УДК 621.391

Ткаченко М. В., канд. техн. наук[1] (ORCID: 0000-0003-2929-3495);
Федоренко Р. М., канд. екон. наук[1] (ORCID: 0000-0001-9433-5458);
Берестов Д. С., канд. техн. наук[2] (ORCID: 0000-0002-3918-2978)

[1] – Київський національний університет імені Тараса Шевченка, Київ;
[2] – Центр воєнно-стратегічних досліджень Національного університету оборони України імені Івана Черняховського, Київ

Сучасні методи автоматичної ідентифікації диктора за голосом

Резюме. Проведено аналіз методів автоматичного розпізнавання диктора за голосом, на підставі якого здійснено вибір методу для розв'язання задачі текстонезалежного розпізнавання.

Ключові слова: голосовий сигнал, диктор, розпізнавання, динамічна трансформація часової шкали, приховані марковські процеси, векторне квантування, опорні вектори, гаусові суміші.

Постановка проблеми. Мова є формою спілкування людей за допомогою фонетичних конструкцій, яка склалася у процесі історичного розвитку людини. За її допомоги людина пізнає навколишній світ, передає свої знання і вміння іншим людям. Усна складова мови виявляється у вигляді висловлювань у звуковій формі, які можливі завдяки голосовому апарату людини. Кожна людина має індивідуальні голосові характеристики, які визначаються особливостями будови його голосових органів [1]. У процесі спілкування люди здатні на підсвідомому рівні розрізняти голоси інших людей, однак для обчислювальної техніки ця задача є нетривіальною.

Обробка голосу була однією з найзахоплюючих областей дослідження. Сигнали зазвичай обробляються в цифровому вигляді, тому обробку голосу можна розглядати як окремий випадок цифрової обробки сигналу. На сьогодні актуальним є пошук нових рішень в даній області, але не слід забувати про існуючі алгоритми і їх оптимізацію.

Нині існує висока потрібність у системах ідентифікації дикторів, в їх розробленні та поліпшенні зацікавлені багато представників з організацій і структур з найрізноманітнішим вектором діяльності (безпека, оборона, торгівля, промисловість тощо). Це багато в чому пов'язано з тим, що такі системи використовуються або можуть бути корисні в достатньо широкому спектрі галузей. Насамперед такі системи застосовуються для забезпечення безпеки, найчастіше для розмежування доступу до будь-якого фізичного або електронного об'єкта. Наприклад, у військових установах для обмеження проходу до будь-якого службового приміщення для всіх крім конкретних осіб. Використання голосової ідентифікації замість електронної або разом з нею, наприклад, по ключ-карті, яку зловмисник може просто вкрасти або підробити, значно підвищує рівень безпеки в установі з обмеженим доступом і знижує ймовірність перебування сторонніх на її території.

Крім забезпечення безпеки фізичних систем або об'єктів, набагато більшого поширення набуло застосування систем розпізнавання дикторів по голосу під час розмежування доступу до різних інформаційних систем або об'єктів, таким як спеціалізовані бази даних з обмеженим доступом, радіо або телефонним каналам зв'язку. В останньому випадку актуальним є ідентифікація суб'єкта-диктора противника, який знаходиться в ефірі. Ця технологія може бути використана під час проведення радіорозвідки, контррозвідки або антитерористичного моніторингу.

Сучасні методи ідентифікації мови в реальному часі висувають високі вимоги до обчислювальних ресурсів, але часто їх обсяг обмежений. Так, в мобільних пристроях неможливо застосовувати велику кількість існуючих алгоритмів, що змушує шукати ефективніші методи. У виконанні завдання ідентифікації диктора зацікавлені державні установи, бізнес-структури та інші категорії різних користувачів інформаційних послуг. На сьогодні ведуться інтенсивні наукові дослідження ідентифікації людини за голосом, проте реальне застосування таких систем на практиці обмежена обчислювальними ресурсами і складністю різних алгоритмів, що підтверджується регулярними річними звітами логістичної компанії Gartner Group [2]. За даними компанії, лише невелика кількість користувачів (до 1 % від загального числа)

задоволена ефективністю систем розпізнавання голосових характеристик диктора.

