

Винахід відноситься до аналізу мовних сигналів з метою ідентифікації диктора і може використовуватись в судових експертизах матеріалів звукозапису.

Відомі способи ідентифікації особи за параметрами усного мовлення, які базуються на порівнянні спектру мовного сигналу і частоти основного тону (Дж. Р. Додінгтон. Распознавание дикторов. Идентификация людей по голосу // ТИИЭР. - 1985. - Т.73. - №11. - С.137 - 140). Але серед ознак, які використовують ці способи, немає параметрів мови, що відображають динаміку мовотворення.

Для зменшення часу аналізу проводять автоматичну сегментацію мовного сигналу на власно мову й паузи по середньо квадратичному значенню сигналу на ділянці, що аналізується. З обраних таким чином ділянок мови в автоматичному режимі за допомогою алгоритму Нола відбирають вокалізовані ділянки, на яких акустичні ознаки більш стабільні (А.М. Noll. Cepstrum pitch determination // The Journal of the Acoustical Society of America. - 1967. - Vol.41. - №2. - P.293 - 309). З вокалізованих ділянок вилучають, параметри мови, а отримані значення усереднюють. З метою спрощення використовуємої апаратури спектральний аналіз і вилучення параметрів з мовного сигналу проводиться не аналоговими, а цифровими методами. Для отримання огинаючої спектру мови застосовують гомоморфну обробку мови.

Найближчим до запропонованого є обраний за прототип спосіб ідентифікації, заснований на визначенні дзвінких ділянок мовного сигналу, вилученні на них спектральної огинаючої мови у декількох спектральних смугах в області вищих формант (2 - 4кГц) з усередненням її у часі (Авт. св. СРСР №518791, кл. G10L1/00). Крім спектральної огинаючої визначається частота основного тону і темп мови. Параметри, які отримані із зразкової і тестової фонограми, порівнюються.

Даний спосіб має суттєві вади. Спектральна огинаюча визначається за допомогою гребінки аналогових фільтрів. Рішення про вокалізованість (дзвінкість) ділянки мови приймається по її енергії. Темп мови вельми залежить від емоційного стану диктора і не є стійкою ознакою. Ці недоліки знижують імовірність правильної ідентифікації. Вони усуваються в запропонованому способі, заснованому на порівнянні вилученого з мовного сигналу комплексу параметрів.

Усунення вказаних недоліків досягається тим, що з метою збільшення ймовірності правильної ідентифікації з мовного сигналу вилучають і застосовують як ідентифікаційну ознаку усереднені у часі значення зміни частоти основного тону. З метою врахування динаміки мовотворення з мовного сигналу вилучають і застосовують як ідентифікаційну ознаку усереднені у часі значення функції артикуляції. Для отримання більш гладкої огинаючої спектру мови застосовують гомоморфну обробку мови із звуженим кепстральним часовим вікном. З метою врахування геометрії мовного тракту з мовного сигналу вилучають і застосовують як ідентифікаційну ознаку усереднені у часі значення нахилу огинаючої спектру мови до верхніх частот. З метою врахування значимості параметрів мови для ідентифікації при порівнянні векторів параметрів, вилучених з мовного сигналу, використовують евклідову відстань із зваженням параметрів за допомогою F-відношення. З метою врахування якості сигналу при відсутності або значному спотворенні завадами корисного сигналу в частотній смузі при обчисленні нахилу огинаючої спектру мови до верхніх частот, функції артикуляції і евклідової відстані відповідні спектральні відліки не враховують.

Поставлена мета досягається в запропонованому способі, який включає 4 етапи:

1) Попередня обробка фонограм (покращання якості записів, усунення шумів та завад, відбір мовного матеріалу достатньої якості і кількості), перетворення мовного сигналу у цифрову форму (його оцифровування). Попередня обробка записів можлива до оцифровування (в аналоговому виді) або після оцифровування (тобто в цифровому виді).

2) Обробка оцифрованих мовних сигналів для виміру комплексу параметрів.

