

РОЗПІЗНАВАННЯ ВІКУ ТА СТАТІ ЗА ГОЛОСОМ ЛЮДИНИ В УМОВАХ ДІЇ КОРЕЛЬОВАНИХ ЗАВАД

Омельченко Сергій Васильович,

к.т.н., доцент

Харківський національний університет радіоелектроніки,
м. Харків, Україна

Вступ. Визначення віку та статі людей може використовуватись як допоміжна інформація при вирішенні ряду задач. Ефективність реклами та діалогових довідкових систем за телефоном зростає, якщо вона враховує вік та стать людей. Знання віку та статі людини може використовуватися як додаткова інформація для аутентифікації та верифікації людей, розпізнавання слів мови.

Однак відомі методи визначення віку та статі людей не є достатньо надійними в умовах дії завад різного типу.

Мета роботи. Метою роботи є створення методів визначення статі та віку людини за її мовою, які є стійкими до дії корельованих завад.

Матеріали і методи. Важливим для вирішення задачі визначення віку та статі людей за мовою є вибір стійких до дії завад ознак, які застосовуються для класифікації, та використання методів придушення завад [1-3].

Для захисту від завад ефективним є гранчастий фільтр. Коефіцієнти відбиття такого фільтру обчислюють за відліками вхідного сигналу x_n при відсутності розмови та присутності завад.

Коефіцієнти відбиття оцінюємо використовуючи помилки передбачення d^{m_n} , b^{m_n} наступним чином

$$k_m = \frac{\sum_{n=1}^N (d^{m-1}_n \cdot b^{m-1}_{n-1})}{\left(\sum_{n=1}^N (d^{m-1}_n)^2 \cdot \sum_{n=1}^N (b^{m-1}_{n-1})^2 \right)^{1/2}},$$

де $m = 1, 2, \dots, p$.

Начальні значення задаємо наступним чином

$$d^0_n = x_n, b^0_n = x_n, n = 1, \dots, N.$$

Помилка прямого передбачення обчислюється як

$$d^m_n = d^{m-1}_n - k_m \cdot b^{m-1}_{n-1},$$

Тоді помилка зворотнього передбачення буде знаходитися

$$b^m_n = b^{m-1}_{n-1} - k_m \cdot d^{m-1}_n.$$

На рис. 1 показано пристрій для обчислення коефіцієнтів відбиття.

