РАДІОЕЛЕКТРОНІКА ТА РАДІОЕЛЕКТРОННЕ АПАРАТОБУДУВАННЯ

https://doi.org/10.31649/1997-9266-2023-169-4-68-77

УДК 621.391

О. В. Коваль[1]

ВИЯВЛЕННЯ ГОЛОСОВОЇ АКТИВНОСТІ НА ОСНОВІ КУТА НАХИЛУ АПРОКСИМУВАЛЬНОЇ ПРЯМОЇ ВЛАСНИХ ЗНАЧЕНЬ

[1]Харківський національний університет Повітряних Сил ім. Івана Кожедуба

Розглянуто метод виявлення голосової активності (VAD — Voice Activity Detection) з метою підвищення ефективності методів подавлення шуму в умовах низького співвідношення сигнал-шум. Наявність акустичних перешкод обмежує використання VAD та погіршує їхню продуктивність. Особливу увагу в роботі приділено методам VAD, що працюють в інтересах систем подавлення шуму, для оцінки шуму в зашумленому мовному повідомленні. Висока ефективність підпросторових методів подавлення шуму, основаних на перетворенні Корунена–Лоева, спонукала пошук простого та надійного VAD. Запропонований у статті метод виявлення голосової активності не вимагає додаткових перетворень та обчислень зашумленого мовлення та полегшує виявлення голосової активності в підпросторових методах подавлення шуму.

Як ознака класифікації мовних кадрів під час детектування голосової активності в запропонованому VAD використовується кут нахилу апроксимувальної прямої власних значень. Особливістю реалізації цього підходу є коригований спектр власних значень. За рахунок віднімання з власних значень коваріаційної матриці вхідних даних дисперсії шуму, досягається зменшення енергії шуму в спостереженні. Використання покращеної оцінки дисперсії шуму враховує наявність адитивних компонентів шуму в підпросторі сигналу. Як критерій прийняття рішення в роботі пропонується використання адаптивного порогу, на основі вхідного відношення сигнал-шум.

Проведений порівняльний аналіз роботи запропонованого VAD в умовах впливу кольорових шумів в порівнянні з VAD кодеку G.729. Реалізація моделей VAD проводилась в MATLAB та оцінено з використанням об'єктивних параметрів оцінки помилкових рішень в умовах впливу шуму. Подані результати моделювання, вказують на ефективність запропонованого методу за низьких значень відношення сигнал-шум (до 0 дБ). Запропонований метод VAD збільшує точність виявлення мовлення та зменшує кількість помилкових рішень. Проведене дослідження може бути використане для вдосконалення систем подавлення шуму.

Ключові слова: детектор голосової активності, мовний сигнал, власні значення, подавлення шуму.

Вступ

Впровадження нових систем та методів обробки сигналів є одним з головних напрямків розвитку систем зв'язку. Дослідження та розвиток методів подавлення шуму мови ведуться досить давно і є невід'ємною частиною в телекомукаційних системах. Мова, як природне джерело інформації, має надмірність [1, с.768]. В ній міститься велика кількість даних, що не несе смислового навантаження. Для більшості завдань, пов'язаних з обробкою мови, використовується VAD. Він дозволяє визначити моменти початку та закінчення мовних сегментів. Застосування модуля VAD можна зустріти під час розв'язання задач фільтрації (подавлення шуму) мови для оцінки статистики шуму, за використання методу спектрального віднімання. Важливу роль у виявленні «мовчання» модуль VAD відіграє у разі передачі голосу по каналах радіозв'язку або в пакетних мережах.

Існує величезна кількість різних підходів, які розробляються для конкретного виду розв'язуваних завдань. Це призводить до відсутності універсальних рішень та пошуку нових.

Так, наприклад, VAD може використовуватись сумісно з методом сингулярного спектрального аналізу (SSA — Singular Spectrum Analysis), що має місце у розв'язанні задач в системах зв'язку.

Увага до методу SSA зумовлена, високою ефективністю цих методів в задачах фільтрації мовних сигналів від шуму [2]—[4].

Переважна більшість сучасних методів спектрального аналізу основана на використанні методу головних компонент (PCA – Principal Component Analysis) [3], обчислення яких зведено до розрахунку сингулярного розкладання матриці даних (МД) (SVD — Singular Value Decomposition) чи спектральне розкладання кореляційної матриці (КМ) за власними значеннями (ВЗ) та власними векторами (ВВ) (EVD — Eigenvalue Decomposition) [4], [5]. В технічній літературі, методи пов'язані з розкладанням підпростору даних у вигляді суми взаємно ортогональних власних підпросторів (проєкцій) з відповідними коефіцієнтами у вигляді ВЗ отримали назву підпросторових [2]—[4]. В залежності від сфери застосування, метод головних компонент також має назву перетворення або розкладання Карунена–Лоєва [5]. Проте разом з перевагами підпросторових методів існують і недоліки, які полягають в їхній обчислюваній складності. Тому розробка простого та надійного VAD, який не вимагатиме додаткових складних обчислень і водночас оснований на розкладанні Карунена–Лоєва, є актуальною задачею.

