованої людини і еталона.

До недоліків способу, що приводять до обмеження галузі його застосування, належать: помилки, що виникають при визначенні меж пошуку (приблизної смуги) частот формант; помилки, які присутні при визначенні глобальних максимумів спектра потужності в приблизної смугі частот, що обумовлено наявністю локальних максимумів і мінімумів, які ускладнюють ефективну оцінку формантних частот. Одним із суттєвих недоліків є складність збереження в таємниці мовного сигналу як біометричного образу, а також низький ступінь захисту від імітації голосу за допомогою різних звуковідтворюючих пристроїв.

Найбільш близьким за функціональним призначенням і загальними суттєвими ознаками є спосіб ідентифікації людини на основі аналізу структурних ознак елементарних сегментів звуків мови (патент України № 76770, МПК (2006) G10L 15/00, 17/00, опубл. 15.09.2006 у Бюл. № 9). Суть способу полягає у аналізі структурних ознак релевантних ділянок цифрових мовних

сигналів, елементарних сегментів звуків мови (ЕС), формальний аналіз і вимір яких стали можливими завдяки уявленню об'єкта дослідження, елементарного сегмента, у вигляді деякого геометричного об'єкта зі складною просторово-часовою організацією.

5 Докладніше спосіб ідентифікації людини на основі аналізу звуків мови включає перетворення мовного сигналу у цифрову форму:

- попередньо створюють банк еталонних мовних об'єктів, причому кожний еталонний мовний об'єкт містить в собі запис одного звуку мови, який утворюється множиною елементарних сегментів (ЕС) мови людини;

10 - формують початковий набір геометричних елементів аналітичної геометрії ЕС звуків мови і значень мір близькості, які визначають пороги зіставлення, а після перетворення мовного сигналу у цифрову форму;

- виконують класифікацію еталонних мовних об'єктів для визначення оптимальних наборів геометричних елементів аналітичної геометрії ЕС звуків мови і значень мір близькості;

15 - здійснюють ідентифікацію перетвореного мовного сигналу у цифрову форму та еталонних мовних об'єктів на основі наборів геометричних елементів аналітичної геометрії ЕС звуків мови і значень мір близькості за критерієм максимальної якості розпізнавання.

До недоліків способу, що приводять до обмеження галузі його застосування, належать: безпосередньому аналізу підпадає лише один звук мови; сегментацію звуків мови на ЕС виконують вручну заздалегідь.

20 Задачею корисної моделі є розширення функціональних можливостей способу ідентифікації людини на основі аналізу тривалостей елементарних сегментів звуків мови, підвищення якості та швидкості ідентифікації.

Поставлена задача вирішена наступним чином. У способі ідентифікації людини на основі аналізу елементарних сегментів звуків мови, що включає створення банку еталонних мовних об'єктів, причому кожний еталонний мовний об'єкт утворюється множиною елементарних сегментів (ЕС) мови людини, перетворення мовного сигналу у цифрову форму, ідентифікація невідомого об'єкта з множиною еталонних об'єктів виконують за класичною схемою ідентифікації - "один з множиною" на основі евклідової відстані, згідно з корисною моделлю, у мовному сигналі в автоматичному режимі на ділянках з інтервалом в 1500 семплів за оптимальним критерієм виділяють три елементарних сегменти ЕС, кожен сегмент є структурою, яку утворюють три суміжні півхвилі мовного сигналу, структуру кожного ЕС з трьох суміжних півхвиль мовного сигналу відображають чотири суміжні точки, інформативними ознаками є значення тривалостей ЕС, мовний сигнал в автоматичному режимі поділяють на ділянки з інтервалом в 1500 семплів за правилом так, що початок наступної ділянки є кінець останньої півхвилі трьох ЕС, які знайдені за оптимальним критерієм, формується множина значень тривалостей ЕС порядку 100-2000 значень відповідно до тривалості мовного об'єкта, тривалість мовних об'єктів, що порівнюється, може бути різною.

На фіг. 1 зображено фрагмент мовного об'єкта та інтервал L мовного об'єкта, який дорівнює 1500 семплів ($L=1500$).

40 На фіг. 2 зображено інформативні об'єкти (ЕС) та тривалості (Т) ЕС.

На фіг. 3 наведено геометричне відображення процесу ідентифікації об'єктів.