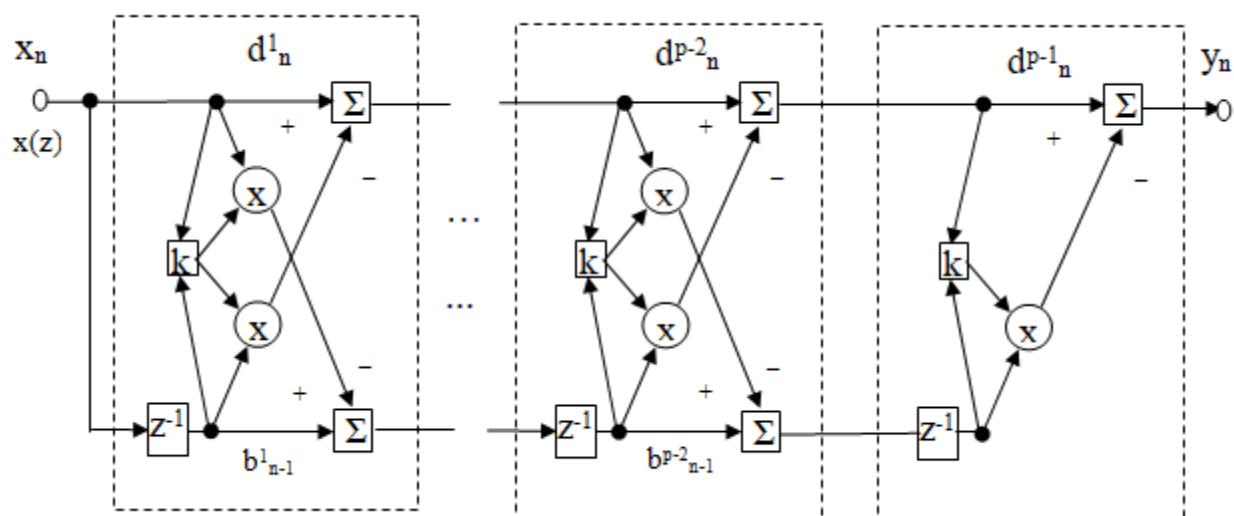


Рис.1. Гратчатий фільтр для оцінювання коефіцієнтів відбиття

Результат роботи фільтра буде

$$y_n = d^p_n .$$

Отримані оцінки коефіцієнтів відбиття $\vec{k} = (k_1, k_2, \dots, k_p)$ цифрового фільтра застосовують для придушення завад фільтром наведеним на рис.2.

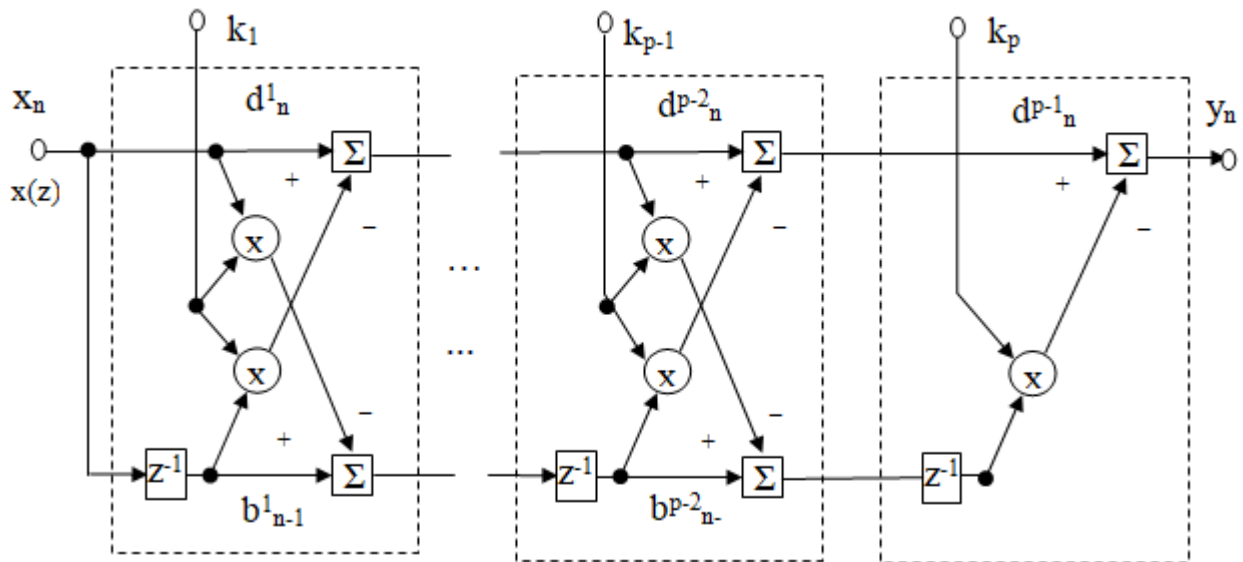


Рис.2. Гратчастий фільтр для захисту від корельованих завад

Статева та вікова класифікація людей може використовувати у якості ознак сукупність різномірних оцінок мовного сигналу на основі моделей слуху та мовотворення.

Вектор ознак для статевої та вікової класифікації включає в себе Мел-кепстральні частотні коефіцієнти (MFCC) які раціонально поєднувати разом з моментними функціями частоти основного тону та формантними ознаками.

Вхідний цифровий сигнал розбивається на часові відрізки однакової тривалості, що дозволяє отримати сукупність оцінок ознак, що змінюються в часі.

