ПРИКЛАДНІ ДОМЕНИ І ПРИКЛАДНЕ ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

УДК 004.934:681.391

Бєлозьорова Я.А.
Національний авіаційний університет

ПОБУДОВА АРХІТЕКТУРИ ПРОГРАМНОЇ СИСТЕМИ ІДЕНТИФІКАЦІЇ ДИКТОРА

Запропонована архітектура програмної системи ідентифікації у вигляді діаграм класів і послідовностей. Досліджено основні критерії оцінки точності ідентифікації диктора та виявлено можливі джерела втрати точності ідентифікації диктора, які можуть бути використані при побудові системи ідентифікації диктора.

Предложена архитектура программной системы идентификации в виде диаграмм классов и последовательностей. Исследованы основные критерии оценки точности идентификации диктора и выявлены возможные источники потери точности идентификации диктора, которые могут быть использованы при построении системы идентификации диктора.

The architecture of the software speaker identification system in the form of diagrams and sequences is proposed. The main criteria for evaluating the accuracy of speaker identification are investigated and possible sources of accuracy loss of speaker identification that can be used in the construction of the speaker identification system are identified.

Ключові слова: *програмна система ідентифікації диктора, вейвлет, діаграми, UML, розпізнавання мови.*

Вступ

Обробка мовного сигналу з метою ідентифікації диктора є найбільш актуальною і популярною в задачах, пов'язаних з обробкою мови. Постійний і високий попит на програмні реалізації систем ідентифікації диктора існує в різних сферах від контролю доступу користувачів до сервісів виявлення злочинця по голосу. Однак зважаючи на відсутність чіткої наукової бази алгоритмів ідентифікації, значні складності їх реалізації, а також точності ідентифікації особистості можна відзначити, що ці завдання в цілому ще досить далекі від свого остаточного вирішення.

Огляд останніх досліджень

Завданням автоматичної верифікації диктора прийнято вважати створення математичної моделі, набору алгоритмів і як результату їх застосування – програмної або програмно-апаратної реалізації, яка дозволила б виконати ідентифікацію особистості, з тією ж точністю і вірогідністю, як це доступно людині.

Дослідницькі зусилля в сфері мовних технологій привели до появи великої кількості комерційних систем розпізнавання мови. Такі компанії як Nuance, IBM, ScanSoft пропонують великий набір програмних рішень як для серверних, так і для десктопних додатків.

Для аналізу роботи програмних систем ідентифікації диктора необхідно розглянути основні підходи до виконання оцінок роботи подібних систем. Національний Інститут Стандартів і Технологій США (National Institute of Standards and Technology – NIST) координує проведення оцінок різних систем аналізу мовного сигналу: систем автоматичного розпізнавання мови, виділення ключових слів з мови, а також розпізнавання диктора. Опис деяких щорічних оцінок систем можна знайти в [1]. Інститут розробляє дослідницькі методології порівняння різних систем, які включають в себе чітку постановку завдання, визначення оціночної метрики, ретельно підібрані і єдині для всіх учасників набори тренувальних і тестових даних, чіткі вимоги до проведення та надання результатів тестів.

Одним з факторів, за якими визначають ефективність біометричних систем, є частота появи помилок. Існує два типи помилок: система може, як помилково відмовити людині, яку потрібно пропустити (клієнт) або помилково пропустити людину, якій потрібно було відмовити в доступі (зловмисник). Частота появи помилок відмови FRR (False Reject Rate) пропорційна частоті появи помилок пропуску клієнта, які відкидаються FAR (False Accept Rate) пропорційна числу спроб зловмисника, які допускаються. У більшості біометричних систем виставляється гнучкий поріг, який управляє балансом між цими двома видами помилок. У кожній програмі оптимальний поріг знаходиться емпіричним шляхом (рис. 1).