Аналіз останніх досліджень і публікацій. Завдання щодо розпізнавання особи за голосом постало більш 40 років тому, але дослідження в цій області продовжуються й досі. До того ж, не зважаючи на значне підвищення якості розпізнавання голосової інформації, проблема автоматичного розпізнавання голосу диктора в будь-якому середовищі потребує розв'язання [3-8].

Метою статті є проведення аналізу існуючих методів автоматичного розпізнавання голосу та визначення їх слабких і сильних сторін для вибору найкращого і адаптивного методу розпізнавання диктора за голосом.

Виклад основного матеріалу. Розглянемо загальне формулювання задачі ідентифікації за голосом [4]. З одного боку, вона є окремим випадком загальної задачі біометричної ідентифікації. З іншого боку, вона є частиною напряму зі створення комплексних систем біометричної ідентифікації [5].

Отже, можна стверджувати, що зазначена задача є не тільки актуальною, але і найскладнішою серед всіх завдань ідентифікації.

Специфіка ідентифікації по голосу полягає в тому, що обробляється інформація в форматі звукових файлів. Як відомо [6, 7], основною особливістю цього виду інформації є її часова протяжність. Ця особливість накладає обмеження на застосовувані для вирішення цього завдання методи. Дійсно, об'єкт для розпізнавання потрібно спочатку розбити на елементарні одиниці – деякі стаціонарні ділянки, такі як фонеми (для ідентифікації за голосом). Однією з ключових проблем подібної інформації є її варіативність.

Існують такі проблеми і обмеження задачі розпізнавання особистості за голосом, які слід враховувати під час побудови рішення:

емоційний стан диктора;

складна акустична обстановка (шуми і перешкоди);

різні канали зв'язку у процесі навчання і розпізнавання нейронних мереж;

природні зміни голосу диктора.

Мовлення являє собою складний сигнал (рис. 1), що утворюється унаслідок перетворень, які відбуваються на кількох рівнях: семантичному, голосовому, артикуляційному (рівні голосового апарату людини) і акустичному (рівні фізичних властивостей звуку). Відмінності в цих перетвореннях тягнуть за собою відмінності у властивостях голосового сигналу. Під час розвя'зання задачі розпізнавання диктора за голосом всі ці відмінності можуть бути використані для того, щоб виділити індивідуальні характеристики голосу кожної людини.

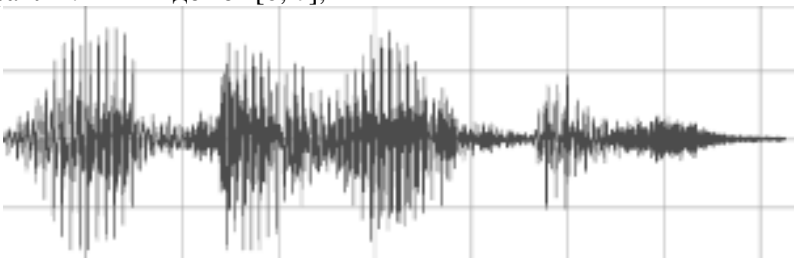

Рис. 1. Вхідний голосовий сигнал

Незважаючи на те, що методи багато чим відрізняються, загалом можна виділити такі основні етапи, характерні для кожного з розглянутих методів:

Рівень обробки сигналів. На цьому рівні сигнал обробляється для виділення ознаки, яка є істотною для завдання розпізнавання. Голосовий сигнал представляється за допомогою послідовності векторів ознак.

Рівень моделей. Під час реєстрації користувача цей рівень використовує отриману від рівня обробки сигналів послідовність векторів ознак для побудови моделі. Моделювання може полягати як у простому копіюванні векторів ознак, так і у побудові імовірнісних моделей або інших структур. Після чого стає можливим за даних ознаках обчислити ступінь подібності між ознаками і збереженою моделлю.

