

УДК 681.327.12

М. М. Биков, к. т. н., проф.; В. В. Ковтун, к. т. н., доц.; Н. Г. Савінова
ОЦІНЮВАННЯ ВПЛИВУ ЗАВАД НА ДОСТОВІРНІСТЬ РОБОТИ
ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНОЇ СИСТЕМИ РОЗПІЗНАВАННЯ
ГОЛОСУ

Оцінено вплив завад у мовленнєвому сигналі на достовірність роботи інформаційно-вимірювальної системи розпізнавання голосу. Отримано аналітичні вирази для оцінювання помилки розпізнавання, наведено результати обчислення імовірності помилки класифікації двох класів голосів за різних рівнів шумів у сигналі.

Ключові слова: розпізнавання голосу, інформаційно-вимірювальна система, достовірність, шум, завади, математична модель, передатний канал, розподільна гіперплощина, нормальний закон розподілу, правило мінімуму відстані.

Передатний канал, через який відбувається поширення мовленнєвих сигналів, які використовуватимуться для розпізнавання голосу, знаходиться під впливом різного роду шумів [1], основними серед яких є шуми апаратури та навколишнього середовища. Указаний вплив призводить до зменшення достовірності роботи інформаційно-вимірювальної системи розпізнавання голосу. Отже, визначення цього впливу з урахуванням конструктивних особливостей інформаційно-вимірювальної системи цієї категорії є актуальним завданням.

Скориставшись лінійною моделлю [2, 3, 4], мовленнєвий сигнал можна розглядати як квазідетермінований процес для одних і тих же самих класів на часовому відрізку $\tau = 10 \div 30$ мс. За наявності адитивні завади, мовленнєвий сигнал матиме такий вигляд:

$$y^*(t, \tau) = y(t, \tau) + \xi(t, \tau) \quad (1)$$

Здебільшого завада $\xi(t, \tau)$ має нульове середнє значення та некорельована з мовленнєвим сигналом $y(t, \tau)$. Отже, вектор сигналу можна показати як:

$$X = X_y + X_\xi, \quad (2)$$

де $X_y = (x_{y_1}, x_{y_2}, \dots, x_{y_n})^T$, $X_\xi = (x_{\xi_1}, x_{\xi_2}, \dots, x_{\xi_n})^T$ – вектори опису мовленнєвого сигналу та завади відповідно;

T – символ транспонування;

$x_i = x_{y_i}^2 + x_{\xi_i}^2$ – спектральна потужність на i -й частотній смузі.

Для нівелювання впливу на результати розпізнавання гучності мовленнєвого сигналу, використаємо нормований вектор опису

$$\|X\| = \sum_{i=1}^n x_i = \sigma_c^2, \quad (3)$$

де σ_c^2 – дисперсія на вході системи розпізнавання.

У результаті нормування вектора (2) відповідно з (3) одержимо:

$$\tilde{X} = \frac{r^2}{r^2 + 1} \tilde{X}_y + \frac{1}{r^2 + 1} \tilde{X}_\xi,$$

де \tilde{X} , \tilde{X}_y , \tilde{X}_ξ – нормовані вектори X , X_y , X_ξ відповідно;

$r = \sqrt{\frac{\sigma_y^2}{\sigma_\xi^2}}$ – відношення «сигнал/шум» на вході розпізнавальної системи.

Враховуючи, що на практиці $r^2 \gg 1$ рівняння матиме вигляд:

$$\tilde{X} = \tilde{X}_y + \frac{1}{r^2} \tilde{X}_\xi. \quad (4)$$

Достовірність роботи класифікатора інформаційно-вимірювальної системи розпізнавання голосу визначимо на прикладі двох класів мовців Ω_1 та Ω_2 , що потім не важко узагальнити на випадок більшої кількості класів. Типовим класифікатором, який використовується в інформаційно-вимірювальній системі для розпізнавання голосу, є класифікатор «за мінімумом відстані» [5]. Отже, для даного типу класифікатора правило класифікації має вигляд:

$$d_e(\tilde{X}, \mu_i) = \min d_e(\tilde{X}, \mu_j) \Rightarrow X \in \Omega_i, \quad i, j = 1, 2, \quad (5)$$

де $\mu_i = E^0\{\tilde{X} \mid \tilde{X} \in \Omega_i\}$ – середнє значення векторів \tilde{X} , які відносяться до класу Ω_i (Ω_i – еталон класу);

$d_e(\tilde{X}, \mu_i) = [(\tilde{X} - \mu_i)^T (\tilde{X} - \mu_i)]$ – евклідова відстань від вектора \tilde{X} до вектора μ_i .

