

УДК 681.327.12

М. М. Биков¹
В. В. Ковтун¹

ВИКОРИСТАННЯ МНОЖИНИ МІКРОФОНІВ У АВТОМАТИЗОВАНІЙ СИСТЕМІ РОЗПІЗНАВАННЯ МОВЦЯ КРИТИЧНОГО ЗАСТОСУВАННЯ

¹Вінницький національний технічний університет

Досліджено залежність якісних показників роботи автоматизованої системи розпізнавання мовців критичного застосування від використання множини мікрофонів для розпізнавання інформації. Для прийняття рішення щодо особи мовця використано суміші гаусових розподілів. Запропоновано правила формулювання остаточного рішення щодо особи мовця, які узагальнюють результати розпізнавання за інформацією від кожного мікрофону окремо.

Ключові слова: автоматизована система розпізнавання мовців критичного застосування, розпізнавання образів, кепстральний аналіз, суміш гаусових розподілів, мовний сигнал

Вступ та постановка задачі дослідження

Задача створення автоматизованої системи розпізнавання мовців критичного застосування (АСРМКЗ) вимагає надзвичайної уваги щодо якісних показників роботи системи, оскільки програмне забезпечення критичних систем має функціонувати з високою надійністю і зберігати прогнозований її рівень протягом всього життєвого циклу функціонування автоматизованої критичної системи незалежно від будь-яких зовнішніх обставин. Під час створення програмного забезпечення критичних систем віддають перевагу перевіреним та знаним методам і технологіям перед новітніми розробками, які не пройшли всебічної емпіричної перевірки. Ресурсовитратні технології, використання яких для розробки поточних систем є економічно не вигідним, допустимі для створення критичних систем, для яких головним є надійність функціонування. Такий підхід передбачає комплексність у створенні всіх структурних компонентів АСРМКЗ, таких як: блок попереднього оброблення мовних сигналів, блок виділення інформативних ознак, блок прийняття рішень. Окрім згаданих структурних елементів АСРМКЗ обов'язковим її компонентом є мікрофон. Враховуючи призначення створюваної системи, одним із можливих підходів до підвищення її надійності є використання не одного мікрофона для запису мовних сигналів, які використовуватимуться для розпізнавання, а множини. Авторами доведено ефективність такого рішення для задачі сегментації мовних сигналів за ознакою «мовний сигнал/пауза» у шумному середовищі [1], проте залежність між якісними показниками роботи АСРМКЗ, кількістю та розташуванням мікрофонів досі не досліджувалася. Також, враховуючи структуру створюваної системи, зростання кількості джерел надходження вхідної інформації відбивається на всіх інших її структурних елементах, що вимагає створення відповідного математичного апарату, зокрема для прийняття рішень.

Отже, в контексті критичного застосування автоматизованої системи розпізнавання мовця необхідним є дослідження доцільності використання множини мікрофонів в структурі таких систем та синтез математичного апарату, який описуватиме процес прийняття рішень щодо особи мовця з урахуванням таких структурних змін у системі.

Загальна інформація про суміші гаусових розподілів

Для опису мовного сигналу у частотній області серед найефективніших класичних систем розпізнавання мовців, що подаються на оцінювання, яке проводить Національний інститут стандартів і технологій США (National Institute of Standards and Technology, NIST), застосовується математичний апарат гаусових сумішей. Суміш гаусових розподілів (СГР) [2, 3] — це зважена сума M

функцій гаусових щільностей імовірності $b_i(x)$, що описується рівнянням

$$p(x|\lambda) = \sum_{i=1}^M w_i b_i(x), \quad (1)$$

де x — вектор інформативних ознак розмірністю D ; w_i — вага i -ї функції суміші.

Кожний компонент функції щільності обчислюється як

$$b_i(x) = \frac{1}{(2\pi)^{D/2} \sqrt{|\Sigma_i|}} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)}, \quad (2)$$

де μ_i — середній вектор розмірністю $D \times 1$; Σ_i — коваріаційна матриця розмірністю $D \times D$.