На фіг. 4 зображено пристрій для реалізації способу.

Розглянемо більш докладніше запропонований спосіб. Основними положеннями корисної моделі є:

45 1) спектральне перетворення мовного сигналу не застосовують;

2) інформативними ознаками є тривалості ЕС;

3) у мовному сигналі в автоматичному режимі на ділянках з інтервалом в 1500 семплів за оптимальним критерієм виділяють три ЕС, кожен сегмент є структурою, яку утворюють три суміжні півхвилі мовного сигналу, структуру кожного ЕС з трьох суміжних півхвиль мовного сигналу відображають чотири суміжні точки 1-2-3-4 (фіг. 2);

50 4) мовний сигнал в автоматичному режимі поділяють на ділянки з інтервалом в 1500 семплів за правилом: початок наступної ділянки є кінець останньої півхвилі трьох ЕС, які знайдені за оптимальним критерієм;

55 5) невідомий об'єкт порівнюють з множиною еталонних об'єктів за класичною схемою ідентифікації - "один з множиною";

6) до мовного матеріалу не пред'являють вимоги порівнянності, тобто мовні сигнали можуть бути отримані в різних умовах і різними пристроями запису; тривалість мовних об'єктів, що порівнюються, може бути різною.

60 У корисній моделі мовним сигналом (об'єктом) є цифровий мовний сигнал (об'єкт) з форматом: частота дискретизації - 22050 Гц, рівень квантування - 8 біт, канал запису - моно.

Новизну способу ідентифікації визначають тим, що:

1) для формального перетворення мовних об'єктів не використовують спектральний математичний апарат;

2) у мовному сигналі в автоматичному режимі на ділянках з інтервалом в 1500 семплів за оптимальним критерієм виділяються три ЕС, кожен сегмент є структурою, яку утворюють три суміжні півхвилі мовного сигналу, структуру кожного ЕС з трьох суміжних півхвиль мовного сигналу відображають чотири суміжні точки 1-2-3-4 (фіг. 2);

3) мовний сигнал в автоматичному режимі поділяється на ділянки з інтервалом в 1500 семплів за правилом: початок наступної ділянки є кінець останньої півхвилі трьох ЕС, які знайдені за оптимальним критерієм;

4) як інформативні ознаки використовуються значення тривалостей ЕС (фіг. 2). Зазначимо, що у найближчому аналозі використовуються значення площі та довжини хорди складної плоскої фігури ЕС як інформативні ознаки;

5) невідомий об'єкт порівнюють з множиною еталонних об'єктів за класичною схемою ідентифікації - "один з множиною";

6) до мовного матеріалу не пред'являють вимоги порівнянності, тобто мовні сигнали можуть бути отримані в різних умовах і різними пристроями запису; тривалість мовних об'єктів, що порівнюються, може бути різною; нема потреби в порівнянні мов, на яких розмовляють люди.

У способі, що заявляється, механізм ідентифікації складається з трьох етапів:

1) етапу оптимального пошуку ЕС;

2) етапу формування інформативних ознак - множини тривалостей ЕС;

3) етапу ідентифікації невідомого об'єкта і еталонних мовних об'єктів. Еталонний об'єкт - об'єкт, приналежність якого заздалегідь відома до відповідного класу еталонних об'єктів.

Невідомий об'єкт - об'єкт, приналежність якого заздалегідь невідома до будь-якого класу еталонних об'єктів.

Кожен еталонний й невідомий об'єкт є звуковим файлом.

Метою першого етапу, етапу оптимального пошуку ЕС, є пошук ЕС за оптимальним критерієм. Оптимальний пошук триває на довжині всього мовного об'єкта на інтервалі, який дорівнює 1500 семплів ($L=1500$, фіг. 1).

У мовному сигналі в автоматичному режимі на ділянках з інтервалом в 1500 семплів за оптимальним критерієм виділяються три ЕС, кожен сегмент є структурою, яку утворюють три суміжні півхвилі мовного сигналу, структуру ЕС з трьох суміжних півхвиль мовного сигналу відображають чотири суміжні точки 1-2-3-4, які зображені на фіг. 2. Наприклад, тривалість всього фрагмента мовного об'єкта, який зображено на фіг. 2, дорівнює 600 семплів.

