

АВТОМАТИЗАЦІЯ КРИМІНАЛІСТИЧНИХ ДОСЛІДЖЕНЬ МОВЛЕННЄВИХ СИГНАЛІВ

В.М.Магера І.І.Горбань, С.В.Левий

Державний науково-дослідний експертно-криміналістичний центр МВС України, Україна, 01024, м. Київ, вул. Богомольця, 10, тел./факс.: 291-39-53.

Іститут проблем математичних машин і систем Національної Академії наук України, Україна, 01187, м. Київ, проспект Глушкова, 42, електронна пошта: gorban@immsp.kiev.ua, Тел.: 38(044) 266-61-74, факс.: 38(044) 446-8129.

Галузева науково-дослідна лабораторія методів і засобів спеціального призначення радіотехнічного факультету Національного технічного університету України, кафедра радіотехнічних приладів і систем, Україна, 01056, м. Київ, пр-т Перемоги, 37 (2103), електронна пошта: Levyi@immfr.kiev.ua, тел./факс.: 274-89-84.

АНОТАЦІЯ

Описано програмно-апаратний комплекс та програму, що дозволяють досліджувати мовленнєві сигнали у автоматичних та автоматизованих режимах. Висока ефективність досліджень досягається на сигналах із рівнем спотворення до 35 дБ та відношенням сигнал-шум – до 10-12 дБ. Наводиться методика і результати тестування програм.

ВСТУП

Сучасний стан розвитку апаратури запису мовлення вимагає адекватного розвитку засобів і методик криміналістичних досліджень. Програмно-апаратні засоби, які зараз використовуються для криміналістичних досліджень, вже не задовольняють сучасним вимогам. Наприклад, програми для запису та редагування звукових сигналів "Cool Edit Pro" фірми "Syntrillium Software Corporation" США [1] і "Sound Forge 4.5" фірми "Sonic Foundry, Inc." США [2] дозволяють, головним чином, здійснювати редагування звукових сигналів та знижувати рівень шумів. Однак дослідження характеристик звукових сигналів в них представлено дуже слабо. Апаратно-програмні засоби "Computerized Speech Lab" фірми "Kay Elemetrics Corp." США [3] і "SIS" – ТОВ "Центр речевих технологій" м. Санкт-Петербург [4] дозволяють всебічно досліджувати різні характеристики мовленнєвих сигналів, проводити статистичний аналіз, але вони мають низький рівень автоматизації проведення досліджень. Програма "Диалект" м. Москва [5] дозволяє в автоматизованому режимі проводити ідентифікаційний аналіз мовленнєвих сигналів. Але великий обсяг підготовчих операцій і оцінка деяких

проміжних результатів обчислень покладається на оператора.

Зазначені програмні засоби та деякі інші демонструють гарні результати лише при роботі в умовах малих частотних спотворень і низького рівня перешкод. Коли деструктивні фактори значні, ці системи не ефективні.

Намагання створити нове покоління систем дослідження мовлення істотно більш ефективних, ніж існуючі, привело до появи нових ідей, методів і алгоритмів обробки, орієнтованих на підвищення стійкості роботи в умовах спотворень і зашумлення мовленнєвих сигналів [6-12]. Аналіз і тестування декотрих нових алгоритмів розпізнавання особи по мовленню виявили загальний їх недолік: алгоритми достатньо ефективні, коли частотні спотворення не перевищують 10-15 дБ і спектр реальної завади співпадає з очікуваним. У наведеному нижче матеріалі описані дві системи, що в значній мірі позбавлені вказаних недоліків. Це автоматизований програмно-апаратний комплекс "Phonograph" та автоматична система CASVI (Crime-detection Automatic Speaker Verification and Identification System).

1. Автоматизований програмно-апаратний комплекс "Phonograph"

Автоматизований програмно-апаратний комплекс "Phonograph" призначений для криміналістичних досліджень в мовленнєвому діапазоні частот: голосу, сигналів акустичної і неакустичної природи виникнення, шумів і завад. Він входить до складу автоматизованої криміналістичної системи "Логос" для дослідження матеріалів та засобів звуко- і відеозапису [13].

Комплекс складається із персонального комп'ютера типу Pentium II, звукової карти ExpertColor MED 3201 InterWave AMD

2. Система CASVI

Система CASVI забезпечує ідентифікацію і верифікацію особи за голосом у автоматичному режимі. Загальний опис цієї програми і її можливостей наведено в роботах [14-17]. Система ефективного вирішує задачі, що на неї покладаються, при наявності великих частотних спотворень і зашумленні мовленнєвих сигналів. Головним її елементом є підсистема ідентифікації (рис. 2). Вона складається із системи функціональних програм (СФП), системи управління (СУ) і бази даних (БД).

СФП забезпечує обробку і порівняння сигналів і включає програму розрахунку спектрів (РС), програму розрахунку інформаційних ознак (РІО) і програму порівняння інформаційних ознак (ПІО).