3) Порівняння мови на зразковій і тестовій фонограмах на основі вимірюваного комплексу параметрів з використанням міри близькості - зваженої евклідової відстані між векторами параметрів.

4) Прийняття рішення про ідентичність мови на зразковій і тестовій фонограмах.

З мовного сигналу вилучають 4 групи параметрів: спектральні відліки, кепстральні відліки, кепстральні коефіцієнти, група неоднорідних параметрів (частоти основного тону, перших трьох формант, антиформанти, відношення частоти першої форманти до частоти основного тону, відношення частот другої і першої формант, нахил огинаючої спектру мови до верхніх частот, зміна частоти основного тону, функція артикуляції).

Для отримання більш гладкої огинаючої спектру мови застосовано гомоморфну обробку мови із звуженим кепстральним часовим вікном (Гусєв К. і Новосельський О. Гомоморфна обробка мови з використанням зменшеного часового вікна // Матеріали 3 - ої Міжнародної всеукраїнської конференції "Оброблення сигналів і зображень та розпізнавання образів". - К., 1996. - С.123 - 124). Отриману огинаючу спектру мови застосовано для оцінки спектру мови, формантних частот, нахилу спектру мови до верхніх частот.

Нахил огинаючої спектру мови до верхніх частот аналізується через обчислення по методу найменших квадратів коефіцієнта к прямої  $y = kx + b$ , яка апроксимує згладжений гомоморфною обробкою спектр мови у смузі частот до 4000Гц.

Функція артикуляції  $M(t)$  обчислюється на ділянках мови за формулою (Тимошенко П.Ю., Салтевський М.В., Жаріков Ю.Ф. Теория и практика использования следов памяти (идеальных отображений) в расследовании преступлений. - К.: Украинская академия внутренних дел, 1991. - С.40).

$$M(i\Delta t) = \frac{1}{n_j} \sum_{j=1}^n [lgS(j, i\Delta t) - lgS(j, (i-1)\Delta t)] ,$$

де  $\Delta t$  - відрізок часу між суміжними спектральними зрізами;

S - миттєвий логарифмічний спектр мови;

i - дискретні відліки часу;

j - номер спектрального відліку ( $j = 1, 2, \dots, n$ ).

Усереднене значення зміни частоти основного тону обчислюється як середня різниця між значеннями частоти основного тону на сусідніх вокалізованих ділянках.

Кепстральні коефіцієнти отримуються з коефіцієнтів лінійного передбачення (КЛП). КЛП обчислюються з використанням асинхронного з основним тоном автокореляційного методу (Маркел Дж. Д., Грей А.Х. Линейное предсказание речи. - М.: Связь, 1980. - 308с.).

Після виміру комплексу параметрів всі 4 групи параметрів об'єднуються в один вектор. Потім обчислюється зважена евклідова відстань ED між векторами параметрів, вилучених із зразкової та тестової фонограми. Голоси вважаються ідентичними, якщо евклідова відстань виявляється менше встановленого порога.

$$ED = \sqrt{\sum_{i=1}^m \frac{(x_i - e_i)^2}{D_i}} F_i,$$

де  $e_i$  - компоненти вектора-зразка ( $i = 1, 2, \dots, m$ ),

$x_i$  - компоненти тестового вектора,

$m$  - розмірність вектора-зразка і тестового вектора,

$D_i$  - дисперсія  $i$ -ої компоненти вектора-зразка  $e_i$ ,

$F_i$  -  $F$ -відношення  $i$ -ої компоненти.