Аналіз проблеми та постановка задачі

Серед наявних традиційних підходів з виділення голосової активності найбільшу популярність отримали методи [6]—[9], побудовані за схемою (рис. 1). В основі їхньої роботи лежить спектральне розкладання, короткочасна енергія, періодичність і спектральна щільність, виявлення енергетичного порогу, розрахунок частоти перетину «нуля» (ZCR — Zero Crossing Rate), лінійне передбачення (LPC — Linear Prediction Coefficients) та ін.

Слід відмітити достатню ефективність запропонованих рішень в умовах стаціонарних шумів з високим SNR (більше 10 дБ). Хоча на практиці такі умови обмежують сферу використання, особливо коли ця система повинна працювати в задачах фільтрації мовних сигналів від шуму. Для розв'язання цієї проблеми в [10] додатково використано складнішу систему ознак, таких як мел-частотні кепстральні коефіцієнти (MFCC — Mel-frequency Cepstral Coefficients), чим ускладнюють систему, в інтересах якої працює VAD, до рівня складності самої системи.

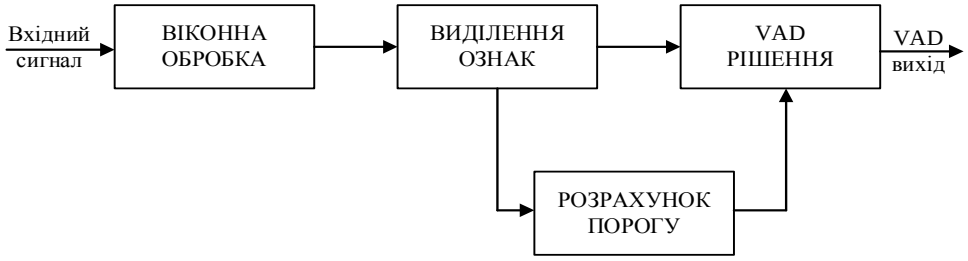


Рис. 1. Структурна схема VAD

Особливу увагу останнім часом приділяють дослідженням в сфері нейромережних технологій [11], [12], що працюють на основі отримання ознак вхідних мовних сигналів з використанням MFCC, отриманих за допомогою алгоритму швидкого перетворення Фур'є (FFT — Fast Fourier Transform) чи дискретного вейвлет-перетворення (DWT — Discrete Wavelet Transform). Однак ці реалізації також мають низку обмежень, пов'язаних з потребою великої кількості даних для навчання, наявністю високопродуктивних комп'ютерів та графічних процесорів.

Різновидом традиційних підходів з виявлення голосової активності стали моделі VAD на основі EVD [13], що використовують кадрову обробку кількох мікрофонів для обчислення енергії сигналу.

Відомі також поєднання SVD та FFT [14]. Проте, спектральне розкладання за сингулярними значеннями не дало змогу розділити мовні кадри від кадрів тиші, тому додатково проводилась апроксимація енергетичного спектра функцією Гауса для ідентифікації шумових кадрів. Узагальнена схема роботи варіантів реалізацій VAD показана на рис. 2.

Варіанти використання сингулярного спектра або спектра ВЗ як критерію прийняття рішення були запропоновані в роботах [15], [16] та зазнавали певної модифікації в [17], де розв'язується задача нестаціонарності шумових процесів. Головною проблемою запропонованих підходів є виділення мовної компоненти в умовах шуму. Це зумовлено впливом шумової складової на спектр мовних кадрів з низьким енергетичним рівнем. Приклад вибору (виділення) ознак з використанням підпросторових методів спектрального аналізу наведений в [18] для оцінювання кутових координат джерел.

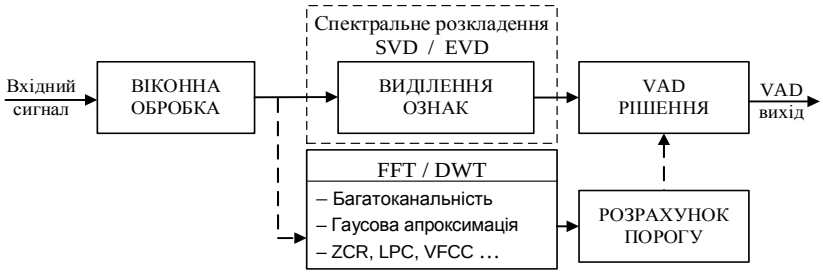


Рис. 2. Варіанти реалізацій VAD

Метою роботи є підвищення точності виявлення голосової активності на основі KLT та зменшення відсотка помилкових рішень VAD.