Рівень прийняття рішень. Функції прийняття рішень традиційно виділяють в окремий рівень, він може виконувати тривіальні функції або відсутній, якщо на рівні моделей обчислюються кінцеві рішення. Для прийняття рішень використовуються ступені подібності, обчислені на рівні моделей, та, якщо необхідно, задані пороги.

Dynamic Time Warping (DTW) – метод динамічної трансформації часової шкали дає змогу знайти близькість між двома

послідовностями вимірювань за деякий проміжок часу. Загалом ці послідовності можуть бути різної довжини, і вимірювання можуть проводитися з різною швидкістю [3]. Основною перевагою алгоритму DTW є простота реалізації. Проте цей алгоритм непридатний для розв'язання задачі текстонезалежної ідентифікації диктора.

Hidden Markov Model (HMM) – прихована марківська модель - статистична модель, яка може використовуватися для розв'язання задачі класифікації прихованих параметрів на основі спостережуваних. HMM є кінцевий автомат, в якому переходи між станами здійснюються з певною ймовірністю, і задано стартовий стан, з якого починається процес. Через дискретні моменти часу може здійснюватися перехід в нові стани. До того ж кожному прихованому стану із заданою ймовірністю відповідає стан, що спостерігається. Крім того, поточний стан автомата залежить тільки від кінцевого числа попередніх, а закон зміни станів не змінюється в часі [4]. Налаштування (навчання) HMM полягає в поєднанні її параметрів у напрямі максимізації апостеріорної ймовірності правдоподібності сигналів, відповідних векторах набору еталонів. Налаштовану HMM можна розглядати як джерело деякого випадкового сигналу з цілком певними характеристиками. Невідомий образ (голосовий сигнал) представляється у вигляді послідовності "спостережень" (список). Потім для кожної моделі знаходиться ймовірність того, що побудована послідовність могла бути генерована саме цією моделлю. Рішення на користь деякого класу приймається за найбільшою ймовірності. Якщо найбільша ймовірність менше деякого фіксованого порога, то робиться висновок про те, що ідентифікований голос не відноситься ні до одного з існуючих користувачів. HMM дає змогу моделювати істотні зміни голосового сигналу і структуру розмовної мови в апараті статистичного моделювання. CMM використовує марківський ланцюг для опису множини реалізацій фонем у реченні. HMM мають високу точність розпізнавання, але, як і DTW, застосовуються в основному для задач текстозалежної ідентифікації диктора.

Vector Quantization (векторне квантування) – розбиття простору можливих значень векторної величини на кінцеве число областей. Цей метод обробки сигналу дає змогу моделювати щільність ймовірності, функції розподілу векторів. Спочатку цей метод використовувався для стиснення даних. Він працює за допомогою поділу великого набору векторів на групи, що мають приблизно однакові значення. У методі векторного квантування вибірка з навчальних векторів перетворюється в фіксовану множину кодових векторів. Одним з поширених методів формування подібної множини, званого також кодовою книгою, є алгоритм К-середніх. Алгоритм К-середніх розбиває вихідну множину на K кластерів, де K – попередньо задане число. Для цього спочатку значення середніх ініціалізується деякими векторами з початкової множини. Потім на кожній ітерації алгоритму відбувається розподіл векторів у найближчі до них кластери (для цього обчислюється відстань між вектором і поточними значеннями середніх) і перерахунок середнього в кожному кластері. Алгоритм завершується після того, як на черговій ітерації стану кластерів не змінилися або після досягнення заданої максимальної кількості ітерацій. Отримані значення середніх є кодовими векторами, які використовуються для побудови шаблону. Метод векторного квантування простий в реалізації, може бути застосований до задачі текстонезалежної ідентифікації диктора, проте не завжди дає високу точність розпізнавання.