Для ваг сумішей w_i виконується співвідношення $\sum_{i=1}^M w_i = 1$. Отже, кожен мовець описується СГР

λ , яка включає у себе множини ваг сумішей w_i , середні вектори μ_i і коваріаційні матриці Σ_i

$$\lambda = \{w_i, \mu_i, \Sigma_i\}. \quad (3)$$

Формування правил прийняття рішень з використанням інформації від множини мікрофонів

Віднесення мовця, який виголосив мовний сигнал $X = \{x_1, x_2, \dots, x_T\}$, до одного з класів мовців $S = \{1, 2, \dots, N\}$ на основі моделей СГР пропонується здійснювати з використанням басового правила

$$\hat{S} = \arg \max_{1 \leq k \leq S} p(\lambda_k | X) = \arg \max_{1 \leq k \leq S} \frac{p(X|\lambda_k) p(\lambda_k)}{p(X)}, \quad (4)$$

де кожен мовець описується відповідною СГР λ_k .

Якщо припустити, що апріорна імовірність $p(\lambda_k)$ однакова для всіх k мовців, то рівняння (4) можна спростити до вигляду

$$\hat{S} = \arg \max_{1 \leq k \leq S} p(X|\lambda_k), \quad (5)$$

яке можна, застосувавши процедуру логарифмування та враховуючи незалежність реалізацій процедури розпізнавання окремих мовців, перетворити до вигляду

$$\hat{S} = \arg \max_{1 \leq k \leq S} \sum_{t=1}^T \log p(x_t | \lambda_k). \quad (6)$$

Нехай у автоматизованій системі розпізнавання мовця критичного застосування є можливість використати для запису паролльної фрази C мікрофонів, формуючи множину точок отримання інформації системою розпізнавання $X = \{X_1, X_2, \dots, X_C\}$. В такому випадку для кожного мовця будуватиметься власна СГР λ_k

$$\hat{S} = \arg \max_{1 \leq k \leq S} P(\lambda_k | X_1, X_2, \dots, X_C). \quad (7)$$

Згідно з теоремою Баеса, апостеріорну імовірність $P(\lambda_k | X_1, X_2, \dots, X_C)$ можна представити як

$$P(\lambda_k | X_1, X_2, \dots, X_C) = \frac{p(X_1, X_2, \dots, X_C | \lambda_k) P(\lambda_k)}{p(X_1, X_2, \dots, X_C)} \quad (8)$$

з подальшим представленням величини $P(X_1, X_2, \dots, X_C)$ у вигляді

$$p(X_1, X_2, \dots, X_C) = \sum_{k=1}^S p(X_1, X_2, \dots, X_C | \lambda_k) P(\lambda_k). \quad (9)$$

Якщо інформація, яка надходить у систему розпізнавання від входів $X = \{X_1, X_2, \dots, X_C\}$, є

незалежною, то результуюча ймовірність $P(X_1, X_2, \dots, X_C | \lambda_k)$ буде представлена добутком ймовірностей $p(X_c | \lambda_k)$

$$p(X_1, X_2, \dots, X_C | \lambda_k) = \prod_{c=1}^C p(X_c | \lambda_k). \quad (10)$$

Якщо підставити вираз (10) у (8) та вважати, що ймовірність $P(X_1, X_2, \dots, X_C)$ однакова для усіх мовців, то вираз (7) можна представити у вигляді

$$\hat{S} = \arg \max_{1 \leq k \leq S} \prod_{c=1}^C p(X_c | \lambda_k). \quad (11)$$

Правило (11) дозволяє приймати рішення щодо особи мовця з урахуванням інформації від множини незалежних вхідних джерел. Прологарифмувавши його, отримаємо вираз

$$\hat{S}_D = \arg \max_{1 \leq k \leq S} \prod_{c=1}^C \log p(X_c | \lambda_k), \quad (12)$$

який є першою із запропонованих авторами схемою прийняття рішень щодо особи мовця за значенням добутку ймовірностей розпізнавання мовця АСРМКЗ за записами мовних сигналів із множини незалежних джерел-мікрофонів.