Мовний сигнал

Мовний сигнал є нестаціонарним, тому на практиці мовні сигнали піддають короткочасному спектральному аналізу. В межах короткочасного фрейму чи вікна тривалістю 10…35 мс, він може бути розглянутий як стаціонарний в широкому сенсі випадковий процес [1, с. 227]. В цій роботі взято за основу представлення мовного сигналу загальною лінійною моделлю [3], [4]

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta} = \sum_{i=1}^{p} h_i \theta_i, \tag{1}$$

де $\mathbf{s} = (s_1, s_2, ..., s_N)^T$ — послідовність відліків сигналу; $\mathbf{H} = \left[\mathbf{h}_1\ \mathbf{h}_2\ ...\ \mathbf{h}_p\right] \in \mathbb{R}^{m \times p}$ — m-мірні комплексні базисні вектори перетворення Карунена–Лоєва; $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_p)^T$ — вектор випадкового коефіцієнта з нульовим середнім, отриманий із багатовимірного розподілу.

В лінійній моделі стовбці $\mathbf{h}_i$ знаходяться в сигнальному підпросторі $\mathbf{H}$

$$\mathbf{H} = \left[\mathbf{h}_1 \quad \mathbf{h}_2 \quad ... \quad \mathbf{h}_p\right]. \tag{2}$$

У випадку коли стовпці $\mathbf{H}$ лінійно незалежні, тобто $\mathbf{H}$ має повний ранг, розмірність підпростору сигналу дорівнює $p < m$. Таким чином, коли $p < m$ множина всіх можливих векторів сигналу $\mathbf{s}$ обмежена знаходженням у підпросторі сигналу, що може бути використане для відновлення сигналу у разі шумових спотворень.

Шум спостереження для початку розглянемо, як адитивний стаціонарний випадковий процес з нульовим середнім, тоді зашумлений випадковий вектор $\mathbf{x} = (x_1, x_2, ..., x_N)^T$ розміром $N \times 1$, в такому разі може бути поданий як

$$\mathbf{x} = \mathbf{s} + \mathbf{n}, \tag{3}$$

де $\mathbf{s}$ — чистий мовний компонент, що лежить у підпросторі $\mathbf{H}$; $\mathbf{n}$ — адитивне спотворення білого шуму.

Таким чином, мовний сигнал, як відомо, лежить у сигнальному підпросторі рангу p, проте підпростір невідомий і точне його відновлення неможливе. Завдання методів подавлення шуму полягає в тому, щоб якомога точніше оцінити підпростір $\mathbf{H}$, тобто його розмір і відповідну основу та використовувати цю інформацію у процедурі фільтрації.

Коваріаційна матриця вектора $\mathbf{s}$ визначається як [4]

$$\mathbf{R}_s = E\left\{\mathbf{s}\mathbf{s}^T\right\} = \mathbf{H}\mathbf{R}_\theta\mathbf{H}^T, \tag{4}$$

де $\mathbf{R}_\theta$ — коваріаційна матриця вектора $\boldsymbol{\theta}$.

Відповідно $\mathbf{R}_s$ має ранг p і ця матриця має $m - p$ ненульових ВЗ. Подібним чином, нехай КМ шуму буде позначена $\mathbf{R}_n$, що дорівнює $\mathbf{R}_n = E\left\{\mathbf{n}\mathbf{n}^T\right\}$ та має повний ранг. Визначаючи статистичні характеристики, припускають, що елементи $\mathbf{s}$ та $\mathbf{n}$ не корельовані, а шум білий з дисперсією ν_{noise}^2, тому

$$\mathbf{R}_{sn} = \mathbf{R}_{ns} = 0, \tag{5}$$

$$\mathbf{R}_n = \nu_{noise}^2 \mathbf{I}_m . \quad (6)$$

Останнє припущення базується на тому факті, що КМ шуму вважається відомою, тоді, кольоровий шум завжди можна відбілити. Таким чином, КМ зашумленого вектора (3) визначається як

$$\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{R}_s + \mathbf{R}_n = \mathbf{H}\mathbf{R}_\theta \mathbf{H}^T + \nu_{noise}^2 \mathbf{I} . \quad (7)$$

Очевидно, що потужність шуму рівномірно розподілена у всьому просторі зашумленого сигналу, тоді як мовний сигнал обмежений розмірами p. На практиці точне знання статистики шумового вектора недоступне та оцінюється за сигналом зашумленої мови. Оцінку КМ $\mathbf{R}_x$ можна отримати з МД Ганкелевої структури $\mathbf{X}$ як

$$\hat{\mathbf{R}}_x = \frac{1}{K}\mathbf{X}\mathbf{X}^T \in \mathbb{R}^{m\times m} , \quad (8)$$

де $K = N - m + 1$.