У термінах гіперплощини, яка розділяє класи мовців Ω_1 та Ω_2 , правило (5) для сигналу без завад матиме вигляд:

$$H_{12}(\tilde{X}) = (\tilde{X}_y - \mu_1)^T (\tilde{X}_y - \mu_1) - (\tilde{X}_y - \mu_2)^T (\tilde{X}_y - \mu_2) = 0, \quad (6)$$

або

$$H_{12}(\tilde{X}) = 2\tilde{X}_y^T (\mu_2 - \mu_1) + \mu_1^T \mu_1 - \mu_2^T \mu_2 = 0. \quad (7)$$

Підставимо (4) в (7). Якщо є завади, сформуємо рівняння гіперплощини враховуючи, що навчання класифікатора проводилося на сигналі, в якому завади були відсутні

$$H_{12}^r(\tilde{X}) = 2\tilde{X}_y^T (\mu_2 - \mu_1) + \frac{2}{r^2} \tilde{X}_\xi^T (\mu_2 - \mu_1) + \mu_1^T \mu_1 - \mu_2^T \mu_2 = 0, \quad (8)$$

або

$$H_{12}^r(\tilde{X}) = H_{12}(\tilde{X}) + \frac{2}{r^2} \tilde{X}_\xi^T (\mu_2 - \mu_1) = 0.$$

З рівняння (8) видно, що вплив завади призводить до зміщення гіперплощини до одного з класів Ω_1 або Ω_2 залежно від знака вектора $(\mu_2 - \mu_1)$. Оскільки для квазідетермінованого мовленнєвого сигналу потужність миттєвого спектру в i -й частотній смузі є випадковою функцією [6], то спектральний вектор опису \tilde{X} є випадковим вектором, який характеризується багатовимірним нормальним розподілом. Тоді (7), (8) визначаються щільністю одновимірного нормального розподілу. Середнє значення для дискримінатної функції (7) матиме вигляд:

$$E^0\{H_{12}(\tilde{X})\} = 2E^0\{\tilde{X}_y\}(\mu_2 - \mu_1)^T + \mu_1^T \mu_1 - \mu_2^T \mu_2, \quad (9)$$

А для дискримінатної функції (1.8), приймаючи $\frac{2}{r^2} E^0\{\tilde{X}_\xi^T\} \approx \frac{2}{r^2} \tilde{X}_\xi^T$,

$$E^0\{H_{12}(\tilde{X})\} = 2E^0\{\tilde{X}_y\}(\mu_2 - \mu_1)^T + \frac{2}{r^2}\tilde{X}_\xi^T(\mu_2 - \mu_1) + \mu_1^T\mu_1 - \mu_2^T\mu_2, \quad (10)$$

Дисперсія дискримінантної функції (7) матиме вигляд:

$$\sigma_n^2 = E^0\left\{2\tilde{X}_y(\mu_2 - \mu_1)^T + \mu_1^T\mu_1 - \mu_2^T\mu_2\right\} - \left[2E^0\{\tilde{X}_y\}(\mu_2 - \mu_1)^T + \mu_1^T\mu_1 - \mu_2^T\mu_2\right]^2, \quad (11)$$

спростивши який одержимо

$$\sigma_n^2 = 2(\mu_2 - \mu_1)^T \sum_i^0 (\mu_2 - \mu_1), \quad i = 1, 2, \quad (12)$$

де \sum_i^0 – коваріаційна матриця i -го класу.

Для спрощення виразів візьмемо, що $\sum_1^0 = \sum_2^0 = \sum^0$.

Підставивши (8) в (11), одержимо рівняння дисперсії дискримінантної функції $H_{12}^r(\tilde{X})$, аналогічне рівнянню (12).