З позиції апостеріорної ймовірності вираз (11) у випадку застосування множини мікрофонів представимо таким чином:

$$\hat{S} = \arg \max_{1 \leq k \leq S} P(\lambda_k)^{-C+1} \prod_{c=1}^C \log P(X_c | \lambda_k). \quad (13)$$

Вважаючи, що відхилення значень апостеріорної ймовірності від апріорної є незначним, припускаємо

$$P(X_c | \lambda_k) = P(\lambda_k)(1 + \delta_{kc}), \quad (14)$$

де $\delta_{kc} \ll 1$ [4].

Підставивши вираз (14) для апостеріорної ймовірності у (13) і не враховуючи члени рівняння, які піднесені до ступеня, відмінного від першого, вираз (12) набуде вигляду

$$P(\lambda_k)^{-C+1} \prod_{c=1}^C P(\lambda_k | X_c) = P(\lambda_k) \prod_{c=1}^C (1 + \delta_{kc}) \approx P(\lambda_k) + P(\lambda_k) \sum_{c=1}^C \delta_{kc}. \quad (15)$$

Враховуючи вирази (14) та (15), утворимо друге правило прийняття рішень:

$$\hat{S}_C = \arg \max_{1 \leq k \leq S} \sum_{c=1}^C P(\lambda_k | X_c). \quad (16)$$

Максимізуючи апостеріорну ймовірність в (16) отримаємо третє альтернативне правило прийняття рішень щодо особи мовця:

$$\hat{S}_{MC} = \arg \max_{1 \leq k \leq S} \max_{1 \leq c \leq C} P(\lambda_k | X_c). \quad (17)$$

І, нарешті, виконавши згортання функції апостеріорної ймовірності $P(\lambda_k | X_c)$ з використанням бінарної функції Δ_{kc} , вираз (16) подамо у вигляді четвертого правила прийняття рішень щодо особи мовця:

$$\hat{S}_\Delta = \arg \max_{1 \leq k \leq S} \Delta_{kc}, \quad (18)$$

де $\Delta_{kc} = \begin{cases} 1, & \text{якщо } P(\lambda_k | X_c) = \max_{1 \leq k \leq S} P(\lambda_k | X_c); \\ 0, & \text{в іншому випадку.} \end{cases}$

Експериментальні дослідження

Експериментальне дослідження ефективності запропонованих авторами правил узагальнення інформації від множини мікрофонів автоматизованої системи розпізнавання мовців критичного

застосування проводилося на основі бази еталонних записів безкоштовної бази даних NOIZEUS [5] — спеціалізованої бази даних Школи інжинірингу та комп'ютерних наук Еріка Джонсона при Університеті Техасу в Далласі (США). Вона використовується для дослідження алгоритмів покращення звуку і складається з 30 речень англійської розмовної мови, вимовлених трьома чоловіками та трьома жінками (по 5 на кожного диктора, частота дискретизації записів складає 25 кГц, але задля додавання шуму була зменшена до 8 кГц) та записів типових побутових та техногенних шумів. В ході експерименту автоматизовану систему розпізнавання мовців критичного застосування навчали як записами чистих парольних фраз, так і парольними фразами із додаванням шумів. В першому випадку навчальна вибірка містила 18 парольних фраз, у другому — 576, де до чистого сигналу додавався штучний шум з рівнями шум/сигнал 0 дБ, 5 дБ, 10 дБ, 15 дБ, відповідно. Для навчання використано 60 % обсягу бази аудіозаписів, у яку увійшли екземпляри записів без шумів та з різним рівнем шум/сигнал (5; 10; 15 дБ, відповідно). Тестувальна вибірка складала решту 40 % аудіозаписів. Розташування джерел звуку відносно множини мікрофонів системи розпізнавання показано на рис. 1. Джерела звуку розташовувалися на умовному радіусі кола з центром у множині мікрофонів у точках, віддалених від центру кола на 0,1, 0,5 і 1 м, та на радіусі, розташованому під кутом 45° відносно першого радіуса. Мовні сигнали виголошувалися однотипними джерелами звуку з гучністю 50 дБ.