Оптимальний пошук ЕС обчислюють за формулами (1-4). Пошук того стану, коли три структури є трьома ЕС, здійснюють за критерієм F_3 , який представлено формулою (4): три структури є трьома ЕС, коли значення критерію F_3 досягає максимуму. На інтервалі $L=1500$ виконується аналіз енергетичних характеристик трьох суміжних структур:

$$f_1(a, \Delta t) = \frac{1}{\Delta t} \sum_i (a_i - \min(a_i)) \quad (1)$$

$$F_1 = \sum_{n=1}^3 (f_1^n(a, \Delta t)) \quad (2)$$

$$F_2 = \sum_{i,j=1}^3 \left(\left| f_1^i(a, \Delta t) - f_1^j(a, \Delta t) \right| \right), i \neq j \quad (3)$$

$$F_3 = \frac{F_1}{F_2} \cdot 100 \quad (4)$$

де $f_1(a, \Delta t)$ - функція енергетичної характеристики елементарного коливання, Δt - довжина елементарного коливання $\Delta t = [1: 4]$, яке утворено чотирма суміжними точками 1-2-3-4; a_i - амплітуда сигналу (семпл), $a_i \cdot$ - \min амплітуда сигналу (семпл) на інтервалі Δt .

Мовний сигнал в автоматичному режимі поділяють на ділянки з інтервалом в 1500 семплів за правилом: початок наступної ділянки є кінець останньої півхвилі трьох ЕС, які знайдені за оптимальним критерієм.

Метою другого етапу, етапу формування інформативних ознак, є формування первинного опису ЕС у вигляді вектора V в n -мірному просторі ознак:

$$V = (v_1, v_2, \dots, v_n) \quad (5)$$

Інформативними ознаками є значення тривалостей ЕС (фіг. 2). Тривалість ЕС є відстань між двома суміжними структурами, які вибрані за критерієм (4). Відстань вимірюється як сума кількості семплів між точками 1 двох суміжних структур. Виконується автоматична обробка кожного мовного об'єкта процесу ідентифікації (еталонного й невідомого), у результаті якої:

5 1) вимірюють значення T тривалостей ЕС трьох структур трьох суміжних хвиль за формулою (6):

$$T = |p(1)_i - p(1)_{i+1}|, \quad (6)$$

де $p(1)$ - адреса точки 1 відповідного складу трьох суміжних хвиль кожної з трьох структур (фіг. 2);

2) кожен елемент вектора V (5) в n -мірному просторі ознак обчислюють за формулою (7):

$$V_i = |I_i - I_i^*|, \quad (7)$$

10 Елементи вектора V (5) ранжуються за правилом: перший елемент - мінімальне значення, останній - максимальне.

Вектор V (5) утворює первинний образ звуку мови i , відповідно, мовного об'єкта та відображає індивідуальні мовні особливості людини.

15 Метою третього етапу, етапу ідентифікації невідомого об'єкта і еталонних мовних об'єктів, є проведення ідентифікації невідомого об'єкта з множиною еталонних об'єктів за правилом "один з множиною" за однією системою ознак. Опис невідомого (вектор V) та еталонного (вектор W) об'єктів є аналогічним зазначеним вище процедурам.

Критерієм порівняння двох порівнюваних об'єктів є обчислена між ними відстань r . Відстань обчислюється в метричному евклідовому просторі R^n за формулою (8):

$$r = r(O_1, O_2) = \sqrt{\sum_i (v_i - w_i)^2}, \quad (8)$$

20 де O - об'єкт розпізнавання, V_i, W_i - i -ті ознаки (елементи векторів V та W), n - кількість ознак.

Значення усіх відстаней між порівнюваними об'єктами зберігають у спеціальній таблиці результатів ідентифікації - один стовпець таблиці зберігає результати ідентифікації невідомого об'єкта і множини еталонних об'єктів за однією системою ознак.

25 Невідомий об'єкт вважають тотожним еталонному об'єкту, тобто приналежним до якого-небудь одного визначеного класу еталонних об'єктів банку, якщо значення критерію (8) досягає мінімуму:

30 Оскільки кількість ознак n двох порівнюваних об'єктів, як правило, не є рівною, то відстань r обчислюється як оптимальне значення для двох об'єктів за правилом: вектор з меншою кількістю ознак проходить з шагом один по всьому вектору з більшою кількістю ознак для визначення оптимального значення r . На фіг. 3 наведено геометричне відображення процесу ідентифікації об'єктів.