Розкладення зашумленого простору на сигнальний та шумовий підпростір можна виконати шляхом застосування розкладання КМ [4] зашумленого сигналу (3), таким чином:

$$\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T = \sum_{i=1}^{m} \lambda_i \mathbf{q}_i \mathbf{q}_i^T , \quad (9)$$

де $\mathbf{Q} = \begin{pmatrix}\mathbf{q}_1 & ... & \mathbf{q}_m\end{pmatrix} \in \mathbb{R}^{m\times m}$ — дійсна ортогональна матриця власних векторів (ВВ);

$\mathbf{\Lambda} = diag(\lambda_1,...,\lambda_m) \in \mathbb{R}^{m\times m}$ — діагональна матриця власних значень (ВЗ), що розміщені в спадному порядку.

Рівняння (9) називається розкладанням КМ за ВЗ та ВВ. Сукупність ВЗ в такому разі називається спектром ВЗ [3], [4]. Для того, щоб виконати розділення матричного спектра, необхідно визначити порядок моделі p. Порядок моделі змінюється від кадру до кадру. Найвідоміші підходи, які розроблені та використані для оцінки порядку моделей — це мінімальна довжина опису (МДО) [4], [19] та інформаційний критерій Акаіке (ІКА) [4], [20]. Аналіз технічної літератури вказує на те, що для мовних сигналів порядок моделі p може бути встановлений на рівні 14 [2], [21], [22].

Для зменшення обмежень у побудові надійного VAD, що використовує EVD (SVD), запропоновано коригування сингулярних (власних) значень підпростору сигналу, як в [21], [22], а саме очищення від впливу шуму шляхом віднімання дисперсії шуму з ВЗ КМ матриці вхідних даних

$$\mathbf{R}_s = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T = \sum_{i=1}^{p} (\lambda_i - \hat{\sigma}_n)\mathbf{q}_i \mathbf{q}_i^T , \quad (10)$$

де $\hat{\sigma}_n$ — оцінка дисперсії шуму.

В цьому випадку спектр ВЗ з матриці $\mathbf{\Lambda}$ мовних ділянок буде позбавлений негативного впливу шумової компоненти, тоді як спектр ВЗ ділянок з кадрами, де мова відсутня, буде наближений до нульових значень. Головною умовою залишається оцінка дисперсії шуму, що буде мати мінімальну похибку в порівнянні з реальним значенням дисперсії [22].

Оцінка шуму

Оскільки доступ до чистого сигналу відсутній, оцінка SNR обчислюється на основі відношення потужності зашумленого мовлення до розрахункової потужності шуму. За таких обмежень в більшості методів VAD оцінка спектра потужності шуму відбувається у кадрах, де відсутня голосова активність, саме тому задається припущення про відсутність мови на перших 150…1200 відліках вхідного сигналу. При цьому оцінка шуму може виконуватись методом експоненціального усереднення [1, с. 25], що підвищує залежність від вхідного SNR та типу шуму. Іншим підходом, стійкішим до зміни рівня шуму, є метод мінімальної статистики [21], [23], який відслідковує рівень шуму в кожній спектральній компоненті, проте має певну затримку на період спостереження у разі збільшення рівня шуму. Однак такий підхід, щодо відсутності голосової активності протягом перших відліків вхідного сигналу може давати помилкову оцінку дисперсії шуму спостереження.

Маючи в своєму розпорядженні результати розкладання КМ за ВЗ згідно з виразом (9), оцінювання порядку моделі сигналу $\hat{p}$ виконано з використанням МДО, щоб не ускладнювати алгоритм VAD. Зі свого боку базова оцінка дисперсії шуму в спостереженні може бути розрахована як

$\hat{\sigma}^2_{base} = \left(1/(m-\hat{p})\right) trace\left(\hat{\Lambda}_n\right)$ [2]. Проте ця оцінка не враховує наявність адитивних компонентів шуму в підпросторі сигналу, що призводить до заниженого значення дисперсії шуму. Тому, запропоновано використання покращеної оцінки дисперсії шуму [22]

$$\hat{\sigma}^2_n = \hat{\sigma}^2_{base} + \frac{\frac{1}{K}\sum_{i=1}^{\hat{p}} \frac{\lambda_i \hat{\sigma}^2_{base}}{\lambda_i - \hat{\sigma}^2_{base}}}{\left(1-\hat{p}/K\right)}. \quad (11)$$