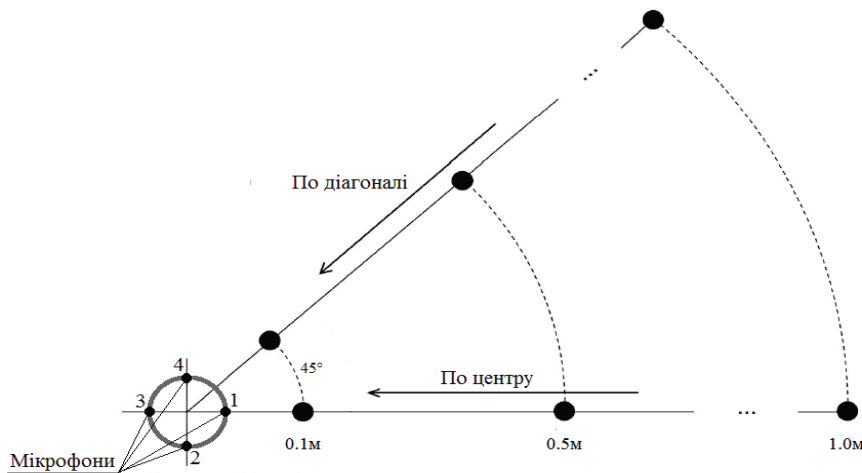


Рис. 1. Розташування мовця відносно 1...4 мікрофонів АСРМКЗ під час проведення дослідів

Створена авторами автоматизована система розпізнавання мовців критичного застосування складалася з блока попереднього опрацювання мовного сигналу, блока виділення інформативних для розпізнавання мовця ознак та блока прийняття рішень на їх основі. У блоці попереднього опрацювання мовного сигналу виконується попередня фільтрація мовного сигналу фільтром із характеристикою $H(z) = 1 - 0,97z^{-1}$, проводиться нормалізація тривалості звучання і сегментація на часові відрізки-фрейми тривалістю 20 мс із перекриттям 10 мс. У блоці виділення інформативних для розпізнавання мовця ознак для множини фреймів обчислюється енергетичний спектр швидкого перетворення Фур'є (ШПФ, FFT) з використанням вікна Хеммінга [6], після чого застосовується математичний апарат PLP-аналізу [7], а саме, частотний діапазон $[0, F_s]$, де F_s — частота дискретизації, розбивається на 10 частотних смуг, границі яких встановлено за Барк-шкалою

$$\left(z = 6 \log \left(\frac{f}{600} + \sqrt{\left(\frac{f}{600} \right)^2 + 1} \right) \right).$$

Для кожної з них будемо трапецеїдальний фільтр із амплітудно-частотною характеристикою

$$h(z) = \begin{cases} 10^{z-z_0+0,5}, & \text{якщо } z < z_0 - 0,5; \\ 1, & \text{якщо } z_0 - 0,5 \leq z \leq z_0 + 0,5; \\ 10^{-0,25(z-z_0+0,5)}, & \text{якщо } z > z_0 + 0,5, \end{cases}$$

де z_0 — центр частотної смуги.

Обчислимо суму енергій у всіх частотних смугах $\log E_i = \log \sum_{j=1}^{L_{FFT}/2+1} h_i^j X_j$,

де i — номер частотної смуги; j — номер фрейму; X_j — спектр потужності попереднього фрейму.

Далі пропускаємо всі отримані спектральні траєкторії $\log E_i$ крізь RASTA-фільтр [8] із передавальною функцією $R(z) = 0,1z^4 \frac{2+z^{-1}-z^{-3}-2z^{-4}}{1-0,94z^{-1}}$ і повертаємо їх у лінійний масштаб взяттям експоненти. Вже лінійний спектр множимо на криву рівної гучності [7]

$$H(f) = \frac{f^4}{(f^2 + 1,6 \cdot 10^5)^2} \cdot \frac{f^2 + 1,44 \cdot 10^6}{f^2 + 9,61 \cdot 10^6},$$

підносимо до степеня 0,33 (імітуючи закон слухового сприйняття), і виконуємо обернене перетворення Фур'є спектра, отримуючи автокореляційну функцію $R(k)$, $k = 0, \dots, L_{FFT} - 1$. Отримуємо рекурсію Левінсона-Дарбіна для розрахунку p коефіцієнтів лінійного прогнозу

$$E^{(0)} = R^{(0)}; \quad k_i = \frac{1}{E^{(i-1)}} \left(R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j) \right),$$

де $1 \leq i \leq p$; $a_i^{(i)} = k_i$; $a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)}$; $1 \leq j \leq i-1$; $E^{(i)} = (1 - k_i^2) E^{(i-1)}$.