Отже, при $\tilde{X} \in \Omega_1$ дискримінантні функції $H_{12}(\tilde{X})$ та $H_{12}^r(\tilde{X})$ визначаються нормальними законами $N_1(m_1^0, \sigma_n)$ та $N_1^r\left(m_1^0 + \frac{2}{r^2}\tilde{X}_\xi^T(\mu_2 - \mu_1), \sigma_n\right)$ відповідно. Середнє значення m_1^0 одержимо, підставивши в (11) значення $E^0\{\tilde{X}_y\} = \mu_1$:

$$m_1^0 = \mu_2^T(2\mu_1 - \mu_2) - \mu_1^T\mu_1. \quad (13)$$

Аналогічно, при $\tilde{X} \in \Omega_2$, дискримінантні функції $H_{12}(\tilde{X})$ та $H_{12}^r(\tilde{X})$ визначаються нормальними законами $N_2(m_2^0, \sigma_n)$ та $N_2^r\left(m_2^0 + \frac{2}{r^2}\tilde{X}_\xi^T(\mu_2 - \mu_1), \sigma_n\right)$ відповідно. Середнє значення m_2^0 одержимо, підставивши в (9) значення $E^0\{\tilde{X}_y\} = \mu_2$:

$$m_2^0 = \mu_1^T(\mu_1 - 2\mu_2) - \mu_2^T\mu_2. \quad (14)$$

Отже, імовірність виникнення помилок першого та другого роду для дискримінантної функції (17) за умови відсутності завад визначається рівнянням

$$P(e) = P(\Omega_1)P(H_{12}(\tilde{X}) < \theta |_{\tilde{X} \in \Omega_1}) + P(\Omega_2)P(H_{12}(\tilde{X}) > \theta |_{\tilde{X} \in \Omega_2}), \quad (15)$$

де θ – поріг розпізнавання.

Враховуючи (13) та (14), одержимо:

$$P(H_{12}(\tilde{X}) < \theta |_{\tilde{X} \in \Omega_1}) = \int_{-\infty}^{\theta} \frac{1}{\sigma_n \sqrt{2\pi}} e^{-\frac{[H_{12}(\tilde{X}) - m_1^0]^2}{2\sigma_n^2}} dH_{12}(\tilde{X}), \quad (16)$$

$$P(H_{12}(\tilde{X}) > \theta |_{\tilde{X} \in \Omega_2}) = \int_{\theta}^{\infty} \frac{1}{\sigma_n \sqrt{2\pi}} e^{-\frac{[H_{12}(\tilde{X}) - m_2^0]^2}{2\sigma_n^2}} dH_{12}(\tilde{X}). \quad (17)$$

Підставивши (16), (17) в (15), одержимо:

$$P(e) = P(\Omega_1) \Phi^* \left(\frac{\theta - m_1^0}{\sigma_n} \right) + P(\Omega_2) \Phi^* \left(\frac{m_2^0 - \theta}{\sigma_n} \right), \quad (18)$$

де $\Phi^*(a)$ – функція Лапласа.

Обираючи бінарну функцію втрат (0 – правильне розпізнавання, 1 – помилка), порогове значення θ визначатиметься відношенням:

$$\frac{P(H_{12}(\tilde{X}) = \theta | \tilde{X} \in \Omega_1)}{P(H_{12}(\tilde{X}) = \theta | \tilde{X} \in \Omega_2)} = \frac{P(\Omega_1)}{P(\Omega_2)} = \theta_0. \quad (19)$$

Підставивши (15) та (16) в (19), одержимо:

$$\theta = \frac{m_1^0 + m_2^0}{2} + \frac{\sigma_n^2 \ln \theta_0}{m_1^0 - m_2^0}. \quad (20)$$

Замінюючи в (20) величини m_1^0 , m_2^0 та σ_n^2 їхніми значеннями, одержуємо порогове значення θ

$$\theta = \sum_0 \ln \theta_0. \quad (21)$$

Якщо наявні завади, порогове значення θ запишемо як:

$$\theta_\xi = -\frac{4}{r^2} X_\xi^T (\mu_2 - \mu_1) + \sum_0 \ln \theta_0 = \Delta_\xi + \theta. \quad (22)$$

Враховуючи (22), імовірність виникнення помилок першого та другого виду для дискримінантної функції (18), якщо наявні завади, матиме вигляд:

$$P(e_\xi) = P(\Omega_1) \Phi^* \left(\frac{\theta + \Delta_\xi - m_1^0}{\sigma_n} \right) + P(\Omega_2) \Phi^* \left(\frac{m_2^0 - \Delta_\xi - \theta}{\sigma_n} \right). \quad (23)$$

Рівняння (23) визначає залежність похибки розпізнавання від наявності в мовленнєвому сигналі шуму. До переваги запропонованого способу врахування впливу наявних у мовленнєвому сигналі шумів на достовірність роботи інформаційно-вимірювальної системи розпізнавання голосу можна віднести те, що немає потреби обчислювати інтеграл щільності розподілу ймовірностей інформативних ознак у n -вимірному просторі.