Оскільки сума ВЗ дорівнює енергії сигналу вздовж відповідного ВВ, розрахунок значення SNR проведено за використання рівняння [2]

$$SNR = 10\log_{10}\frac{trace\left(\mathbf{Q}^T\mathbf{R}_s\mathbf{Q}\right)}{trace\left(\mathbf{Q}^T\mathbf{R}_n\mathbf{Q}\right)} = \frac{\sum_{i=1}^{\hat{p}}\lambda_\Sigma(i)}{K}, \quad (12)$$

де $\lambda_\Sigma(i) = E\left\{\left|\mathbf{q}_i^T\mathbf{s}\right|^2\right\}$.

Використання кута нахилу апроксимувальних прямих ВЗ як ознаки прийняття рішень

Розглянемо можливість виявлення факту присутності мови за рахунок використання ВЗ матриці $\mathbf{\Lambda}$ мовних ділянок, отриманих за виразом (10). Для цього визначимо коефіцієнт нахилу апроксимувальних прямих (АП) ВЗ кадрів відрізку мовного сигналу (рис. 3).

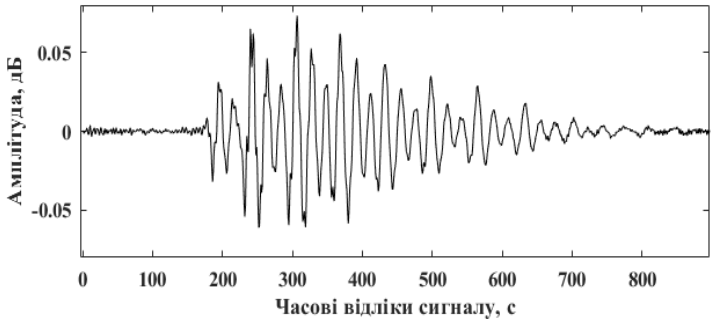


Рис. 3. Часова реалізація мовного сигналу

Розставимо ВЗ в порядку їхнього спадання і визначимо рівняння прямої $y(\lambda_i) = a + b(\lambda_i)$, що апроксимує сукупність точок λ_i, взятих із матриці $\mathbf{\Lambda}$. Далі проводиться розрахунок кута нахилу $\varphi = \text{arctg}(b)$.

Для визначення поведінки кута нахилу АП ВЗ проведено моделювання в умовах впливу білого шуму від 0 до 10 дБ (рис. 4).

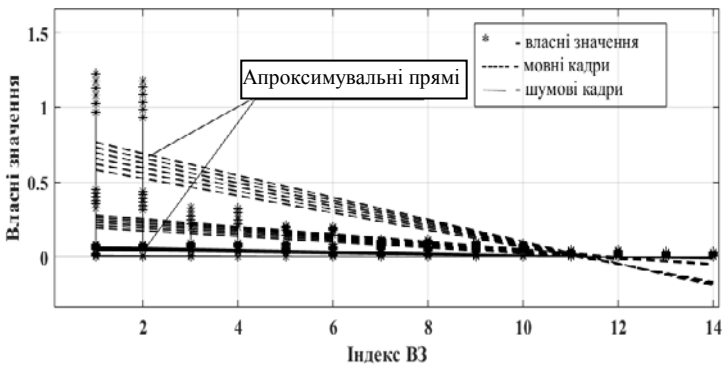


Рис. 4. Апроксимувальні прямі ВЗ

В ході проведених розрахунків встановлено, що кількість ВЗ, використаних для побудови АП, в мовних сигналах впливають на коефіцієнт нахилу прямої. Враховуючи порядок моделі мовного сигналу, кількість ВЗ в розрахунку кута φ зафіксовано на рівні, запропонованому в [21].

В табл. 1 подано значення модуля кута нахилу апроксимувальних прямих ВЗ $|\varphi|$ кадрів тиші (№ 16) та кадрів з мовою (№ 2–5).

Відповідні діаграми поведінки кутів нахилу АП ВЗ $|\varphi|$ в умовах впливу білого шуму від 0 до 10 дБ показано на рис. 5.

З візуального аналізу діаграм (рис. 5) видно, що параметр кута нахилу прямих, які апроксимують ВЗ матриць $\mathbf{\Lambda}$ зашумлених кадрів, залежать від загального рівня зашумленості. Також слід відмітити, що значення $|\varphi|$ можуть бути використані як ознаки класифікації мовних кадрів в VAD. Отже, такий VAD зможе працювати без використання додаткових операцій з виділення додаткових ознак для прийняття рішень у VAD.

