УДК 681.12

А. В. Матіящук
Ю. Я. Таранюк

РОЗРОБКА ERP-ЗАСТОСУНКУ ДЛЯ ДІАРИЗАЦІЇ МОВЛЕННЄВИХ КОМАНД

Вінницький національний технічний університет

Анотація

Авторами розроблено консольний додаток для діаризації мовних сигналів на мові python. В основу додатка покладено попередньо створена загальна схема голосової біометрії GMM+i-vec+DNN. Точність діаризації для нашого додатку на вибірці 80 зразків 40-ка різних дикторів склала 93%. Щодо швидкості, то середня тривалість обробки голосу при навчанні системи склала 22 секунди (оброблювалося файл з вимовою тривалістю 20 секунд).

Ключові слова: діаризація, розпізнавання диктора.

Abstract

The authors have developed a console application for diarization of speech signals in Python. The application is based on the previously created general scheme of voice biometrics GMM + i-vec + DNN. The accuracy of diarrhea for our application in a sample of 80 samples from 40 different speakers was 93%. In terms of speed, the average duration of voice processing when learning the system was 22 seconds (processed a file with a pronunciation duration of 20 seconds).

Keywords: diarization, speaker recognition.

Вступ

Задача діаризації тісно пов’язане з розпізнаванням диктора за його голосом і була поставлена понад 40 років тому, і дослідження в цій області все ще тривають [1-3]. Вирішення цього завдання може знайти застосування в криміналістиці, радіорозвідці, контррозвідки, антитерористичному моніторингу, забезпечення безпеки доступу до фізичних об'єктів, інформаційних і фінансових ресурсів. Залежно від конкретної завдання розрізняють верифікацію та ідентифікацію диктора. У першому випадку користувач вказує свій ідентифікатор, і потрібно або підтвердити його або відмовити в підтвердженні. У другому випадку необхідно ідентифікувати диктора серед безлічі інших дикторів. У більшості робіт для діаризації диктора використовуються параметри у вигляді коефіцієнтів кепстра, який обчислюється по обвідної спектра, отриманого через перетворення Фур'є, за допомогою гребінки фільтрів, або по передавальної функції мовного тракту, знайденої методом лінійного передбачення. Перевага такого підходу полягає в обчислювальної простоті, а також в тому, що в кепстра відображаються індивідуальні характеристики голосового джерела та анатомія мовного тракту. Разом з тим, розрізняльна здатність такого опису обмежена, і тому значні зусилля сконцентровані на розробці вирішальних правил. Найбільш популярні методи гаусівських сумішей (GMM) і опорних векторів (SVM). Використовуються також штучні нейронні мережі і приховані Марковськими моделі (HMM). З метою порівняння різних методів діаризація диктора введений показник рівний помилки (EER), що визначає помилку розпізнавання за умови рівності ймовірності пропуску самозванця і відмови законному користувачеві.

Результати дослідження

Задачу діаризації, як згадувалось раніше можна розділити на верифікацію (перевірку) диктора (SV) та ідентифікацію диктора (SID). Метою верифікації диктора є перевірка особи за допомогою голосу суб'єкта. При ідентифікації диктора, метою є забезпечення ідентичності суб'єкта. Розпізнавання гучномовців має багато реальних додатків, включаючи аутентифікацію користувачів, контроль доступу, а також допомогу при розділенні та розпізнаванні мовлення. Недавній успіх глибоких нейронних мереж (DNN) для автоматичного розпізнавання мови (ASR) мотивував застосування DNN до розпізнавання диктора. Деякі методи готують DNN для безпосереднього розпізнавання диктора і використовуються в основному для розпізнавання диктора за сталими

парольними фразами. Інші підходи використовує DNN, підготовлену для ASR за допомогою універсальної моделі фону (UBM) для і-векторної системи, підвищуючи її здатність захоплювати шаблони вимови. Це призводить до значного поліпшення точності діаризації в порівнянні з традиційною гаусівською моделлю суміші (GMM) UBM. Однак ці дослідження спрямовані на відносно лабораторні умови. См створили фонетично адаптивну систему i-vector в умовах, коли вимова спотворена сильним адитивним шумом. Запропоновано алгоритм аналізу вимови для видалення або послаблення фонового шуму. Зокрема, DNN навчається для оцінки маски, і маска, створена DNN, використовується для вилучення спотворених складових вимови (Рис. 1).