Використовуючи формули (18) та (23), проведено дослідження залежності помилки розпізнавання голосу від присутності в мовленнєвому сигналі шуму. Результати дослідження наведено на рис. 1.

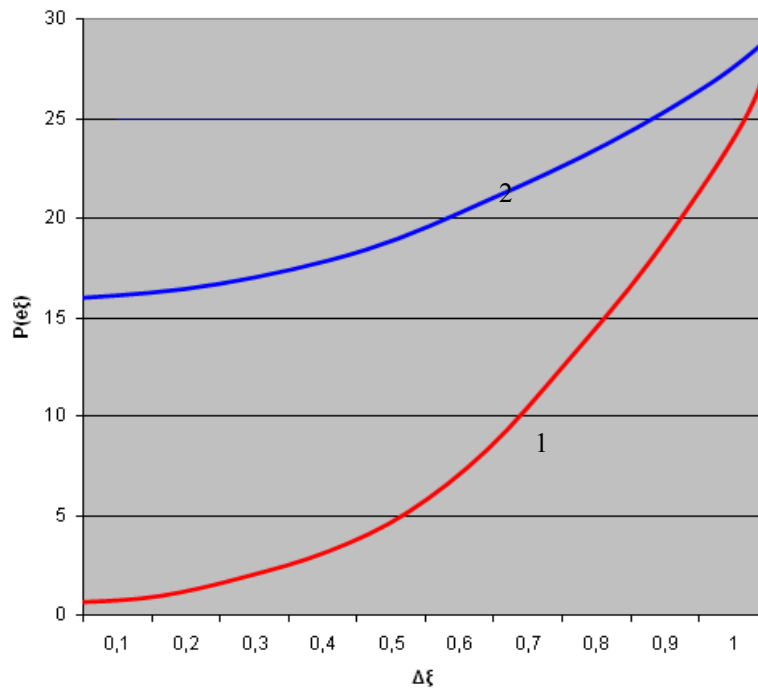


Рис. 1. Залежність помилки розпізнавання від величини шуму, присутнього в мовленнєвому сигналі: 1: $m_1^0 = 1$, $\sigma_n = 1$; 2: $m_1^0 = 1$, $\sigma_n = 0,5$

ВИСНОВКИ

У роботі отримано математичну модель впливу завад на достовірність розпізнавання голосу, яка усуває необхідність інтегрування сфері, помилкових рішень у багатовимірному ознаковому просторі. Проведені аналітичні викладки й експериментальні результати показали:

- зменшення впливу завад на достовірність роботи інформаційно-виміральної системи для розпізнавання голосу методом їхньої фільтрації є ефективним для інформативних ознак з високою роздільністю класів мовців, тобто за умови $\sigma_n \leq \frac{m_1^0}{2}$;
- фільтрація завад з великою амплітудою дає значний ефект під час розпізнавання голосів, ніж фільтрація завад з малою амплітудою.

СПИСОК ЛІТЕРАТУРИ

1. Давенпорт В. Б., Рут В. Л. Введение в теорию случайных сигналов и шумов. – М.: ИЛ, 1970. – 498 с.
2. Рамишвили Г.С. Автоматическое опознавание говорящего по голосу. – М.: Радио и связь, 1981. – 224 с.
3. Фант Г. Акустическая теория речеобразования. – М.: Наука, 1964. – 284 с.
4. Рабинер Л. Р. Шафер Р. В. Цифровая обработка речевых сигналов: Пер. с англ. – М.: Радио и связь, 1981. – 496 с.
5. Ротштейн А. П. Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети. – Вінниця: УНИВЕРСУМ-Вінниця, 1999. – 320 с.
6. Харкевич А. А. Спектры и анализ. – М.: Физматиз, 1962. – 320 с.

Биков Микола Максимович — к. т. н., проф., кафедри комп'ютерних систем управління.

Ковтун В'ячеслав Васильович — к. т. н., доц., кафедри комп'ютерних систем управління.

Савінова Наталія Геннадіївна — аспірант кафедри комп'ютерних систем управління, e-mail: savinova1987@mail.ru.

Вінницький національний технічний університет.