35 Якість ідентифікації за новим способом ідентифікації людини на основі аналізу тривалостей ЕС дорівнює 80-98 %, перевірка результатів ідентифікації виконувалась на множені з 228 еталонних об'єктів та 1 невідомим об'єктом, час ідентифікації дорівнює 30 с

40 Запропонований спосіб може бути реалізований, наприклад, за допомогою технічного пристрою, блок-схему якого наведено на кресленні (фіг. 4). До складу пристрою входять 6 функціональних елементів: 1 - блок керування; 2 - банк еталонних мовних об'єктів; 3 - банк невідомих мовних об'єктів; 4 - блок оптимального пошуку ЕС; 5 - блок ідентифікації мовних об'єктів; 6 - блок аналізу результатів ідентифікації.

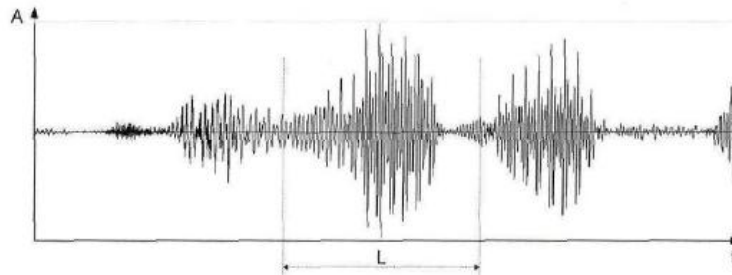
45 З зовнішнього сигналу "Запуск" починається робота пристрою. З його приходом блок керування пристроєм 1 генерує спеціальні сигнали, які керують роботою всього пристрою. Зв'язок між функціональними елементами пристрою здійснюють за допомогою двонаправленої шини даних і керування. На початковому етапі блок керування 1 активує банк еталонних мовних об'єктів, банк невідомих мовних об'єктів. Банк еталонних об'єктів 2 формує таблицю всіх мовних еталонних об'єктів. Банк невідомих об'єктів 3 формує таблицю всіх невідомих мовних об'єктів, відносно яких виконують ідентифікацію. Усі мовні об'єкти зберігають в електронній пам'яті пристрою у вигляді цифрових файлів. Після виконання цих процедур банк еталонних мовних об'єктів та банк невідомих мовних об'єктів генерують керуючі сигнали, які повідомляють блок керування про готовність до подальшої роботи. Блок керування 1 активує роботу блока оптимального пошуку ЕС 4, в якому виконується пошук ЕС на довжині всього мовного об'єкта на інтервалах, які дорівнюють 1500 семплів, та опис об'єктів у вигляді інформативних ознак, якими є значення тривалостей ЕС. Після завершення процесу пошуку ЕС й опису об'єктів у блоці

ідентифікації мовних об'єктів 5 виконується попарне порівняння невідомого об'єкта з множиною еталонних об'єктів за правилом "один з множиною" за однією системою ознак. Результати ідентифікації аналізуються й зберігаються у блоці аналізу результатів ідентифікації 6 невідомого і еталонних мовних об'єктів. Після того, як виконано ідентифікацію невідомого об'єкта і множини еталонних об'єктів, блок аналізу результатів ідентифікації 6 інформує блок керування 1 про завершення ідентифікації за допомогою інформаційно-керуючого сигналу "Ідентифікація завершена". Коли блоком керування 1 прийнято сигнал про завершення ідентифікації, блок керування 1 активує блок аналізу результатів ідентифікації 6, який видає користувачу звіт про підсумкові результати ідентифікації у вигляді спеціального протоколу. Для відображення об'єктивної сторони процесу ідентифікації в пристрої передбачено перегляд інформації, яка сформована на проміжних етапах його роботи. Перегляд цієї інформації ініціалізується зовнішнім сигналом "Інформація".