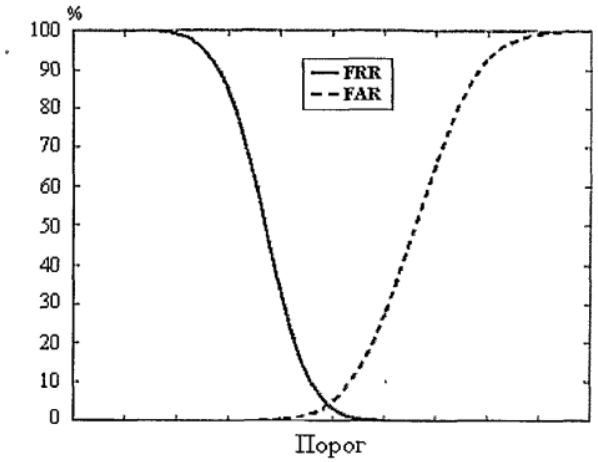

Рис. 1. Залежність FRR(частота появи помилок відмови) та FAR (частота появи помилок) від вибраного порогу

Проведені оцінки NIST різних систем ідентифікації диктора показали [5]:

- Порівняння голосів дикторів на основі обмеженого набору даних – Точка рівновірогідності помилок I і II роду лежить в межах 5-10%. Ступінь впевненості класифікатора в отриманому результаті становить приблизно 95%.
- Верифікація диктора на основі розширеного набору даних – точка рівновірогідності помилок I і II роду лежить значно нижче, в області 1,3 – 2%, що приблизно відповідає відносному зменшення кількості помилок на 74-80%.
- Порівняння голосів дикторів на основі розширеного набору даних – точка рівної ймовірності помилок – 12-15 %

З огляду на представлені помилки ідентифікації можна зробити висновок, що існуючі системи ідентифікації диктора викликають справедливі нарікання користувачів, пов'язаних з об'єктивністю результатів експертиз. Проведені дослідження [2,3] показали, що висловлюються сумніви цілком обґрунтовані. Цей висновок обумовлений головним чином тим, що в більшості сучасних засобах проведення ідентифікаційних досліджень голосових сигналів використовується перетворення Фур'є, що є штучним математичним прийомом розкладання складного сигналу на періодичні складові. Але механізм сприйняття і перетворення звукових коливань слуховим апаратом людини влаштований інакше і в ньому не можуть існувати подібні штучні перетворення. Також встановлено, що основні процеси передачі інформації в мозок, що містяться в звукових сигналах, носять імпульсний характер, а тривалості цих імпульсів лежать в межах від десятків до сотень мілісекунд [3,4], в зв'язку з вищевикладеним був зроблений висновок про необхідність використання мультифрактального підходу для побудови системи ідентифікації диктора [5].

Основна частина

Основним завданням роботи було дослідження факторів, що впливають на достовірність результатів ідентифікації. А рішення цього завдання вимагало вивчення тонкої структури мовних сигналів. Отже, потрібно створити нові програмні засоби, необхідні для проведення цих досліджень. Але в першу чергу було необхідно осмислити

результати сучасних наукових досліджень в області фізіологічного і психофізичного вивчення слухової системи і адаптувати ці знання до задачі криміналістичної ідентифікації особистості по її голосу з точки зору можливостей технічної реалізації поставлених завдань. Вирішення цього завдання забезпечувало вибір необхідного математичного апарату, що дозволяє спочатку створити, а потім і відпрацювати необхідні програми і методики. Таким чином, був потрібний комплексний системний підхід до вирішення виниклої проблеми, що поєднує в своїх рамках її нейрофізіологічні, математичні, технічні (в тому числі акустичні) і правові аспекти. Вивчення і осмислення нейрофізиологічного аспекту проблеми показало необхідність дослідження сигналів тривалістю до сотень мілісекунд. При цьому було висунуто припущення, що необхідна достовірність результатів ідентифікації може бути забезпечена при дослідженні окремих фонем в тимчасовій і частотній області подання сигналів. Вивчення сигналів у часовій області необхідно тому, що всі фонеми мають чітко виражений фрактальний характер, що зберігається та є індивідуальним для кожної з фонем, тобто форма сигналу фонеми в тимчасовій області однакова у всіх мовах і приблизно однакова при її вимові будь-яким індивідуумом [3]. Саме ця одноманітність дозволяє нам розпізнавати мову будь-якої людини. Основна відмінність, визначає індивідуальність мовця, полягає в індивідуальності частотного складу сигналів, які складають цей звук при його вимові конкретною особистістю. Ця індивідуальність, на нашу думку, визначається частотою основного тону (ОТ) і модулюються параметрами цієї частоти. І частота ОТ, і ці параметри визначаються індивідуальністю складових голосового тракту будь-якої людини [2].