Support vector machine (метод опорних векторів) – метод полягає в побудові оптимальної поділяючої гіперплощини. Під оптимальною розуміється гіперплощина, яка перпендикулярна найкоротшому відрізку, що з'єднує опуклі оболонки різних класів, і проходить через середину цього відрізка. Іншими словами, оптимальна гіперплощина повинна максимізувати ширину поділяючої смуги між класами. Метод опорних векторів дає високу точність класифікації, має теоретичне обґрунтування, дає змогу застосовувати різні підходи до класифікації згідно з вибором функції. Серед недоліків слід відзначити проблему повільного навчання нейромережі в разі завдання багатокласового розпізнавання.

Gaussian Mixture Model (модель гаусових сумішей) являє собою параметричну функцію щільності ймовірності. Ця модель є вдалою варіацією стохастичної моделі для побудови систем розпізнавання [8]. До того ж необхідно відмітити, що стохастичні методи ґрунтуються на припущенні, що аналізовані дані є реалізаціями випадкового процесу. Це дозволяє, використовуючи наявні вимірювання як зафіксовані значення (умовне моделювання), отримати нескінченно багато значень (реалізацій) змінної в точці

оцінювання. Побудовані таким чином стохастичні реалізації мають ту ж функцію розподілу і таку ж просторову кореляційну структуру, що і вихідні дані. Наявність декількох рівно ймовірних оцінок в одній точці дає змогу визначити невизначеність оцінювання і побудувати ймовірнісні карти – оцінки ймовірності перевищення заданого рівня значень або оцінки, які можуть бути перевищені дійсними значеннями із заданою вірогідністю. Насамперед модель гаусових сумішей є моделлю стохастичних процесів у вигляді параметричної ймовірної функції, що дає змогу отримати приблизне значення розподілу ймовірностей вхідного стохастичного процесу.

Ці функції зручні для моделювання характеристик голосу диктора, каналу звукозапису, навколишнього середовища. Кожна з компонент моделі відображає деякі загальні, але індивідуальні для кожного диктора особливості голосу. Саме тому цей підхід можна успішно застосовувати для розв'язання задачі ідентифікації диктора. Звичайні системи розпізнавання мови використовують моделі гаусових сумішей на основі HMM. Модель гаусових сумішей описується виразом (1) у вигляді зваженої суми M компонент

$$P(x/\lambda)=\sum_{i=1}^{M} w_i b_i(x), \quad (1)$$

де x – D-мірний вектор випадкових величин;

λ – модель диктора;

$w_i, 1 \le i \le M$ - вага компонентів моделі, $\sum_{i=1}^{M} w_i = 1$;

$b_i, 1 \le i \le M$ - функції щільності розподілу компонент моделі (так званий Гауссіан [8]):

$$b_i(\vec{x})=\frac{1}{(2\pi)^{D/2}\left|\Sigma_i\right|^{1/2}}\times\exp\left[-\frac{1}{2}(\vec{x}-\overline{\mu_i})^T\Sigma_i^{-1}(\vec{x}-\overline{\mu_i})\right], \quad (2)$$

де $\overline{\mu_i}$ – вектор математичного сподівання;

Σ_i – коваріаційна матриця.

На рис. 2 наведено представлення щільності розподілу одновимірної гаусової суміші, яка містить 3 компоненти з різними коефіцієнтами ваги w_i.