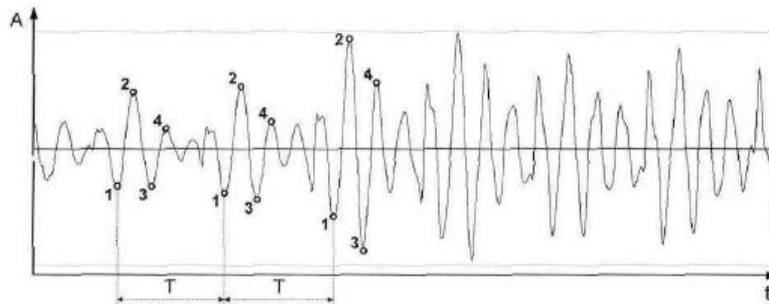
Таким чином, досягнуто вирішення задачі корисної моделі, а саме, підвищення якості та швидкості ідентифікації людини на основі аналізу тривалостей елементарних сегментів звуків мови.

ФОРМУЛА КОРИСНОЇ МОДЕЛІ

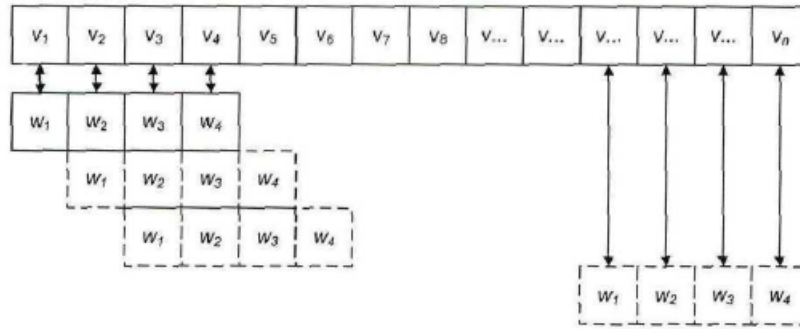
Спосіб ідентифікації людини на основі аналізу елементарних сегментів звуків мови, що включає створення банку еталонних мовних об'єктів, причому кожний еталонний мовний об'єкт утворюється множиною елементарних сегментів (ЕС) мови людини, перетворення мовного сигналу у цифрову форму, ідентифікацію невідомого об'єкта з множиною еталонних об'єктів виконують за класичною схемою ідентифікації - "один з множиною" на основі евклідової відстані, який **відрізняється** тим, що у мовному сигналі в автоматичному режимі на ділянках з інтервалом в 1500 семплів за оптимальним критерієм виділяють три елементарних сегменти ЕС, кожен сегмент є структурою, яку утворюють три суміжні півхвилі мовного сигналу, структуру кожного ЕС з трьох суміжних півхвиль мовного сигналу відображають чотири суміжні точки, інформативними ознаками є значення тривалостей ЕС, мовний сигнал в автоматичному режимі поділяють на ділянки з інтервалом в 1500 семплів за правилом так, що початок наступної ділянки є кінець останньої півхвилі трьох ЕС, які знайдені за оптимальним критерієм, формується множина значень тривалостей ЕС порядку 100-2000 значень відповідно до тривалості мовного об'єкта, тривалість мовних об'єктів, що порівнюється, може бути різною.



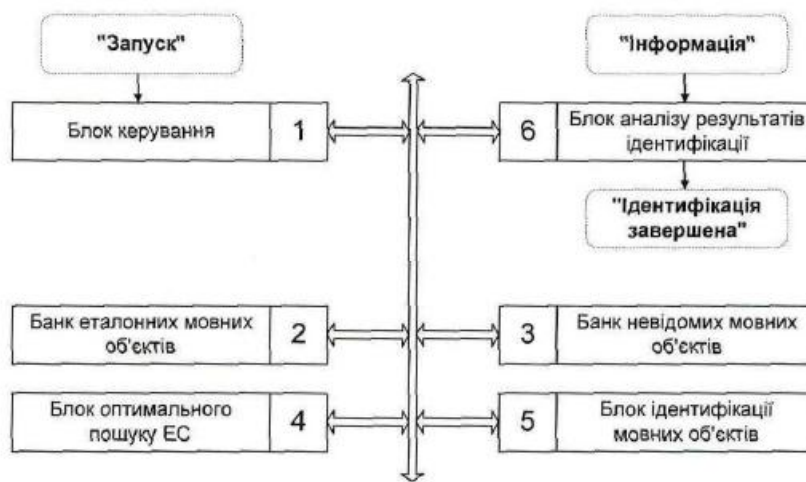
Фіг. 1



Фіг. 2



Фіг. 3



Фіг. 4