При реалізації програмного забезпечення виникли дві взаємопов'язані завдання – автоматичної сегментації фонограми на фонеми і виділення, підрахунку і визначення міри близькості фрактальних утворень, що містяться в досліджуваних сигналах спірною і зразковою фонограм. Обидві ці завдання вирішені в [5,6,7].

В рамках задачі ідентифікації диктора можна виділити два взаємопов'язані завдання ідентифікації та верифікації диктора [6]. У першій задачі мета полягає в ідентифікації аудіокомпонента як виголошеного одним з дикторів з розглянутого безлічі, у другому – встановлення належності аудіокомпонента конкретному еталонному дикторові.

На підставі цих задач системи розбиваються на три частини:

1. визначення індивідуальних ознак МС (мовного сигналу);

2. уявлення характерного еталона диктора;

3. прийняття рішення про індивідуальність диктора.

На підставі вищевикладеного можна виділити наступні основні етапи реалізації системи розпізнавання диктора:

Вимірювання фрактальної розмірності компонентів сигналу. Простий у реалізації етап, але досить ефективний в наборі всіх заходів розрізнюваності. Реалізація його можлива як з постійним вікном, так і з адаптивним типом вікна.

Визначення границь фрази. Для вирішення даного завдання найбільш раціонально використовувати алгоритми сегментації мови на основі мультифрактального підходу. На основі цього підходу в тих елементах сигналу, де зміна фрактальної розмірності перевищує деякий встановлений поріг, передбачається, починається фраза.

Виділення основного тону. Для вирішення завдання виділення основного тону існує необхідність розробки перешкодостійкого методу виділення основного тону для кожного періоду. В якості базового алгоритму виділення основного тону може бути взятий алгоритм, заснований на використанні апроксимації сигналу вейвлетом Морле з подальшим статистичним аналізом розподілу вейвлет-максимумів, що фізично пояснюється наявністю самоподібних структур характерних для сигналів, пов'язаних з резонаторами.

На етапі вимірювання основного тону на ділянках сигналу має сенс порівнювати ні абсолютні величини, а нормовані - це дозволяє більш точно розрізняти дикторів по інтонаційному забарвленню.

Виділення характерних параметрів основного тону. Для вирішення цього завдання можна скористатися знаходженням тільки деяких з розглянутих параметрів при аналізі для кожного фрагменту: середньою частотою і дисперсією основного тону; розподілом періодів основного тону; амплітудною

модуляцією основного тону; частотної модуляцією періодів основного тону.

Порівняння параметрів сигналу з еталонними параметрами. Після здійснення процесу порівняння параметрів мови з еталонними потрібно вибрати з бази найбільш «близького» диктора. Для цього необхідно порівняти виділені параметри основного тону з бази на основі імовірнісного підходу.

В процесі проведення досліджень було знайдено наступна методика проведення експертизи [4,7]:

1. Для проведення досліджень ідентифікації диктора надаються дві і більше фонограм для проведення відповідності дикторів, як правило в наборі фонограм мінімум одна фонограма має чітку приналежність голосу конкретного диктора.

2. Кожна фонограма сегментується на фрагменти на основі фрактальної розмірності [4].

3. Для кожного фрагмента кожної фонограми обчислюється розподіл частот основного тону по всій довжині фонограми на основі розподілів частот основного тону, отриманих для фрагментів.