Таблиця 1

Значення модуля кута нахилу $|\varphi|$ апроксимувальних прямих ВЗ

SNR, дБ / № кадру	0	1	2	3	4	5	6	7	8	9	10
	Кут нахилу ВЗ, $\|\varphi\|$ (град.)										
1	1,98	1,75	1,55	1,37	1,12	1,11	1,53	1,71	1,86	1,99	2,09
2	5,35	6,35	7,57	9,08	10,70	12,38	13,76	13,55	13,44	13,42	13,41
3	2,99	4,09	5,44	6,99	8,74	10,70	12,36	12,48	12,47	12,44	12,41
4	4,62	5,70	6,90	8,21	9,64	11,18	12,29	11,97	11,63	11,40	11,21
5	4,90	6,42	8,12	10,02	12,14	14,56	16,65	16,71	16,71	16,66	16,60
6	0,93	1,11	1,30	1,52	1,63	1,93	2,27	2,37	2,50	2,61	2,71

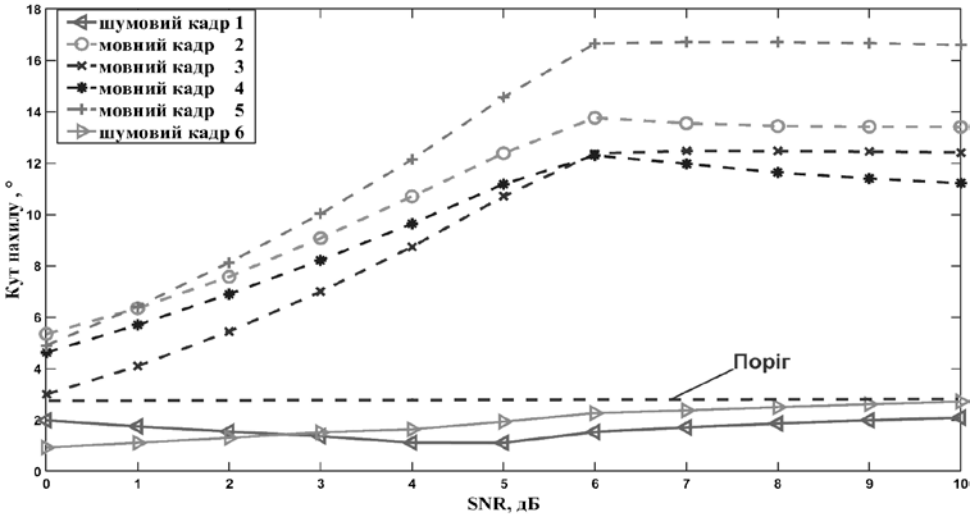


Рис. 5. Діаграми поведінки кутів нахилу АП ВЗ в умовах впливу білого шуму (0…10 дБ)

Для встановлення початкового порогу прийняття рішень припускається відсутність мови на перших трьох кадрах вхідного сигналу. Таким чином, для встановлення початкового порогу прийняття рішень VAD береться максимальне значення кута нахилу АП ВЗ, усереднених протягом трьох перших послідовних кадрів.

З метою визначення діапазону порогу прийняття рішень VAD в умовах впливу кольорових шумів проведено розрахунок кутів нахилу АП ВЗ ділянок сигналів, де мова відсутня. Для цього рівень SNR вхідного мовного сигналу встановлено в діапазоні від 20 до 5 дБ. Максимальні та мінімальні значення кутів в результаті проведених розрахунків становили 0,8° та 2,14° відповідно до рівня зашумленості. Для коригування встановленого порогу запропоновано рівномірний поділ в діапазоні розрахованих значень кута нахилу АП ВЗ згідно зі схемою, показаною на рис. 6.

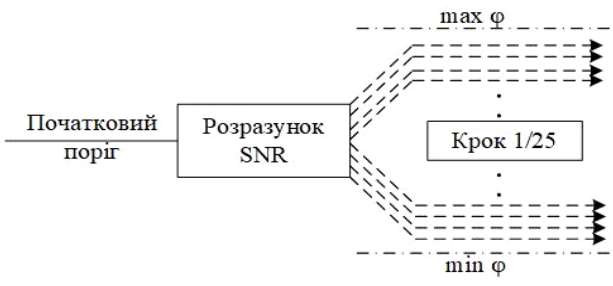


Рис. 6. Коригування встановленого порогу

Адаптація порогу VAD відбувається за рахунок оцінки SNR вхідного сигналу згідно з виразом (12). В основі моделі прийняття рішень, береться відхилення кута нахилу $|\varphi|$ від середнього максимального значення перших кадрів. Загальна схема роботи запропонованого VAD показана на рис. 7.