В результаті застосування цих правил отримуємо коефіцієнти лінійного прогнозу $a_j = -a_j^{(p)}$, де $j = 1, \dots, p$, за якими обчислюємо кепстральні коефіцієнти на основі рекурентного співвідношення $c_n = -a_n - \sum_{k=1}^{n-1} \frac{k}{n} c_k a_{n-k}$, де $n = 1, \dots, p$, до яких застосовуємо процедуру ліфтингу $c'_n = n^{0,6} c_n$. Отже, блок виділення інформативних ознак представляє вхідний мовний сигнал множиною c'_n кепстральних коефіцієнтів. У проведених далі дослідах автори використали 15 кепстральних коефіцієнтів (c'_1, \dots, c'_{15}) .

Для підтвердження ефективності запропонованих авторами правил прийняття рішень щодо особи мовця у блоці прийняття рішень використовуються параметри моделей СГР $\{w_i, \mu_i, \Sigma_i\}$ з (3), побудованих для кожного з мовців, яких розпізнаватиме АСРМКЗ. Для навчання СГР використано ЕМ-алгоритм [9]. Емпірично виявлено, що для використаної бази мовців для навчання СГР достатньо не більше 18 ітерацій ЕМ-алгоритму. Подальше навчання не привело до покращення

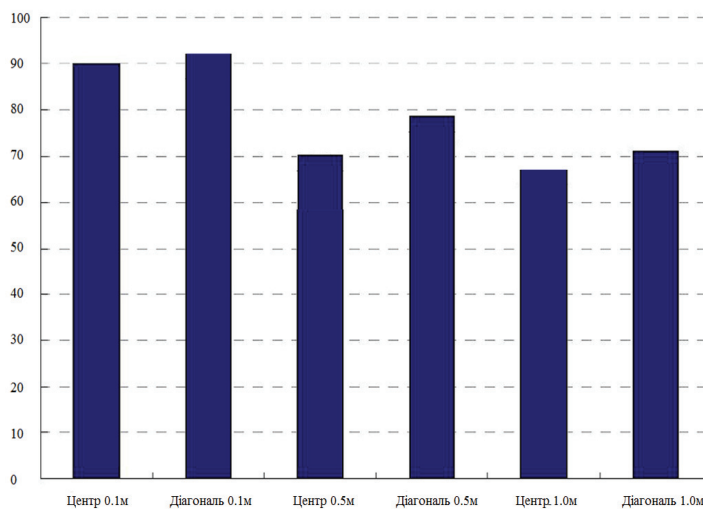


Рис. 2. Середня імовірність правильного розпізнавання для всіх чотирьох узагальнюючих правил в залежності від розташування мовця відносно системи розпізнавання

результатів розпізнавання. Для ініціації процесу навчання застосовано алгоритм К-середніх, який із застосуванням до множини векторів інформативних ознак $x = \{x_1, x_2, \dots, x_T\}$ дозволяє знайти M квантів, як основу для ініціації процесу пошуку оптимальних значень $\{w_i, \mu_i, \Sigma_i\}$.

Зазначимо, що створена авторами система на відміну від класичних, використовує інформацію від множини з чотирьох мікрофонів, що зумовило необхідність формування узагальнюючих правил (12), (16), (17) і (18) для блока прийняття рішень. На рис. 2 показано середню імовірність правильного розпізнавання для всіх чотирьох узагальнюювальних

правил в залежності від розташування мовців відносно системи розпізнавання.

Відмітимо, що діагональне розташування мовця відносно множини мікрофонів зі збільшенням відстані від нього до системи покращує результати розпізнавання.

Проведено дослідження залежності імовірності правильного розпізнавання від розташування мовця відносно множини мікрофонів, без узагальнення інформації, тобто використовуючи для прийняття рішень правила (6). Результати дослідження показані на рис. 3.