4. Дані фонограм точно ідентифікують особу (заданий власник голосу) зберігається в базу ідентифікації.

5. Для кожної з фонограм, для яких необхідно виконати ідентифікацію диктора перевіряється приналежність до одного розподілу для частот основного тону кожного фрагмента, виділеного з фонограми, з аналогічними розподілами, що зберігаються в базі ідентифікації.

6. На підставі проведеної оцінки ступеня близькості між розподілом частоти основного тону встановлюється диктор на основі ступеня близькості до розглянутого розподілу.

Розглянемо архітектуру реалізованої системи ідентифікації диктора на мові UML у вигляді діаграм класів і послідовностей. Діаграма класів відображає статичну структуру системи. Вона складається з опису класів і взаємозв'язків між ними. Діаграма послідовностей відображає динамічні зв'язки в системі, наприклад, послідовність викликів.

На рис.2 представлена діаграма викликів при попередній підготовці до виділення характеристик мови диктора. У режимі попереднього розпізнавання мови система завантажується з підготовленим конфігураційним файлом і вхідним сигналом. Розпізнавання буде здійснюватися через диспетчер конфігурації.

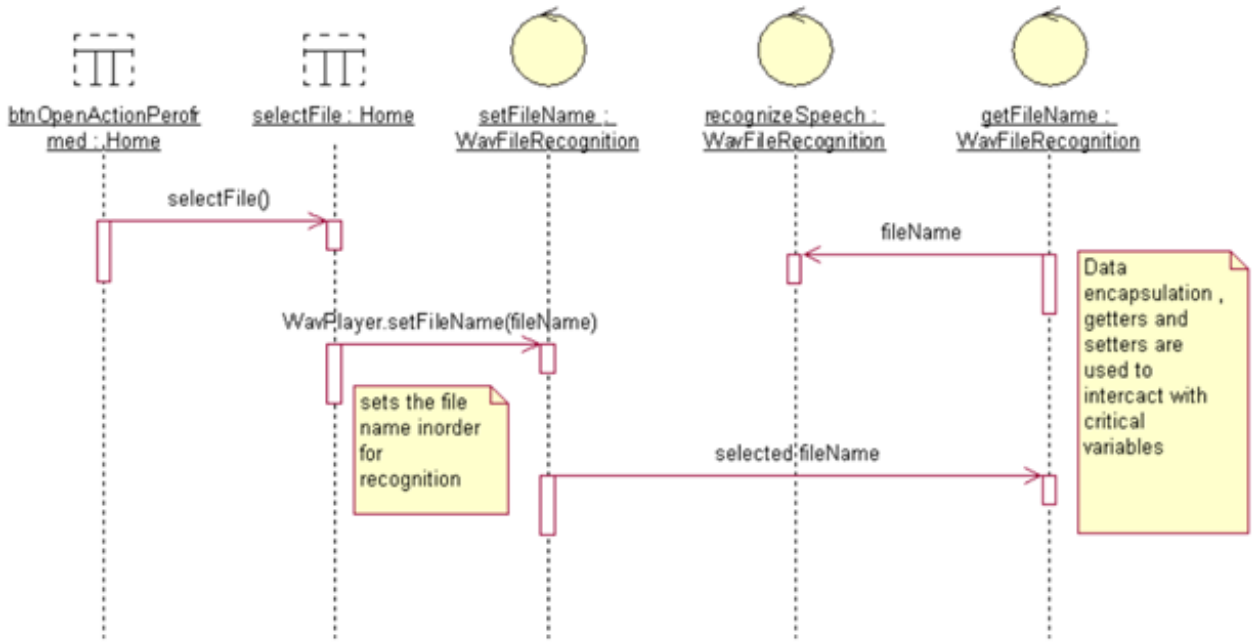

Рис.2 Послідовність викликів попередньої обробки

На рис.3 представлена діаграма виклик при постобробці при розпізнаванні диктора. На цьому етапі вхідний цифровий сигнал буде проходити через процес поділу на вокалізовані і невокалізовані ділянки, розкладання вейвлетом Морлі з подальшим статистичний аналізом розподілу вейвлет-максімумів і визначенням частоти основного тону для сегментів. Класи AudioFileDataSource і Recognizer реалізують функції для виконання цих завдань. Результатом послідовності викликів є мітки класу диктора і мови, до яких класифікатор відніс вхідний мовний сигнал.

