

УДК 621.391

Ткаченко М. В., к.т.н.<sup>1</sup>;  
Федоренко Р. М., к.е.н.<sup>2</sup>;  
Зотова І. Г.<sup>3</sup>

<sup>1</sup> – Київський національний університет імені Тараса Шевченка, Київ;

<sup>2</sup> – Управління інформаційних технологій Міністерства оборони України, Київ;

<sup>3</sup> – Центр воєнно-стратегічних досліджень Національного університету оборони України імені Івана Черняхівського, Київ

## Методи і технології біометричного розпізнавання суб'єкта за голосом

**Резюме.** Розвиток інформаційних технологій, і множина практичних застосувань, в яких використовуються технології ідентифікації та верифікації за голосом, роблять цю область актуальною для теоретичного і практичного дослідження.

**Ключові слова:** розпізнавання мовця, ідентифікація, верифікація, мовний сигнал, індивідуальні характеристики.

**Постановка проблеми.** Розпізнавання мовлення відноситься до біометричних способів ідентифікації особистості нарівні з ідентифікацією особи за відбитками пальців, конфігурації кисті руки, за формою і особливостями обличчя, сітківці ока або структури генів. Одним із головних стимулів щодо дослідження процесів розпізнавання усного мовлення є прагнення до більш достовірного розпізнавання особистості, оскільки унікальність біометричних характеристик забезпечує більш високу надійність ідентифікації.

**Аналіз останніх досліджень і публікацій.** Дослідженню проблем верифікації та ідентифікації біометричних даних, аналізу біологічних особливостей формування звукового сигналу значну увагу приділяли такі провідні вчені, як: Г. С. Рамішвілі [1], М. А. Сапожков [4], Дж. Доддингтон [3] та інші. Однак у працях цих дослідників недостатньо проаналізовано методи і засоби автентифікації біометричних даних залежно від біологічних особливостей формування звукового сигналу.

**Мета статті.** Аналіз різниці між верифікацією та ідентифікацією мовного сигналу, яка визначає різну постановку задачі та різні методи її вирішення. Крім того, проводиться аналіз фізичних особливостей формування голосу, що визначають індивідуальну особливість тональності та речення суб'єкта.

**Виклад основного матеріалу.** Системи розпізнавання суб'єкта за голосом будуються на принципі відмінності голосів різних людей один від одного. Розпізнавання мовця - це

процес визначення на основі характеристик мовного сигналу факту належності цього висловлювання конкретній особі. Системи розпізнавання суб'єкта за голосом поділяються на два види: верифікацію та ідентифікацію мовця. При верифікації отриманий зразок голосу особи порівнюється з еталоном і встановлюється його ідентичність цьому еталону. Необхідно прийняти рішення з двох можливих: чи є суб'єкт тим, за кого себе видає, або ні. Для такого рішення використовується сукупність параметрів, що містять необхідну інформацію про індивідуальність особи та вимірюваних за однією або кількома фразами. Виміряні значення порівнюються з аналогічними параметрами еталонних зразків суб'єкта, що підлягає розпізнаванню.

Завдання ідентифікації суб'єкта значно відрізняється від завдання верифікації. До того ж необхідно виявити із сукупності  $N$  суб'єктів ту особу, яка за своїми голосовими характеристиками буде збігатися з характеристиками заздалегідь збережених еталонів. Для ідентифікації невідомої особи запис його промови повинна почергово порівнюватися з еталонними записами всіх  $N$  можливих суб'єктів, через що величина ймовірності помилки ідентифікації залежить від значення  $N$  [1].

До засобів розпізнавання слід віднести комп'ютерні системи ідентифікації/верифікації, оскільки процес може бути цілком автоматизований і виключати прийняття рішення експертом.

Комп'ютерне розпізнавання суб'єкта починається з цифрової обробки мови. Мовний сигнал перетворюється з використанням таких

методів цифрової обробки, які зберігають індивідуальні особливості особи. Наступний етап - порівняння з наявними еталонними записами зареєстрованого числа суб'єктів в базі даних комп'ютера. Рішення про розпізнавання приймається згідно з подібністю чи розбіжністю порівнюваного зразка з еталоном. Якщо в системах верифікації вхідний мовний сигнал має ступінь відмінності менше заданої величини - ототожнення пройшло успішно, при відмінності більш заданого - система визнає результат розпізнавання негативним. При автоматичній ідентифікації система вибирає того зареєстрованого суб'єкта, чий контрольний зразок найбільш близький до вхідного сигналу. Результат розпізнавання певною мірою залежить від обраних параметрів мовних характеристик, за якими проводиться порівняння [2].