З рис. 3 випливає, що зі зростанням відстані від джерела мовного сигналу до множини мікрофонів інформативнішими виявляються мікрофони 1 і 3, які розташовані на різній відстані від джерела мовного сигналу. Це можна пояснити тим, що мовні хвилі згасають у середовищі їх поширення в міру віддалення від джерела сигналу. Враховуючи, що актуальна для розпізнавання мовця інформація міститься у відмінності аналізованих даних, таке розташування мікрофонів як у 1 і 3 відносно джерела сигналу виявляється ефективнішим, порівняно з точками розташування 2 і 4. Те, що ця ситуація проявилася, не зважаючи на малу відстань між мікрофонами, свідчить про високу чутливість обраних характеристик опису мовних сигналів.

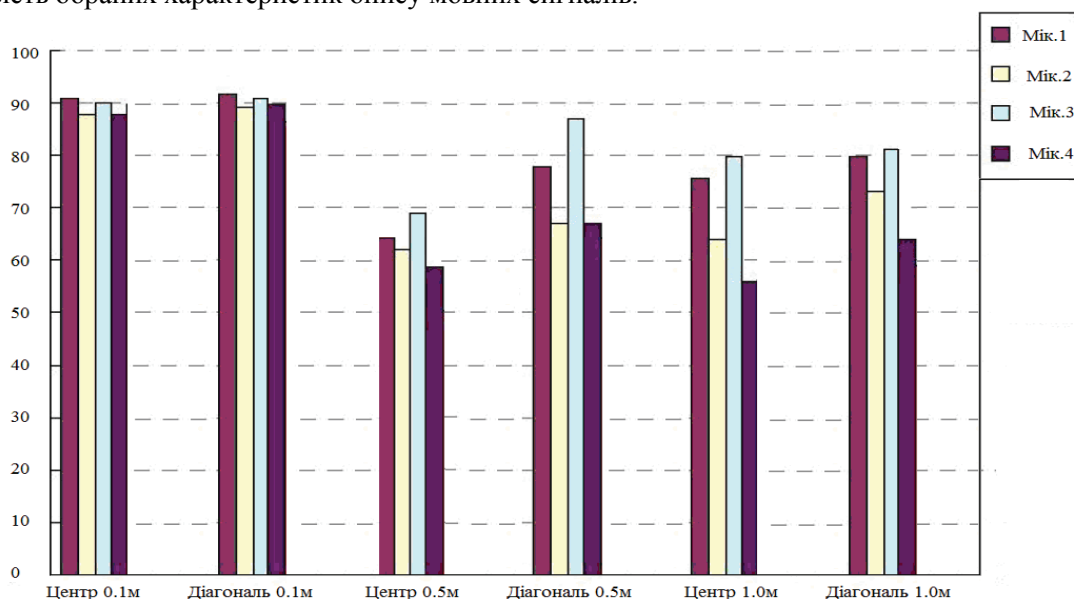


Рис. 3. Імовірність правильного розпізнавання в залежності від розташування мовця відносно системи розпізнавання для кожного з мікрофонів окремо