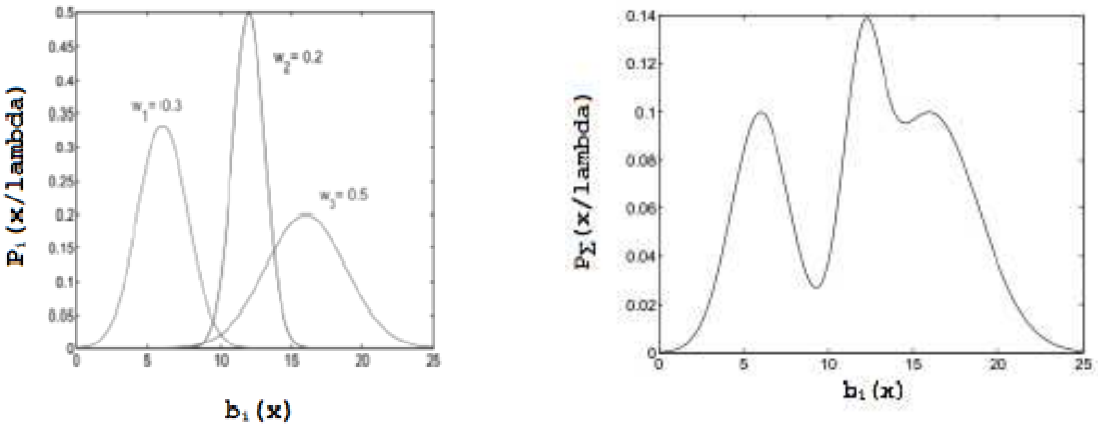

Рис. 2. Щільності розподілу одновимірної гаусової суміші

Вектори математичного сподівання, коваріаційні матриці і ваги сумішей повністю визначають модель гаусової суміші для кожного компонента моделі:

$$\lambda=\{p_i,\mu_i,\Sigma_i\}, i=\overline{1,M}. \quad (3)$$

Під час використання цього методу кожного з дикторів можна представити як модель гаусових сумішей λ.

Для того щоб побудувати модель диктора, необхідно точно оцінити її параметри, які найбільш точно відповідають розподілу векторів ознак навчального висловлювання. Існує певна низка методів для оцінки параметрів моделі. Одним з найпопулярніших, який добре себе зарекомендував є метод оцінювання максимальної правдоподібності [9]. Мета оцінювання полягає у визначенні параметрів моделі, які максимально підвищать ймовірність правдоподібності цієї моделі у разі заданих даних для навчання.

Для детермінованої послідовності векторів $X=\{x_1,x_2,...,x_T\}$ правдоподібність моделі гаусових сумішей може бути записано у вигляді

$$P(X/\lambda)=\prod_{t=1}^{T} P(x_i/\lambda). \quad (4)$$

Вираз (4) являє собою нелінійну функцію від набору параметрів λ. Під час використання цієї функції немає можливості

безпосередньо обчислити правдоподібність гаусових сумішей.

Нехай $S = \{S_1, S_2, ..., S_N\}$-- група дикторів, які представлені набором моделей гаусових сумішей $\lambda_1, \lambda_2, ..., \lambda_N$. Необхідно знайти таку модель, яка під час ідентифікації максимізує значення апостеріорної ймовірності для заданого зразка:

$$S = \arg\max P(\lambda_k / X) = \arg\max P(X/\lambda_k),\ 1 \le k \le N \quad (5)$$

Логарифмуючи отриманий вираз і враховуючи незалежність між спостереженнями, отримуємо систему ідентифікації дикторів:

$$S = \arg\max \sum_{t=1}^{T} \log P(\overline{x_t}/\lambda_k),\ 1 \le k \le N. \quad (6)$$

Моделі гаусових сумішей є ефективним алгоритмом, який дає змогу проводити ідентифікацію з високою точністю розпізнавання [10]. Однак виникає низка проблем, пов'язаних з вибором числа компонентів моделі та ініціалізацією її початкових параметрів.

Для зменшення впливу цих проблем використовується алгоритм ініціалізації початкових параметрів, який заснований на кластеризації векторів ознак голосового сигналу. За алгоритм кластеризації використовується алгоритм пошуку початкових значень центроїдів кластерів *K-means++*, евклідова відстань у якому використовується як запобіжне спотворення. Центр першого кластера вибирається випадково, після цього кожен наступний центр обирається з решти точок даних з ймовірністю, пропорційною квадрату відстані до самого ближнього існуючого центру кластера. Цей метод дає досить хороше зниження похибки підсумкового результату. Хоча початковий вибір в алгоритмі потребує додаткового часу, головна частина алгоритму *K-means* сходиться досить швидко.