При всій можливій схожості завдання ідентифікації/верифікації мовця, вони відрізняються стратегіями прийняття рішення. Зазвичай при текстозалежній верифікації суб'єкта стратегія полягає у порівнянні параметрів мови невідомого мовця з параметрами еталона в еквівалентних точках часової осі. Це означає, що вхідний мовний сигнал поєднується в часі з еталонним сигналом, а потім обчислюється відстань між відповідними точками на осі часу, яке осереднюється за всім мовним фрагментом. Цей метод став основою майже всіх підходів до вирішення завдання текстозалежної верифікації суб'єктів. Винятком з цього правила є підхід, при якому обчислюються усереднені за часом статистичні оцінки параметрів мовних кадрів, і рішення за ототожненням голосу ґрунтуються на обчисленні оцінки максимальної правдоподібності для еталонного голосу [3].

Стратегія текстонезалежної ідентифікації ґрунтується на двох підходах. *Перший підхід* заснований на використанні характеристик, усереднених на великому інтервалі часу. Це означає, що певні характерні ознаки мовного сигналу обчислюються для кожного кадру мовних даних і потім осереднюються за всім фрагментом мови. Рішення щодо розпізнавання приймається на основі обчислення статистичної правдоподібності усередненого ідентифікуючого вектора відповідно до гіпотез приналежності голосу конкретного суб'єкта. Другий метод полягає в знаходженні у прийнятому мовному сигналі характерних фонетичних явищ, а потім співставленні

характеристик, відібраних і визначених фонетичних явищ, з характеристиками відповідних фонетичних явищ в еталонних голосах [3].

Крім вибору стратегії розпізнавання необхідно вибрати параметри мовного сигналу, за якими відбувається порівняння з еталонними зразками.

У кожній мові існують інваріантні структури, які допомагають людям розуміти один одного, але все ж голос і мова кожного з нас індивідуальна, і саме ця теза лежить в основі розпізнавання мовця за голосом.

Індивідуальні голосові якості різних людей визначаються цілим рядом параметрів, що включають анатомічні, артикуляційні, вікові, соціальні характеристики, дефекти мови, а також різні поєднання цих параметрів.

Звук мовлення є результатом послідовної взаємодії чотирьох артикуляційних процесів, на кожному з яких формуються індивідуальні ознаки.

1. З легких виштовхується струм повітря, який є основою формування звучання. Вже на першому етапі формуються відмінності в мові, які залежать від ємності легень і характеру дихання (ключичний, грудний або черевний тип). Людина при цьому робить різну кількість вдихів і видихів і витрачає на них різний час [1]. Це визначає цикл мовного дихання, ритмічну структуру мови і силу звучання.

2. Повітряний потік починає вібрувати, проходячи через голосові зв'язки, розташовані в гортані. На другому етапі також формуються голосові відмінності, які обумовлені розмірами гортані та голосових зв'язок. Голос вже на даному етапі характеризується певною висотою, силою і тембром. Останні дві характеристики з проходженням звукових коливань через глотку, потім ротову і носову порожнини істотно видозмінюються залежно від параметрів цих порожнин-резонаторів. Що ж стосується висоти звуку, то вона зберігається, що є однією з основних особливостей індивідуального голосу.

Ця особливість, тобто висота голосу, напряму залежить від коливання голосових зв'язок, які залежать від довжини, товщини і натягу останніх. Довгі, товсті та слабо натягнуті зв'язки забезпечують низькі за висотою звуки. Збільшення натягу зв'язок, що здійснюється за допомогою м'язового апарату гортані, тягне за собою підвищення висоти звуку [1].

3. Вібрація в струмені повітря знаходить особливу форму завдяки резонаторам,

сформованим в ротовій і носовій порожнинах органами артикуляції.

Тут голосу надається індивідуальне темброве забарвлення, завдяки цілеспрямованій регуляції людиною взаємного розташування рухомих (язик, губи) і нерухомих (м'яке і тверде піднебіння, зуби) органів. Залежно від обсягу повітряного потоку, що йде в носову порожнину або порожнину рота, з'являються носові або назалізовані звуки.

4. Розповсюдження повітряної хвилі особливої форми в навколишнє середовище.

Форма мовної хвилі однозначно визначається джерелом і фільтром. Мовний тракт є складним акустичним фільтром з низкою резонаторів, що створюються порожнинами рота, носа і носоглотки [1].

У системах розпізнавання використовуються "низькорівневі" акустичні параметри, до числа яких відносяться: частота основного тону, частоти максимумів спектра, ширина смуги кожної з них і характерні аномалії голосу. Усі ці характеристики можуть бути виміряні як функції часу. Як змінні розпізнавання можуть також використовуватись відповідні статистичні характеристики, усереднені на великому інтервалі часу [3].

Дослідження частоти основного тону виявили, що його особливістю є зміна в значних межах тривалості інтервалів при вимові окремих фраз, а також наявність у багатьох людей різного основного тону для вимову тих чи інших фраз. Перше явище має назву мелодії основного тону. Вона характерна, наприклад, для питальних і окличних речень. Але у деяких людей і в ряді іноземних мов вона часто зустрічається і в звичайних фразах. З цієї особливості можна впізнати голос мовця. Оскільки в цьому випадку відбувається плавна зміна тону, то воно може бути виділено за допомогою приладів [4].