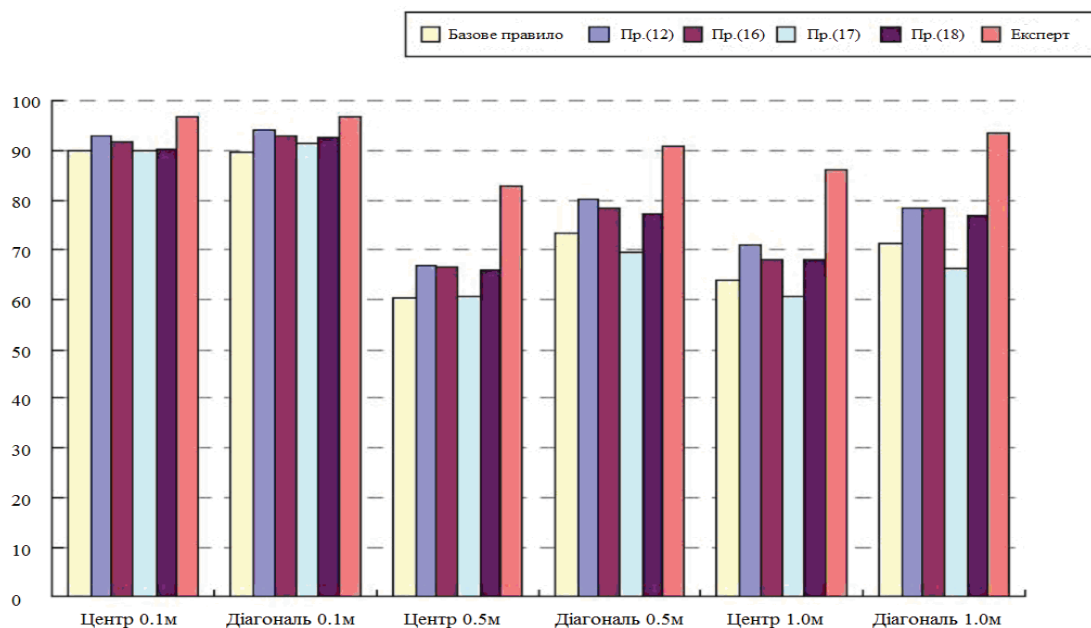


Рис. 4. Імовірність правильного розпізнавання в залежності від розташування мовця відносно системи розпізнавання для правил прийняття рішень (12), (16), (17) і (18)

Насамкінець, результати проведених дослідження впливу узагальнюючих правил на якість роботи автоматизованої системи розпізнавання мовців критичного застосування, показані на рис. 4, емпірично доводять доцільність узагальнення інформації у разі використання множини мікрофонів у АСРМКЗ і дозволяють стверджувати, що для запропонованої системи інформативних ознак найкращими виявилися правила узагальнення (12) та (16).

Висновки

Авторами досліджено питання доцільності застосування множини мікрофонів у структурі автоматизованої системи розпізнавання мовців критичного застосування. Автори запропонували програмну реалізацію такої системи з використанням значень RASTA-PLP кепстральних коефіцієнтів для опису мовних сигналів і математичного апарату гаусових сумішей для прийняття рішень. Враховуючи, що інформація на класифікатор надходить від множини джерел-мікрофонів, автори розробили правила узагальнення результатів розпізнавання, використання яких дозволило підняти відносну імовірність правильного розпізнавання на 10...20 %, яка зростає зі збільшенням відстані від джерела мовного сигналу до множини мікрофонів.

Результати досліджень показали, що зі зростанням відстані від джерела мовного сигналу до множини мікрофонів інформативнішими є ті два мікрофони 1 і 3, які розташовані на різній відстані від джерела звуку на умовній прямій, що проходить через них. Це можна пояснити тим, що мовні хвилі затухають у середовищі їх поширення в міру віддалення від джерела сигналу. Враховуючи, що актуальна для розпізнавання мовця інформація міститься у відмінності аналізованих даних, таке розташування мікрофонів як у 1 і 3 відносно джерела сигналу виявляється ефективнішим, порівняно з точками розташування 2 і 4. Те, що така ситуація проявилася, не зважаючи на малу відстань між мікрофонами, свідчить про високу чутливість обраних характеристик опису мовних сигналів.

Втім, використання апарату СГР виявилось досить ресурсовитратним, що зумовлює необхідність додаткових досліджень для оптимізації типу та порядку СРГ для прийняття рішень щодо особи мовця.