На рис. 4 представлена діаграма класів сутностей, які є об'єктними уявленнями даних, якими керує система ідентифікації.

Home виконує роль графічного інтерфейсу програмної системи, який безпосередньо взаємодіє з DBSpeaker і WavFileRecognizer. DBSpeaker виконує функції уявлення і опису збережених записів дикторів. WavFileRecognizer призначений для реалізації процесу читання звукового сигналу (зі стріму або з файлу) та ідентифікації диктора. AudioFileDataSource реалізує функцію читання звукового сигналу, а Recognizer – ідентифікацію диктора.

Абстрактний клас VoiceFeatures призначений для зберігання і обчислення ознак вхідного мовного сигналу. Клас складається з масиву об'єктів VoiceFeatureValue і методу отримання ExtractFeatures, що виконує витяг ознак з отриманого на вхід мовного сигналу. Спадкоємцями класу є класи, що виконують пофрагментній аналіз: середню частоту і дисперсію основного тону; розподіл періодів основного тону; амплітудну модуляцію основного тону; частотну модуляцію періодів основного тону.

Абстрактний клас PersonClassifier призначений для реалізації класифікуючого алгоритму. Клас складається з методів Train і Classify, а також об'єкта Parameters, який містить всі необхідні для роботи класифікатора параметри. Метод Train приймає на вхід словник, в якому ключем є мітка класу, а значенням – об'єкт типу Features, і повертає об'єкт Parameters. Метод PersonClassify приймає об'єкт VoiceFeatureValue і повертає значення вирішальної функції, а також мітку класу – рішення.