Для перевірки ефективності запропонованого підходу проведено імітаційне моделювання в програмному комплексі Matlab-2019b. Для проведення дослідження використовувались мовні сигнали з бази даних мовних сигналів NOIZEUS [24].

Для порівняльного аналізу роботи запропонованого методу вибрано VAD кодеку G.729 (стандарт ITU-T) з бази Mathworks [25]. Проведено моделювання в умовах впливу білого та рожевого шуму 0 та 5 дБ (рис. 8, 9).

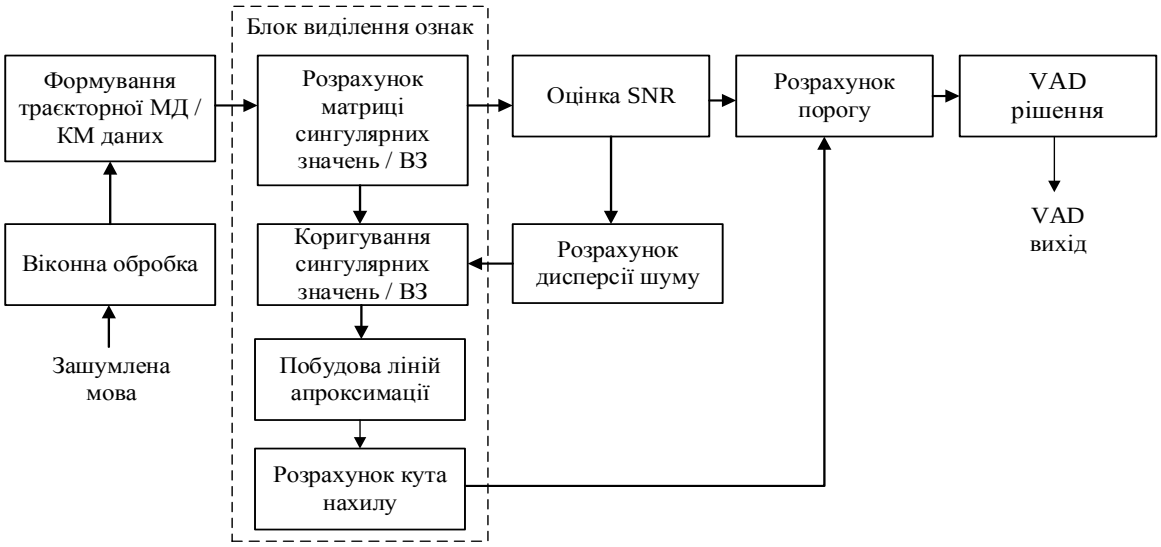
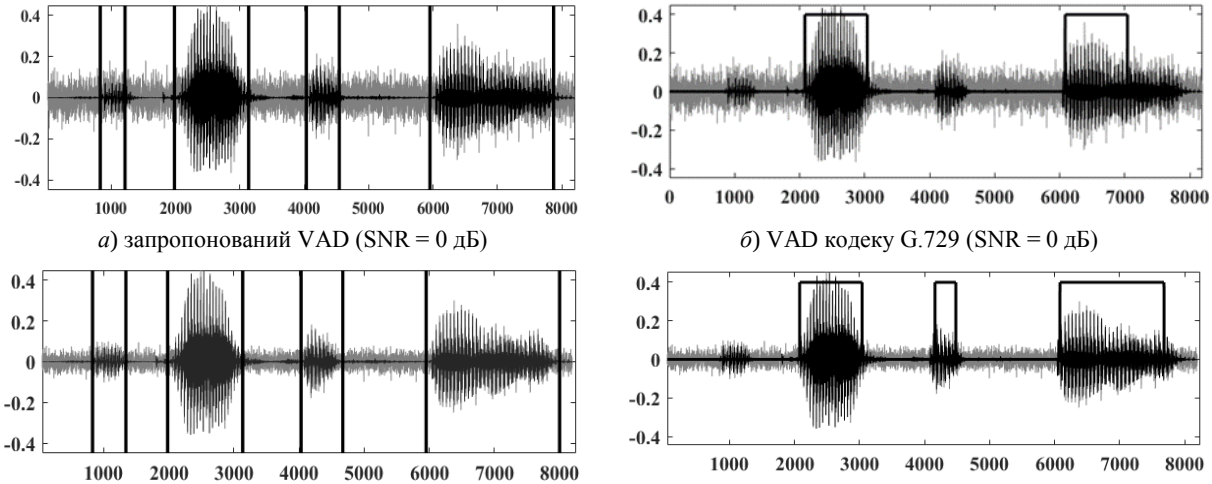