Алгоритм *K-means* застосовується до об'єктів, що є точками в d-вимірному векторному просторі. Отже, ці кластери набору d-мірних векторів $D = \{x_i\},\ i = \overline{1, N},$ де $x_i \in R_d$ є *i*-м об'єктом або "точкою даних". Алгоритм *K-means* пов'язує всі точки даних в *D* так, що кожна точка x_i потрапляє тільки у визначений один з *k* розділів. Можна відстежувати, яка саме з точок знаходиться в конкретному кластері, якщо призначити номер кластера кожній точці. Точки з таким же номером кластера знаходяться в тому самому кластері, водночас, як точки з різними номерами кластера знаходяться в різних кластерах. Це можна позначити як кластерний складовий вектор *m* розміром *N*, де m_i буде номером кластера x_i. До того ж виникає проблема вибору оптимального числа кластерів (значення *k*). Цю проблему пропонується розв'язувати за допомогою використання EM-алгоритму, на кожній ітерації якого обчислюються такі значення:

$$p_i = \frac{1}{T}\sum_{t=1}^{T} p(i/\overline{x_t}, \lambda), \quad (7)$$

$$\mu_i = \frac{\sum_{t=1}^{T} p(i/\overline{x_t}, \lambda)\overline{x_t}}{\sum_{t=1}^{T} p(i/\overline{x_t}, \lambda)}, \quad (8)$$

$$\Sigma_i = \frac{\sum_{t=1}^{T} p(i/\overline{x_t}, \lambda)(\overline{x_t} - \overline{\mu_i})^T}{\sum_{t=1}^{T} p(i/\overline{x_t}, \lambda)}. \quad (9)$$

До того ж значення апостеріорної ймовірності обчислюється за формулою

$$p(i/x_t, \lambda) = \frac{p_i b_i(x_t)}{\sum_{k=1}^{M} p_k b_k(x_t)}. \quad (10)$$

Однією з основних проблем у процесі навчання моделі гаусових сумішей є вибір числа компонентів моделі. Теоретичного вирішення цього завдання не існує. Вибрати оптимальну кількість компонентів моделі можна, перебравши і оцінивши точність розпізнавання у разі заданого числа компонентів. Для тестування були використані тренувальні та тестові записи, які включали в себе 20 голосів різних людей тривалістю 3-5 с. Для визначення числа компонент, що входять до складу моделі, розглядалися моделі з числом компонент від 1 до 10. На рис. 3 наведено залежність ймовірності правильної ідентифікації від числа компонент моделі гаусової суміші. Похибка результатів склала 3 %.

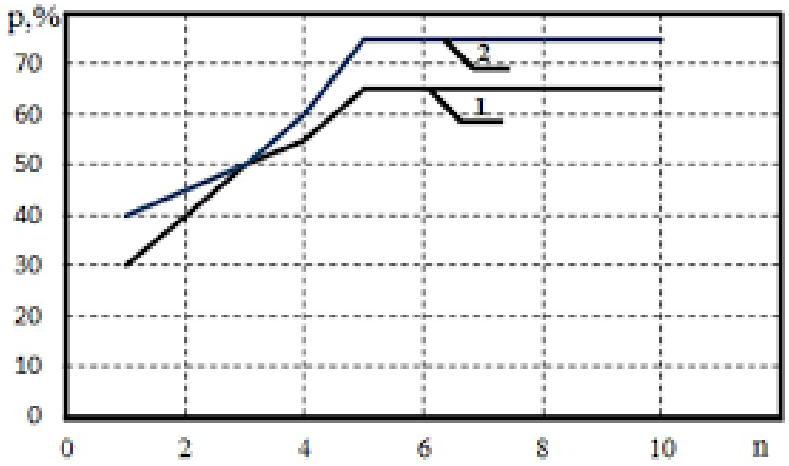

Рис. 3. Залежність ймовірності ідентифікації від числа компонент моделі гаусових сумішей

Аналіз залежності, наведеної на рис. 3 (крива 1), показує, що найефективніше число компонент дорівнює 5, оскільки з подальшим збільшенням числа компонент збільшення ймовірності правильної ідентифікації людини по голосовому сигналу дуже малий. Точність ідентифікації за такої умови досягає 65 %.

