

Робастні алгоритми верифікації особи за голосом, що призначені для роботи в умовах сильних завад та спотворень мовних повідомлень

І. І. Горбань, А. В. Клименко

Інститут проблем математичних машин і систем Національної академії наук України

252187, Україна, Київ, Пр. Академіка Глушкова, 42

Тел. (044) 2666174, Факс (044) 4468129, E-mail: gorban@immsp.kiev.ua, klimenko@immsp.kiev.ua

У цій доповіді представлені нові робастні алгоритми обробки мовних сигналів, спрямовані на підвищення якості роботи алгоритмів автоматичної верифікації особи за голосом в умовах сильних частотних спотворень сигналу та потужних завад. Доповідь складається з чотирьох розділів. У вступній частині наведено обґрунтування необхідності застосування робастних алгоритмів обробки сигналу для здійснення ефективної верифікації в реальних умовах експлуатації систем автоматичної верифікації особи за голосом. Описання робастних алгоритмів верифікації особи за голосом наведено в другому розділі. Третій розділ присвячено результатам тестування цих алгоритмів. Краткі висновки наведено у четвертому розділі.

1. Вступ

Умови експлуатації систем верифікації особи за голосом дуже часто далекі від ідеальних. Мовні повідомлення, за якими здійснюється верифікація, як правило, сильно спотворені та зашумлені. Більшість відомих алгоритмів верифікації здатні працювати лише в умовах незначних завад та частотних спотворень [1-4].

В процесі розробки нової криміналістичної автоматичної системи верифікації та ідентифікації особи за голосом (CASVI) [5-8] авторами цієї доповіді були запропоновані і досліджені нові робастні алгоритми верифікації, що забезпечують стійку роботу в умовах сильних частотних спотворень сигналу та адитивних завад.

Мета доповіді – представити ці алгоритми і описати результати їх тестування.

2. Опис алгоритмів

2.1. Принципова схема роботи системи

При верифікації основними етапами обробки сигналів є: попередня обробка порівнюваних повідомлень (ПОП), розрахунок інформаційних

ознак (РО), порівняння цих ознак (ПО) і прийняття рішення (ПР). На етапі порівняння інформаційних ознак та прийняття рішення використовуються дані з бази даних (БД) (Рис. 1)

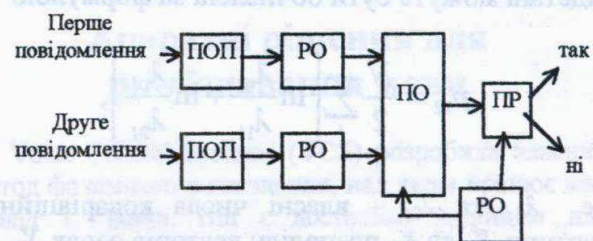


Рис. 1. Принципова схема роботи підсистеми верифікації системи CASVI

2.2. Попередня обробка повідомлень

На етапі попередньої обробки проводиться розбиття кожного повідомлення на фрейми тривалістю 20 мс. Ці фрейми розділяються системою на сигнальні, що підлягають подальшій обробці, і шумові, що виключаються. Ця процедура виконується на основі розрахунку оцінок верхньої та нижньої меж спектру сигналу і робастної обробки спектрів. Завдяки цим алгоритмам забезпечується відбір ідентичних фреймів як при відсутності так і при наявності спотворень та завад.

2.3. Розрахунок інформаційних ознак

Розрахунок інформаційних ознак передбачає для кожного фрейма обчислення 16-ти кепстральних компонент, а також двох перших кепстральних похідних. Для виключення залежності кепстрів від частотних спотворень сигналу введено їх центрування на вектор, що являє собою кепстр оцінки верхньої межі спектру сигналу. Для виключення розбіжностей в ознаках, що зумовлені різними шумовими умовами, введено нормування кепстрів. Суть нормування кепстрів полягає в штучному підвищенні рівня завад в кожному фреймі до

відношення сигнал-завада 12 дБ на кожній частоті з подальшим розрахунком кепстральних компонент.