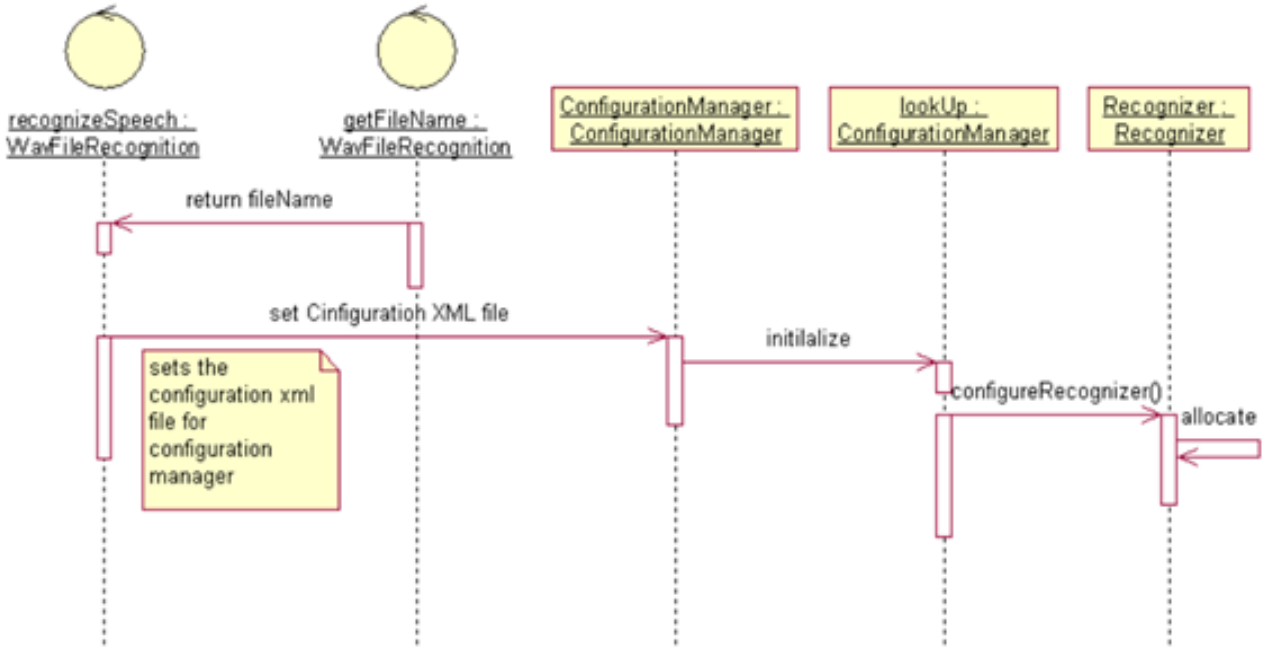

Рис.3 Послідовність викликів процесу розпізнавання диктора

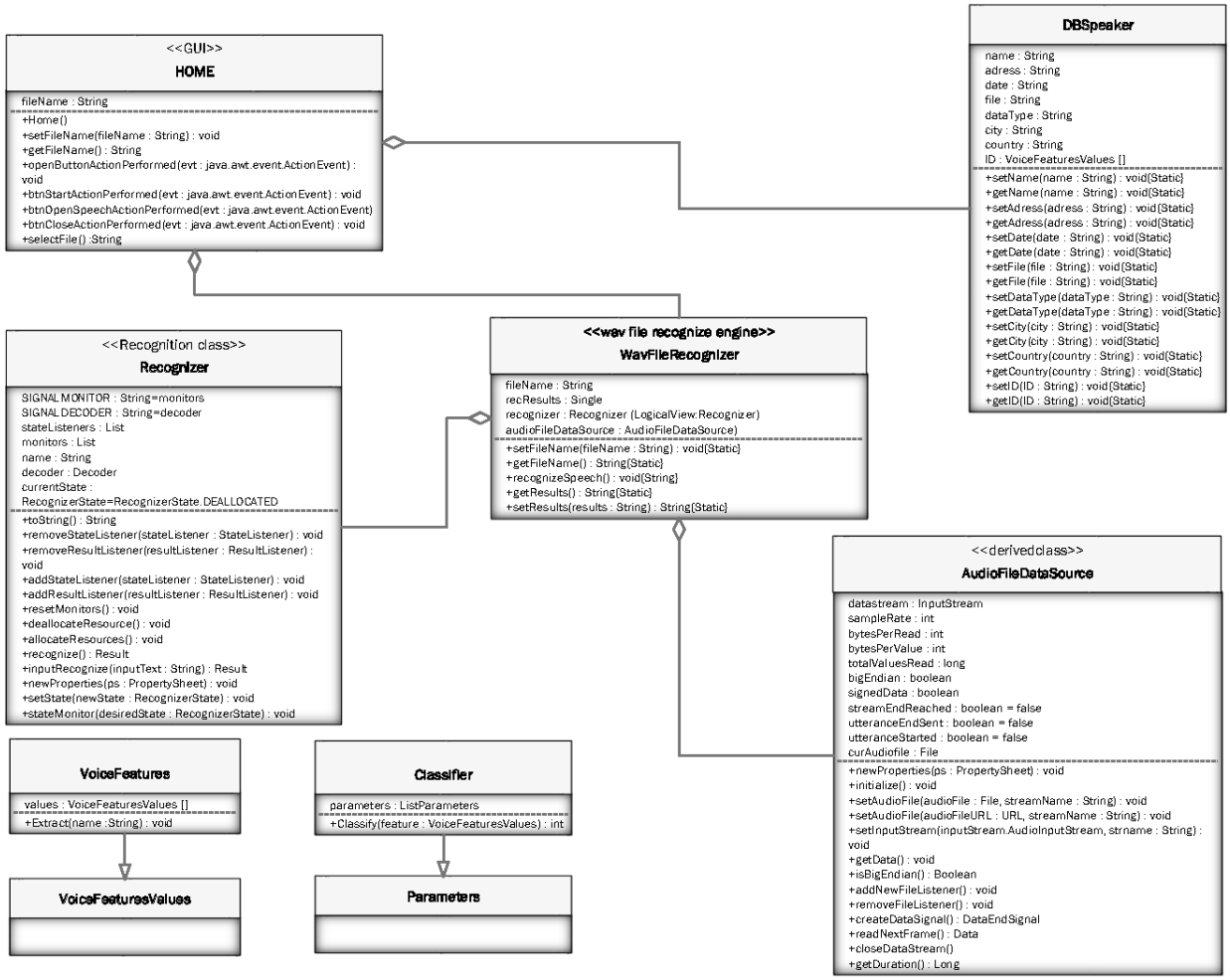

Рис. 4 Діаграма класів – сутностей

Клас SpeechUtils містить допоміжні методи, необхідні для обчислення ознак і класифікації, такі як, наприклад, обчислення поділу на вокалізовані і невокалізовані ділянки та очищення від шуму.

Таким чином, запропонована архітектура програмного забезпечення для задач ідентифікації диктора, що використовує мультифрактальний підхід в описі структури мови. Використання подібної архітектури та використання мультифрактального підходу дозволить в цілому підвищити точність ідентифікації диктора.

Висновки

На підставі виконаного дослідження можна зробити наступні висновки:

1. розглянуті підходи до побудови існуючих систем ідентифікації диктора;

2. досліджено основні критерії оцінки точності ідентифікації диктора та виявлено основні джерела втрати точності при ідентифікації диктора;

3. розглянута структурна побудова системи ідентифікації диктора, що враховує виявлені джерела втрати точності при ідентифікації диктора;

3. запропонована архітектура системи ідентифікації диктора на мові UML у вигляді діаграм класів і послідовностей.

Список використаних джерел

1. NIST Speaker Recognition Evaluation [Електронний ресурс] – Режим доступу: https://www.nist.gov/itl/iad/mig/speaker-recognition

2. Рибальський О.В. Методика ідентифікаційних і діагностичних досліджень матеріалів та апаратури цифрового й аналогового звукозапису зі застосуванням програмного забезпечення «Фрактал» при проведенні експертиз матеріалів та засобів