Потім виконується сегментація, що включає в себе прийняття рішення про виявлення сигналу мовлення та знаходження часових меж початку та закінчення мовлення. Це дозволяє на етапі розпізнавання виключити фрагменти, що не несуть мовну інформацію.

Особливості розподілу частоти основного тону можна використовувати для розпізнавання статі та віку дикторів. Оцінки частоти основного тону для дикторів-чоловіків мали середню частоту 128 Гц з діапазоном можливих значень від 58 Гц до 238 Гц, а для жінок середню частоту - 256 Гц з її зміною в інтервалі від 135 Гц до 522 Гц.

Гістограми які знайдено за сукупністю послідовних в часі оцінок частоти основного тону є несиметричні - щодо своєї моди: у жіночих голосів з боку малих періодів схил крутіший, ніж для великих періодів, тоді як у чоловіків спостерігається зворотня картина. Для розпізнавання статі та віку раціонально використовувати нормовані центральні моменти частоти основного тону аж до 6-го порядку включно.

Нормовані центральні моменти γ_k визначаються таким чином:

$$\gamma_k = \frac{\mu_k}{\mu_2^{k/2}}.$$

де μ_k - центральні моменти.

Так коефіцієнт асиметрії частоти основного тону обчислюється за формулою

$$As = \frac{\mu_3}{\sigma^3},$$

а ексцес частоти основного тону відповідає формулі

$$Es = \frac{\mu_4}{\sigma^4} - 3,$$

де σ – середньоквадратичне відхилення.

Для класифікації людей за статтю та віком можна використовувати різні класифікатори. Це можуть бути : моделі гауссових сумішей, регресія опорного вектора (SVR) , векторну машину підтримки (SVM), k-найближчих сусідів (k-NN).

Результати та обговорення. Вирішення цієї задачі потребує формування бази даних еталонів класів, при цьому в нашому підході застосуємо для класифікації три класи, а зокрема дорослих чоловіків, дорослих жінок та дітей без гендерної дискримінації;

В експериментальних дослідженнях, які отримані для оцінювання віку та статі, використовували 40 зразків мови. У модельних класах (літні, дорослі та діти обох статей) система визнання статі досягла 66% точності розпізнавання. Тоді як вікове визнання досягало 55% точності розпізнавання при використанні лише в якості ознак MFCC (від 4 до 9 коефіцієнта).

Процент правильного розпізнавання віку лише з використанням формантних частот був 50 %, а правильного розпізнавання статі лише з використанням формантних частот (без урахування дітей) становив - 91%.

При використанні функції частот формантів найкращий результат розпізнавання віку був отриманий за допомогою формантів 1 і 2, а найкращий результат для розпізнавання статі, отриманий при використанні всіх чотирьох формантів разом.

Застосування в якості ознак MFCC та оцінок формантних частот поєднане з оцінками частоти основного тону та її нормованими моментами дозволило досягти процент правильного розпізнавання статі 97 %.

Висновки. Створено нові завадостійкі методи визначення статі та віку людини за їх мовою, які сумісно використовують в якості ознак MFCC, оцінки формантних частот та оцінки частоти основного тону та її нормовані моменти. Показана необхідність використання придушення завад. Проведені експериментальні дослідження підтверджують ефективність розроблених алгоритмів.

Список літератури

1. Пресняков І.М., Омельченко О.В., Омельченко С.В. Автоматическое распознавание речи в каналах передачи //Радиоэлектроника и информатика. 2002. №1.С. 26-31.
2. Omelchenko S. Development of the method of Automatic determination of the speaker Gender on the basis of joint evaluation of frequency moments of basic tones and formant frequencies// TECHNOLOGY AUDIT AND PRODUCTION RESERVES. 2018. 3 (2(41)). P.29-33.
3. Омельченко С.В. Алгоритмы сегментации речевого сигнала на фоне коррелированной помехи//ScienceRise 4, Scientific Journal «ScienceRise». 2018. №4(45). С. 31-35.