Ці спостереження показують, що при розпізнаванні суб'єктів за голосом можна опиратися не тільки на фонемний аналіз окремих звукових сегментів, а й на просодичну інформацію, яка несе додаткові індивідуальні ознаки. Просодична інформація може виявитися більш корисною при ідентифікації мовця, ніж при верифікації. Це пов'язано з тим, що ідентифікація спирається на текстонезалежний аналіз, де часто можна використовувати тривалі мовні фрагменти.

До просодичних характеристик мови відносять частоту основного тону, тривалість і

інтенсивність. Аналіз результатів досліджень, проведених на супрасегментних рівнях, показує, що частота основного тону - акустичний корелят висоти тону - є однією з найуніверсальніших супрасегментних характеристик. Майже всі види інтонаційної інформації можуть бути передані за допомогою модифікацій частоти основного тону. Значення частоти основного тону, що фіксуються, несуть інформацію про характер роботи голосового джерела, який обумовлюється як мінімум трьома факторами: квазіперіодичністю коливань голосових зв'язок, індивідуальними особливостями голосового джерела та емоційним станом мовця. До числа параметрів частоти основного тону зазвичай відносять середній рівень частоти, частотний діапазон, частотний інтервал, швидкість зміни (підйому або падіння) частоти основного тону [5].

Мовний сигнал є складним, багатоплановим явищем, параметри якого залежать від цілого ряду особливостей кожної людини. Вибір аналізованих параметрів, за якими може проводитися розпізнавання суб'єкта, повинен спиратися на цілі, методи і завдання розпізнавання. Сучасні системи розпізнавання мовця по голосу прагнуть використовувати якомога більше інформації більш високого рівня, яка тісно пов'язана з інформацією низького рівня. Імовірність правильного розпізнавання мовця, зростає пропорційно тому, які ознаки аналізуються, кількість ознак і їх кореляція.

**Висновки.** Якщо при верифікації рішенням може бути використання автоматичних систем, то при ідентифікації на сьогодні немає чітких характеристик, за якими можна виявити подібності між замаскованим голосом і еталоном, оскільки поки не існує алгоритмів приведення замаскованого голосу до "нормального" зразка. Поєднання суб'єктивних і об'єктивних методів можуть підвищити точність розпізнавання, але ймовірність правильного суб'єкта у разі маскування голосу досі залишається невисокою.

**Подальші дослідження.** Будь-який сигнал, що буде аналізуватися у майбутньому, потребує попередньої обробки - оцифрування з видаленням шумових ефектів, визначення та видалення несуттєвих проміжків. Так і мовний сигнал потребує попереднього оброблення, що є майбутньою темою досліджень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Рамишвили Г. С. Автоматическое опознавание говорящего по голосу. – М.: Радио и связь, 1981. – 224 с.
2. Furui S. Digital speech processing, synthesis, and recognition. – 2nd ed. – USA : Marcell Dekker Ink., 2001. – 453 p.
3. Доддингтон Дж. Распознавание дикторов: Идентификация людей по голосу // ТИИЭР. – 1985. – № 11 (73). – С. 129–146.
4. Сапожков М. А. Речевой сигнал в кибернетике и связи. – М.: Связьиздат, 1963. – 450 с.
5. Общая и прикладная фонетика: учеб. пособие / Л. В. Златоустова, Р. К. Потапова, В. В. Потапов, В. Н. Трунин-Донской. – М.: Изд-во МГУ, 1997. – 416 с.

Стаття надійшла до редакції 23.05.18

**Ткаченко М. В., к.т.н.<sup>1</sup>;**

**Федоренко Р. М., к.е.н.<sup>2</sup>;**

**Зотова І. Г.<sup>3</sup>**

<sup>1</sup> – Киевский национальный университет имени Тараса Шевченко, Киев;

<sup>2</sup> – Управление информационных технологий Министерства обороны Украины, Киев;

<sup>3</sup> – Центр военно-стратегических исследований Национального университета обороны Украины имени Ивана Черняховского, Киев

**Методи и технологии биометрического распознавания субъекта по голосу**

**Резюме.** Развитие информационных технологий и многочисленные практические приложения, в которых используются технологии идентификации и верификации по голосу, делают данную область актуальной для теоретического и практического исследования.

**Ключевые слова:** распознавание говорящего, идентификация, верификация, речевой сигнал, индивидуальные характеристики.

**M. Tkachenko, PhD (Technical)<sup>1</sup>;**

**R. Fedorenko, PhD (Economic)<sup>2</sup>;**

**I. Zotova<sup>3</sup>**

<sup>1</sup> – Kyiv National Taras Shevchenko University, Kyiv;

<sup>2</sup> – Informational technologies department, Kyiv

<sup>3</sup> – Center for Military and Strategic Studies of the National Defence University of Ukraine named after Ivan Cherniakhovskiy, Kyiv

**Methods and technologies of the subject's biometric recognition by voice**

**Resume.** The development of information technologies and numerous practical applications, which use technologies of identification and verification by voice, make this area relevant for theoretical and practical research.

**Keywords:** speaker recognition, identification, verification, speech signal, individual characteristics.