відео та звукозапису. Науково-методичний посібник / О.В. Рибальський, В.І. Соловйов, В.В. Журавель, Т.О. Татарнікова. – К.: ДУІКТ, 2013. – 75 с.

3. Соловьев В.И., Белозерова Я.А. Анализ алгоритмов построения системы идентификации диктора // Вісник Східноукраїнського національного університету ім.В.Даля. – 2013. – №6 (195). – Ч.1. – С. 62-67.

4. Solovjov V.I., Byelozorova Ya.A. Multifractal approach in pattern recognition of an announcer's voice. // Teka. Commission of motorization and energetics in agriculture. – 2014. – Vol. 15. – № 2. – P. 13-21.

5. Byelozorova Ya.A. The allocation of self-similar structures in voice signals for speaker identification tasks. // ScienceRise. – 2017. – № 5. – P.125-142

6. Бєлозьорова Я.А. Ідентифікація диктора на основі кратномасштабного аналізу// Інженерія програмного забезпечення. – 2017. – №1(29). – С. 15-24.

7. Соловьев В.И., Белозерова Я.А. Использование фрактальной размерности аудиофайлов в задаче сегментации звукового файла // Вісник Східноукраїнського національного університету ім.В.Даля. – 2013 – №5(194). – Ч.2. – С. 165-169.

Відомості про автора:

Бєлозьорова Яна Андріївна – асистент кафедри інженерії програмного забезпечення Навчально-наукового інституту комп'ютерних інформаційних технологій Національного авіаційного університету. Наукові інтереси: інженерія программного забезпечення

E-mail: bryuhanova.ya@gmail.com