Додатково проводився експеримент зі збільшенням тривалості сигналу в навчальній вибірці до 60 секунд і збільшенням тривалості тестового висловлювання до 10 секунд. На рис. 3 (крива 2) зображена залежність ймовірності ідентифікації від числа компонент моделі гаусових сумішей.

Аналіз залежності, наведеної на рис. 3 (крива 2), показує, що найефективніше число компонент дорівнює 5, оскільки з подальшим збільшенням числа компонент збільшення ймовірності ідентифікації людини по голосовому сигналу також дуже малий. Точність ідентифікації за такої умови досягає 75 %.

З представлених залежностей можна дійти висновку, що зі збільшенням тривалості навчального висловлювання збільшується ймовірність правильної ідентифікації диктора.

Висновки. На сьогодні існує невелика кількість методів, що дають змогу вирішувати завдання текстонезалежної ідентифікації диктора за голосом, причому кожен з наведених методів має свої переваги і недоліки. Проте найпоширенішим методом є Gaussian Mixture Model. Моделі гаусових сумішей добре себе зарекомендували в якості стохастичної моделі для побудови систем розпізнавання [11, 12]. Вони зручні не тільки для моделювання характеристик голосу диктора, але і каналу звукозапису, навколишнього середовища. Окремі компоненти моделі можуть моделювати окрему множину акустичних ознак. Кожна з компонент моделі відображає деякі загальні, але індивідуальні для кожного диктора особливості голосу. Саме тому цей підхід можна успішно застосовувати для вирішення завдання текстонезалежної ідентифікації диктора.

На точність роботи систем розпізнавання впливає низка факторів. По-перше, необхідно відзначити мінливість самого голосу. Емоційний стан, втома, вікові зміни, застуда і багато інших чинників впливають на голос. По-друге, проблемою для систем розпізнавання є вплив навколишнього середовища та зміна умов запису. Бази даних, що використовуються для експериментального оцінювання, не завжди здатні змоделювати перераховані ситуації. Тому результат істотно залежить від розміру представницької бази і як побудований експеримент. Результат також залежить від тривалості матеріалу, використовуваного в кожному тесті та для створення моделей, від кількості користувачів у базі.

Для оцінювання систем ідентифікації в більшості випадків обмежуються замкнутою множиною користувачів, тобто всі користувачі, що проходять спробу ідентифікації, зареєстровані в системі. Результат залежить від кількості зареєстрованих користувачів і від розміру списку (найчастіше використовують лише один код) або від порога включення в список. Імовірність ідентифікації (істинно-позитивної ідентифікації) оцінюють як частку спроб ідентифікації, унаслідок яких було повернуто список кандидатів, що містить вірний ідентифікатор.

Напрями подальших досліджень. Ефективна система розпізнавання мови має містити в собі такі етапи обробки вхідного