СУ виконує задачі взаємодії з користувачем, управління процесами обробки і порівняння фонограм, формування і поповнення бази даних. Вона містить програму, яка забезпечує інтерфейс для користувача (ІК), програму управління функціональними програмами (УФП) і програму управління базою даних (УБД).

Час, необхідний для прийняття рішення підпрограмою, визначається режимами роботи і можливостями комп'ютеру і може становити від декілька хвилин в простих випадках до десятків хвилин.

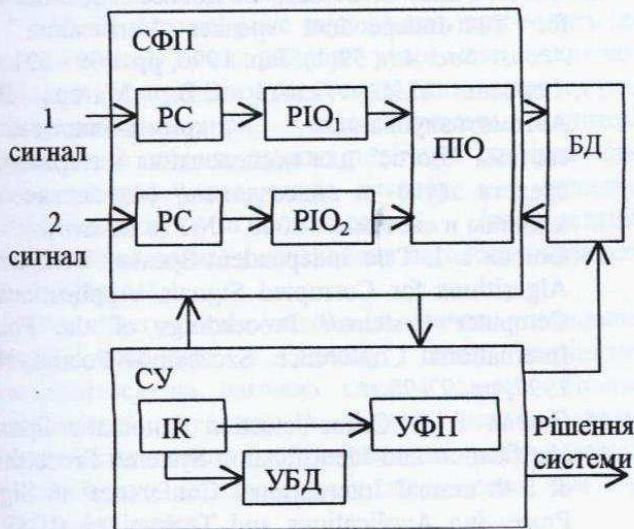


Рис. 2. Структура підпрограми ідентифікації CASVI

Функціональні програми реалізують кепстральні алгоритми.

РС проводить розділення запису на окремі короткі фрагменти (фрейми), для кожного фрейму розраховує спектр, статистичну характеристику корисного сигналу і завади та сортує фрейми на сигнальні і завадові. РІО реалізує формування сигнальних блоків фреймів і розрахунок інформаційних ознак: центрування кепстрів, а також перші і другі похідні кепстрів.

ПІО проводить розрахунок відстані між записами по кожній інформаційній ознаці окремо, розрахунок оцінок щільностей імовірності для однакових і різних голосів із БД, формування підсумкової відстані між голосами, формування

порогу з використанням отриманих оцінок щільностей імовірності, порівняння підсумкової відстані з порогом, формування рішення про ідентичність голосів і розрахунок оцінки імовірності помилкового рішення.

В СУ головними є програми, які забезпечують ІК і УФП. Інтерфейс користувача СУ складається із вікон головного меню, поточного результату та бази даних.

Головне меню дозволяє вибрати файли, оцінити рівень шумів фонограм, прослухати їх і виконати ідентифікацію по одному із описаних вище варіантів.

Вікно поточного результату надає результат ідентифікації у вигляді рішення системи Yes/No (да/нет) і розрахункову імовірність помилки прийнятого рішення. Воно інформує користувача про рівень шумів фонограм.

Вікно бази даних дозволяє відібрати файли для БД, прослухати їх і запустити розрахунок спектрів і кепстральних ознак, які використовуються при ідентифікації.

Підсистема ідентифікації CASVI розрахована на роботу з тексто-незалежним матеріалом. Але, враховуючи низьку якість запису реальних фонограм, які надходять на експертизу, в окремих випадках доцільно використовувати для порівняння заздалегідь підготовлений матеріал. У цьому випадку з допомогою комплексу "Phonograph" у автоматизованому режимі може проводитись відбір подібних наголошених голосних, що стоять між приголосними звуками. Із них формуються масиви для порівняння з допомогою підсистеми ідентифікації CASVI.

Програма орієнтована на операторів з нижчим ступенем кваліфікації, ніж для роботи з комплексом "Phonograph", так як більшість технологічних операцій повністю автоматизована. Це зменшує витрати на спеціальну підготовку експертів.

3. Тестування працездатності комплексу "Phonograph" та системи CASVI

Тестування програмно-апаратного комплексу "Phonograph" здійснювалось згідно з програмою, затвердженою Секцією експертизи матеріалів та засобів звуко- і відеозапису Науково-методичної ради Міністерства юстиції України. Для кожного окремого режиму чи розрахунку були сформовані спеціальні сигнали та входні дані, на яких перевірялась працездатність комплексу. Тестування показало відповідність усіх параметрів розрахунковим і дієздатність всієї системи в цілому. Комплекс "Phonograph" був розглянутий на Науково-координаційній раді МВС України та рекомендований до впровадження у експертну практику.

Алгоритми ідентифікації CASVI тестувались на 32 п'ятихвилинних записах чоловічих голосів. Відношення сигнал-завада для всіх записів складало

не менше 30 дБ. Записи були розділені на дві однакові групи, до яких входили по два комплекти записів восьми дикторів. Записи першої групи використовувались як зразки, а другої - як робочі. Ефективність роботи досліджувалась при сильних частотних викривленнях сигналу і зашумленні, типових при передачі повідомлення по телефонному каналу. Для цього перший комплект робочих записів викривлявся по частоті і зашумлявся на кожній частоті до фіксованого рівня.