$F$  - відношення параметра мови - це відношення факторної дисперсії  $S^2_{\text{факт}}$  до залишкової дисперсії  $S^2_{\text{зал}}$

$$F = S^2_{\text{факт}} / S^2_{\text{зал}}.$$

Факторна дисперсія  $S^2_{\text{факт}}$  характеризує розсіювання параметра мови між дикторами еталонної множини, залишкова дисперсія  $S^2_{\text{зал}}$  характеризує розсіювання для одного диктора (Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. - М.: Высш. шк., 1979. - С.283):

$$S^2_{\text{факт}} = \frac{n}{m-1} \sum_{j=1}^m (\mu_j - \bar{\mu})^2;$$

$$S^2_{\text{зал}} = \frac{1}{m(n-1)} \sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \mu_j)^2,$$

де  $i$  - номер запису ( $i = 1, \dots, n$ );

$j$  - номер диктора ( $j = 1, \dots, m$ );

$x_{ij}$  - величина параметру в  $i$ -му записі  $j$ -го диктора;

$$\mu_j = \frac{1}{n} \sum_{i=1}^n x_{ij} - \text{середнє для } j\text{-го диктора};$$

$$\bar{\mu} = \frac{1}{m} \sum_{j=1}^m \mu_j - \text{загальне середнє}.$$

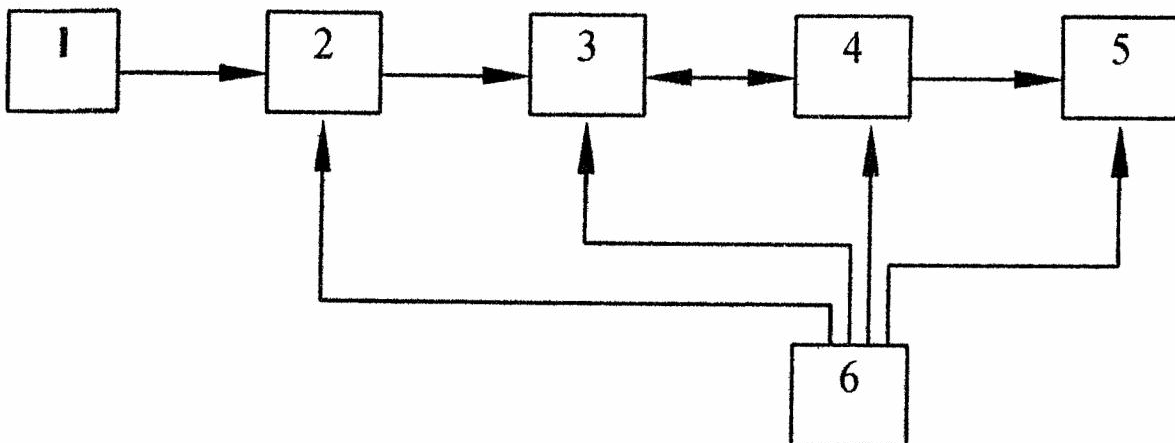
$F$ -відношення для параметрів мови обчислюється по заздалегідь зібраній базі даних мовних сигналів. Чим більше його величина для даного параметру мови, тим більш даний параметр придатний для розрізнення дикторів. Через застосування у обчисленні евклідової відстані  $F$ -відношення враховується значимість параметрів мови для ідентифікації.

Якщо в зразковій або тестовій фонограмі корисний сигнал в якійсь частотній смузі відсутній або спотворений значними завадами, при обчисленні нахилу огинаючої спектру мови до верхніх частот, функції артикуляції і евклідової відстані відповідні спектральні відліки не враховують.

Спосіб може бути реалізовано, наприклад, за допомогою пристрою, блок-схему якого наведено на кресленні (фіг.). Пристрій складається із фільтру низьких частот (ФНЧ) 1, аналого-цифрового перетворювача (АЦП) 2, оперативного запам'ятовуючого пристрою (ОЗП) 3, арифметично-логічного пристрою (АЛП) 4, блоку відображення інформації (БВІ) 5, блоку управління (БУ) 6.

Пристрій, працює наступним чином.

Мовний сигнал подається на ФНЧ, де його спектр обмежується для усунення ефекту накладання спектрів, і перетворюється у цифрову форму за допомогою АЦП. Оцифрований мовний сигнал зберігається у ОЗП. У ОЗП також зберігається програма (алгоритм), за яким АЛП обробляє оцифрований мовний сигнал. Результати обчислень відображаються на БВІ. Роботою АЦП, ОЗП, АЛП і БВІ керує БУ.



Фіг.