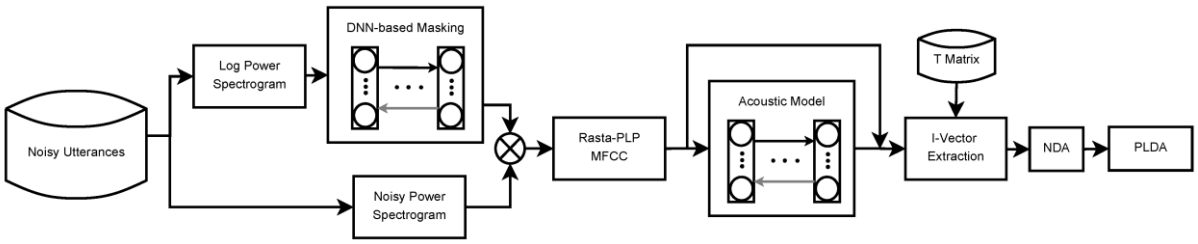

Рис. 1. Архітектура системи діаризація на основі DNN та і-векторів

В створеній системі для вилучення і-векторів ми використовували матрицю 400-вимірної тотальної мінливості, підготовлену з алгоритмом максимізації очікування правдоподібності. Зголдом методами факторного аналізу ми зменшували розмірність кожного вектору до 200 елементів. Далі і-вектори за допомогою моделі ймовірнісного лінійного дискримінантного аналізу (PLDA) готувалися для подачі на глибоку нейромережу DNN. Зауважимо, що кожен фрейм вимови описувався 12 коефіцієнтами MFCC. Результати створеної системи діаризації вимови представлені в Таблиці 1.

Таблиця 1. – EER (%) та ACC (%) для різних архітектур GMM, DNN, і-вектор та маскування

SNR	**-3 dB**		**3 dB**		**8 dB**		**12 dB**		**Average**	
System	ACC	EER	ACC	EER	ACC	EER	ACC	EER	ACC	EER
GMM/i-vector	36.08	11.47	63.27	6.75	67.14	5.97	68.56	5.67	58.76	7.47
GMM/i-vector + Masking	53.61	9.02	75.39	5.41	79.38	4.90	79.61	4.64	72.00	5.99
DNN/i-vector	37.37	11.38	66.75	5.90	72.68	5.78	72.42	5.03	62.31	7.02
DNN/i-vector + Masking	59.28	7.47	76.03	4.60	79.77	4.12	80.28	4.07	73.84	5.07

Висновки

Отже, авторами розроблено консольний додаток для діаризації мовних сигналів на мові python. В основу додатка покладено попередньо створена загальна схема голосової біометрії GMM+i-vec+DNN. Точність діаризації для нашого додатку на вибірці 80 зразків 40-ка різних дикторів склала 93%. Щодо швидкості, то середня тривалість обробки голосу при навчанні системи склала 22 секунди (оброблювалося файл з вимовою тривалістю 20 секунд).

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Аграновский А.В., Леднов Д.А. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов. Москва: Изд-во "Радио и связь", 2004. - 164 с.

2. Keshet J., Bengio S. Automatic Speech and Speaker Recognition. Large Margin and Kernel Methods. Wiley, 2009, - 257 p.

3. Столбов М.Б., Кассу А.-Р.М. Цифровая обработка речевых сигналов. Учебно-методическое пособие по лабораторному практикуму – СПб: НИУ ИТМО, 2016. – 71 с.

Матіящук Артем Володимирович — студент групи ЗАКІТ-20м, факультет комп'ютерних систем і автоматики, Вінницький національний технічний університет, Вінниця, e-mail: matiyashukartem@gmail.com

Таранюк Юлія Ярославівна — студентка групи ЗАКІТ-20м, факультет комп'ютерних систем і автоматики, Вінницький національний технічний університет, Вінниця, e-mail: tysvnyr@ukr.net

Науковий керівник: ***Никитенко Олена Дмитрівна*** — к.т.н, доцент кафедри комп'ютерних систем управління, Вінницький національний технічний університет, м. Вінниця

Matiiashchuk Artem V. – student of group 3AKIT-20m, Faculty of Computer Systems and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: matiyashukartem@gmail.com

Taranyuk Yuliya Y. – student of group 3AKIT-20m, Faculty of Computer Systems and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: tysvnyr@ukr.net

Supervisor: ***Nikitenko Olena D.*** - Cand. Tech. Sciences, Associate Professor of Computer Control Systems Department, Vinnytsia National Technical University, Vinnytsia