Також отримані авторами результати емпіричних досліджень доводять лише адекватність запропонованого математичного апарату. Втім питання оптимізації просторового розташування мікрофонів відносно мовця та визначення їх оптимальної кількості для застосування у конкретній критичній системі вимагає додаткового дослідження.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Биков М. М. Надійний метод виділення складових сегментів у мовному сигналі / М. М. Биков, В. В. Ковтун, Н. Г. Савінова [Електронний ресурс] // Наукові праці Вінницького національного технічного університету. — 2007. — № 1. — С. 1—6. — Режим доступу : <https://praci.vntu.edu.ua/index.php/praci>.
2. A short tutorial on Gaussian Mixture Models [Electronic resource] // 14th Conference on Computer and Robot Vision. May 17-19, 2017. — Edmonton, Alberta. — Access mode : http://www.computerrobotvision.org/2010/tutorial_day/GMM_said_crv10_tutorial.pdf.
3. Reynolds D. A. Speaker verification using adapted Gaussian mixture models / D. A. Reynolds, T. F. Quatieri, R. B. Dunn // Digital Signal Processing. — 2000. — Vol. 10. — P. 19—41.
4. Kittler J. Combining classifiers / J. Kittler, M. Hatef, P.W. Duijn // Proc. 7th German Workshop on Color Image Processing (ICPR'96), Erlangen, Germany, 1996. — Vol. 2. — P. 897—901.
5. NOIZEUS: Noisy speech corpus — Univ. Texas-Dallas [Electronic resource]. — Access mode : <http://ecs.utdallas.edu/loizou/speech/noizeus/>.
6. Рабинер Л. П. Цифровая обработка речевых сигналов / Л. П. Рабинер, Р. В. Шафер. — М. : Радио и связь, 1981. — 593 с.
7. Perceptual Linear Predictive (PLP) Analysis of Speech [Electronic resource]. — Access mode : <http://seed.ucsd.edu/mediawiki/images/5/5c/PLP.pdf>.
8. RASTA Processing of Speech [Electronic resource] // LABrosa. — Access mode : <https://labrosa.ee.columbia.edu/~dpwe/papers/HermM94-rasta.pdf>.
9. Reynolds D. A. Robust text-independent speaker identification using Gaussian mixture speaker models / D. A. Reynolds, R. C. Rose // IEEE Trans. Speech Audio Process. — 1995. — Vol. 3. — P. 72—83.

Рекомендована кафедрою комп'ютерних систем управління ВНТУ

Стаття надійшла до редакції 20.03.2017

Биков Микола Максимович — канд. техн. наук, доцент, професор кафедри комп'ютерних систем управління;

Ковтун В'ячеслав Васильович — канд. техн. наук, доцент, доцент кафедри комп'ютерних систем управління, e-mail: kovtun_v_v@vntu.edu.ua .

Вінницький національний технічний університет, Вінниця

M. M. Bykov¹
V. V. Kovtun¹

Using a Set of Microphones in the Automatic Speaker Recognition System of Critical Use

¹Vinnitsia National Technical University

In the article an investigation result of relationship between the quality indicators of the automated speaker recognition systems of critical use and speech material recorded on the set of microphones have been presented. The authors suggested complex decision rules on the classification of the speaker based on Gaussian mixture model parameters describing information from each microphone separately.

Keywords: automatic speaker recognition system of critical use, pattern recognition, cepstral analysis, Gaussian mixture models, speech signal.

Bykov Mykola M. — Cand. Sc. (Eng.), Assistant Professor, Professor of the Chair of Computer Control Systems;

Kovtun Viatcheslav V. — Cand. Sc. (Eng.), Assistant Professor, Assistant Professor of the Chair of Computer Control Systems, e-mail: kovtun_v_v@vntu.edu.ua

М. М. Быков¹
В. В. Ковтун¹

Использование множества микрофонов в автоматизированной системе распознавания диктора критического применения

¹Винницкий национальный технический университет

Исследована зависимость качественных показателей работы автоматизированной системы распознавания дикторов критического применения от использования множества микрофонов для распознавания информации. Для принятия решения относительно личности диктора использованы смеси гауссовых распределений. Авторы предложили правила формулирования окончательного решения относительно личности диктора, которые обобщают результаты распознавания по информации от каждого микрофона отдельно.

Ключевые слова: автоматизированная система распознавания диктора критического применения, распознавание образов, кепстральный анализ, смесь гауссовых распределений, речевой сигнал.

Быков Николай Максимович — канд. техн. наук, доцент, профессор кафедры компьютерных систем управления;

Ковтун Вячеслав Васильевич — канд. техн. наук, доцент, доцент кафедры компьютерных систем управления, e-mail: kovtun_v_v@vntu.edu.ua