2.4. Порівняння ознак та прийняття рішення

Порівняння ознак полягає в розрахунку відстані між порівнюваними повідомленнями. Ця відстань визначається шляхом об'єднання відстаней між окремими інформаційними ознаками.

Базою для вимірювання відстаней між ознаками обрано відношення правдоподібності. При цьому припускається, що ознаки мають гаусовий розподіл. При такому припущенні відстані можуть бути обчислені за формулою

$$p_{12} = \frac{N}{2} \sum_l \left[\ln \frac{\lambda_{0l}}{\lambda_{1l}} + \ln \frac{\lambda_{0l}}{\lambda_{2l}} \right],$$

де λ_{1l} та λ_{2l} – власні числа коваріаційних матриць K_1 та K_2 розподілу векторів ознак X_{1n} та X_{2n} ($n = 1, \dots, N$) відповідно для 1-го та 2-го повідомлень. λ_{0l} – власні числа коваріаційної матриці K_0 сукупного розподілу векторів X_{1n} та X_{2n} ($n = 1, \dots, N$), l – порядковий номер кепстральної компоненти, N – кількість сигнальних фреймів, що обробляються.

Об'єднання одержаних відстаней здійснюється за допомогою бази даних.

Рішення приймається при порівнянні сукупної відстані між повідомленнями і порогом, що розрахований за допомогою бази даних.

3. Результати тестування алгоритмів

Для проведення комплексного тестування алгоритмів верифікації використовувались 4 набори записів голосів 8 дикторів (усі диктори чоловічої статі) тривалістю від 1 до 2,5 хв. Використовувались записи, зроблені в разний час з перервами між записами від 2 тижнів до 3 місяців.

Базова тривалість фрагментів, по яких проводилась верифікація – 11,6 с чистого сигналу.

Перед початком проведення тестування робочі записи були спотворенні до 35 дБ, а також зашумлені адитивною завадою до відношення сигнал-завада 12дБ.

Всього було проведено 972 тести, з них 849 для різних дикторів і 123 для однакових. При

цьому ймовірність правильного прийняття рішення про співпадіння порівнюваних голосів виявилася рівною 93%, при цьому помилка ложної тривоги зафіксована на рівні 10%.

4. Висновки

Проведені дослідження дозволили сформулювати алгоритми верифікації для системи CASVI. Завдяки використанню нових процедур була забезпечена ефективна робота системи в складних умовах: при наявності великих спотворень й сильних завад.

Література

1. B. S. Atal, "Automatic Recognition of Speakers from Their Voices," *Proc. IEEE*, 64.– April.– 1976.–P. 460-475.
2. S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification," *IEEE Trans. ASSP*, 29(2).–April.–1981.–P. 254-272.
3. H. Gish, M. Shmidt, "Text-Independent Speaker Identification," *IEEE Signal Processing Magazine*, Oct.–1994.–P. 18-31.
4. I. I. Gorban, "Crime Automatic Speaker Verification and Identification (CASVI) System," *134th Meeting of ASA*, 102(5).–Pt.2.– Nov.–1997.–P. 3165.
5. J. Mammone, X. Zhang, R. P. Ramachandran, "Robust Speaker Recognition," *IEEE Signal Processing Magazine*, Sept.–1996.–P. 58-71.
6. И. И. Горбань, Н. И. Горбань, А. В. Клименко, М. С. Хазанович, "Подсистема верификации новой криминалистической автоматической системы верификации и идентификации личности по голосу (CASVI)," *Математические машины и системы*, 2.–1997.–С. 61-64.
7. I. I. Gorban, "Text-Independent Speaker Verification Algorithms for Corrupted Signals," *Proc. Of 4th International Conference on Application of Computer Systems (ACS'97)*.–1997.–Oct.–P8.
8. I. I. Gorban, "Crime-Detection Speaker Verification and Identification System," *Proc. Of 9th annual International Conference on Signal Processing Application and Technology (ICSPAT'98)*.–1998.–Sept.–P5.