Експерименти проводились в два етапи: навчання і тестування. На етапі навчання формувались дані для БД з використанням зразкових записів. Ефективність алгоритмів перевірялась на етапі тестування з використанням робочих записів. Число спроб в разі однакових дикторів було 72, у випадку різних - 504. Тривалість записів дорівнювалась 12 секундам.

Результати експериментів показали, що в умовах частотних викривлень до 35 дБ і зашумленні до відношення сигнал-завада 12 дБ імовірність правильної ідентифікації становить 90% при імовірності помилкової тривоги 10%.

Висновки

Результати тестування програмно-апаратного комплексу "Phonograph" та системи CASVI свідчать про їх ефективну роботу при дослідженні мовленнєвих сигналів.

Програмно-апаратний комплекс "Phonograph" є більш дослідницькою програмою і дозволяє виконувати всі види акустичних та просодичні і фонетичні лінгвістичні дослідження. Значна частина технологічних операцій та режимів автоматизована, що значно скорочує час виконання експертиз.

Практично всі функції підпрограми ідентифікації CASVI працюють у автоматичному режимі при значних частотних викривленнях (до 35 дБ) і відношенню сигнал-шум 12 дБ. Наведені результати суттєво перевищують ті, що до цього часу було досягнуто. Значне покращання технічних параметрів стало можливим завдяки комплексному підходу до проблеми і використанню нових робастних алгоритмів обробки, оптимізованих для роботи в несприятливих умовах.

Комплекс "Phonograph" та система CASVI за результатами порівняння записів голосів наводять не тільки рішення про їх ідентичність, але і оцінку імовірності помилки прийнятого рішення.

ЛІТЕРАТУРА

1. <http://www.syntrillium.com>
2. <http://www.sfoundry.com/index.html>
3. Новосельский А.Ф. Компьютерная система для всесторонней работы с речью CLS MODEL 4300B// Компьютеры+Программы. - 1995. - №7(22). - С.66-71.
4. Рекомендации по эффективному использованию возможностей АРМЭФ SIS при выполнении криминалистических экспертиз. - Центр речевых технологий. - Санкт-Петербург, 1995.
5. Идентификация лиц по фонограммам русской речи на автоматизированной системе "Диалект" / Попов Н.Ф., Линьков А.Н., Кураченкова Н.В., Байчаров Н.В. / Под ред. Фесенко А.В. - М.: В/ч 34435, 1996. - 102 с.
6. Hansen J. H. L., Mammone R. J., Young S. editors// IEEE Transactions on Speech and audio Processing. - 1994. - October.
7. Mansour D. and Juang B. H. The short-time modified coherence representation and noisy speech recognition// IEEE Trans. Acoust., Speech, Signal Processing. -1989. -37. - June. - P. 796-804.
8. Newney L. and Weintraub M. Probabilistic Optimum Filtering for Robust Speech Recognition// Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing. - 1994. -1. - P. 417-420.
9. Reynolds D. A. and Rose R. C. Robust Text-independent Speaker Identification Using Gaussian Mixture Speaker Models// IEEE Trans. Speech, Audio Processing. -1995. - 3. - January. - P. 72-83.
10. Nolasco Flores J. A., Young S. J. Continuous Speech Recognition in Noise Using Spectral Subtraction and HMM Adaptation// Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing. - 1994. -1. - P. 409-412.
11. J. Mammone, X. Zhang, R. P. Ramachandran, "Robust Speaker Recognition," *IEEE Signal Processing Magazine*, Sept. 1996, pp. 58 - 71.
12. M. A. Lund, C. C. Lee, "A Robust Sequential Test for Text-Independent speaker Verification," *J. Acoust. Soc. Am.* 99(1), Jan. 1996, pp. 609 - 621.
13. Горбань И.И., Левый С.В., Марепа В.Н. Автоматизированная криминалистическая система "Логос" для исследования материалов и средств звуко- и видеозаписи// Математические машины и системы. -2000. -№3 (в печати).
14. Gorban I. I. Text Independent Speaker Verification Algorithms for Corrupted Signals. Applications of Computer Systems// Proceedings of the Fourth International Conference. Szczecin - Poland, Nov, 1997, pp. 92-98.
15. Gorban I. I. Crime-detection Automatic Speaker Verification and Identification System// Proceedings of 9-th annual International Conference on Signal Processing Applications and Technology (ICSPAT 98). - 1998. - August.
16. Горбань И.И., Горбань Н.И., Клименко А.В., Хазанович М.С. Подсистема верификации новой криминалистической системы автоматической верификации и идентификации личности по голосу (CASVI)// Математические машины и системы. - 1997. -№2. -С.61-64.
17. Горбань И.И., Клименко А.В. Робастні алгоритми верифікації особи за голосом, що призначені для роботи в умовах сильних завад та спотворень мовних повідомлень// 4 Всеукраїнська міжнародна конференція по обробці та розпізнаванню образів "Укробраз 98". - 1998.