сигналу, як видалення шуму, сегментація, виділення вокалізованих ділянок, параметризація, розпізнавання, коригування за словником з оберненим зв'язком. Зрозуміло, що не один метод не може покрити усі етапи. Ефективна система має поєднувати в собі найкращі методи виконання кожного етапу, використовуючи їх переваги. Отже надалі планується провести аналіз і відібрати ефективні методи обробки сигналу для створення відповідної системи.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Campbell J.P. Speaker Recognition: A Tutorial // Proceedings of the IEEE. 1997. Vol. 85, № 9. pp. 1437-1462.
2. Ing-Jr Ding, Chih-Ta Yen, Yen-Ming Hsu. Developments of Machine Learning Schemes for Dynamic Time-Wrapping-Based Speech Recognition // Mathematical Problems in Engineering. 2013.
3. Daniel Ramage. Hidden Markov Models Fundamentals // CS229 Section Notes. 2007.
4. Mamou J., Mass Y., Ramabhadran B., Sznajder B. Combination of multiple speech transcription methods for vocabulary independent search // In proceedings of the ACM SIGIR Workshop `Searching Spontaneous Conversational Speech. Singapore. 2008. pp. 20-27.
5. Вишняков Р. Ю. Интеллектуальные информационно-поисковые системы. Лингвистический анализ // Перспективные информационные технологии и интеллектуальные системы. 2006. № 4. С. 37-42.
6. Garofolo J., Auzanne G., and Voorhees E. The trec spoken document retrieval track: A success story. // In proceedings of the Recherche d'Informations Assiste par Ordinateur: Content Based Multimedia Information Access Conference, 2000. pp. 1-20.
7. Huijbregts M., Ordelman R., Jong F. Annotation of heterogeneous multimedia content using automatic speech recognition // In Proceedings of the second international conference on Semantics And digital Media Technologies 143 (SAMT). Lecture Notes in Computer Science. Berlin. Springer Verlag. December 2007. pp. 78-90.
8. Методы автоматического распознавания речи: в 2-х кн. / под ред. У. Ли; пер. с англ. О.В. Александровой; под ред. А. А. Воронова. М. : МИР, 1983. Кн. 1. 328 с.
9. Reynolds D.A. Speaker identication and verication using Gaussian mixture speaker models / D.A. Reynolds. Helsinki: Speech Commun, 1995.
10. Кульбак С. Теория информации и статистика. М.: Наука, 1967. 408 с.
11. X. Huang, A. Acero, H. Hon. Spoken languageprocessing: a guide to theory, algorithm, and systemdevelopment. – Prentice Hall PTR, 2001. p. 936.
12. Furui S. Digital Speech Processing, Synthesis and Recognition // Marcel Dekker, New York, 1989.
13. Navratil J., Klusacek D. On linear DETs // Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-07). 2007.
14. Martin A., Doddington G., Kamm T., Ordowski M., Przybocki M. The det curve in assessment of detection task performance // Proc. of Eurospeech. 1997. V. 4. pp. 1895-1898

Стаття надійшла до редакційної колегії 15.03.2019

Ткаченко М. В., канд. техн. наук[1]**;**
Федоренко Р. Н., канд. экон. наук [1]**;**
Берестов Д. С., канд. техн. наук [2]

[1] – Киевский национальный университет имени Тараса Шевченко, Киев

[2] – Центр военно-стратегических исследований Национального университета обороны Украины имени Ивана Черняховского, Киев

Современные методы автоматической идентификации диктора по голосу

Резюме. Проведен анализ методов автоматического распознавания диктора по голосу, на основании которого осуществлен выбор метода для решения задачи текстонезависимого распознавания.

Ключевые слова: речевой сигнал, диктор, распознавание, динамическая трансформация временной шкалы, скрытые марковские процессы, векторное квантование, опорные векторы, гауссовские смеси.

M. Tkachenko, PhD (Technical)[1]**;**
R. Fedorenko, PhD (Economic)[1]**;**
D. Berestov, PhD (Technical)[2]

[1] – Kyiv National Taras Shevchenko University, Kyiv;

[2] – Center for Military and Strategic Studies of the National Defence University of Ukraine named after Ivan Cherniakhovskyi, Kyiv

Modern methods for automatic speaker identification by voice

Resume. The analysis of the methods of automatic recognition of the announcer by voice was carried out, on the basis of which the method was chosen for solving the problem of text-independent recognition.

Keywords: speech signal, speaker, recognition, dynamic time warping, hidden markov model, vector quantization, support vector machine, Gaussian. mixtures.