Рис. 7. Блок схема роботи запропонованого VAD



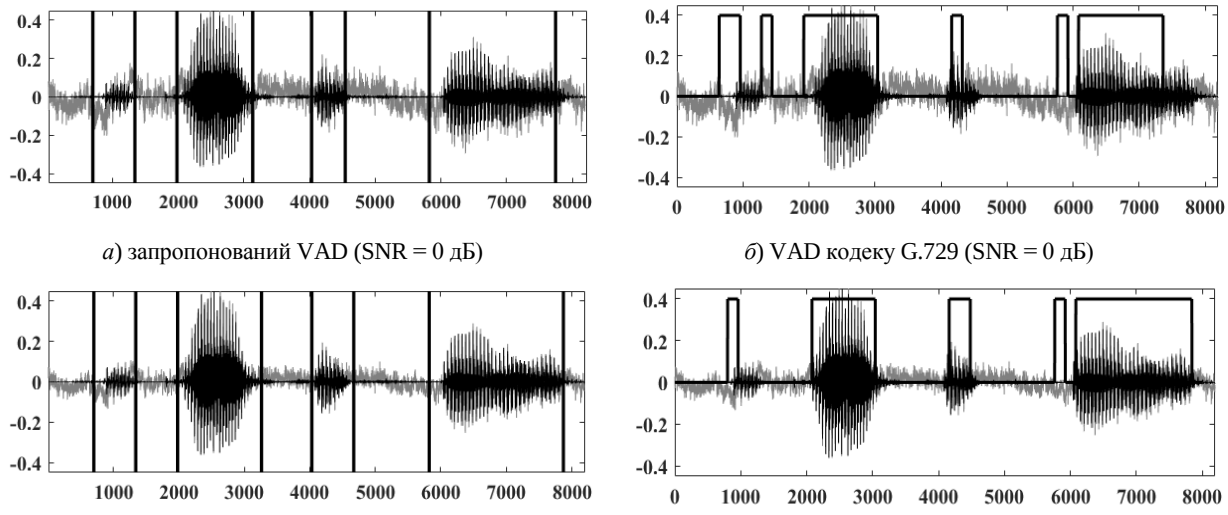
а) запропонований VAD (SNR = 0 дБ)

б) VAD кодеку G.729 (SNR = 0 дБ)

в) запропонований VAD (SNR = 5 дБ)

г) VAD кодеку G.729 (SNR = 5 дБ)

Рис. 8. Порівняння роботи VAD в умовах впливу білого шуму: *а*, *б* —SNR = 0 дБ; *в*, *г* — SNR = 5 дБ



а) запропонований VAD (SNR = 0 дБ)

б) VAD кодеку G.729 (SNR = 0 дБ)

в) запропонований VAD (SNR = 5 дБ)

г) VAD кодеку G.729 (SNR = 5 dB)

Рис. 9. Порівняння роботи VAD в умовах впливу рожевого шуму: *а*, *б* — SNR = 0 дБ; *в*, *г* — SNR = 5 дБ

В ході моделювання до мовного сигналу додавався шум, отриманий за допомогою функції Matlab — dsp.ColoredNoise, яка генерує шум із заданою спектральною щільністю потужності. Вхідний сигнал з частотою дискретизації 8000 Гц розбивався на кадри довжиною 32 мс (256 відліків), що перекривались на 50 %, кожен кадр помножувався на тимчасове вікно Хеммінга. На вхід VAD подавався мовний сигнал (фоновий сірий графік), спотворений білим та рожевим шумом. Часова реалізація «чистого» мовного сигналу представлені чорним кольором, для наочної верифікацій VAD. Вертикальними лініями на графіках нанесені ділянки мови, що були позначені як результат роботи VAD.

Аналіз оцінки ефективності алгоритмів VAD проводився за чотирма об'єктивними параметрами оцінки помилкових рішень [26]:

– FEC (Front End Clipping) — зміщення зони виявлення при переході від шуму до мовного активності;

– MSC (Mid Speech Clipping) — зміщення зони виявлення при помилковій класифікації мови як шум;

– OVER (Over Hang) — шум класифікується як мова, при переході від мови до шуму;

– NDS (Noise Detected as Speech) — шум класифікується як мова в період шуму.

Оцінки помилкових рішень виражені у відсотках в порівнянні з «еталонним» VAD шляхом ручного позначення чистого мовлення, записаного в спокійному середовищі. Параметри OVER і NDS вказують на помилкові рішення.

Порівняльні результати якості роботи детекторів в умовах шуму з SNR від 0 до 10 дБ подані в табл. 2 у вигляді об'єктивних оцінок. Проведений візуальний аналіз роботи моделей детекторів на рис. 8 та рис. 9, також враховуючи об'єктивні показники якості роботи VAD в умовах впливу шуму (табл. 2) вказують на те, що запропонований VAD є надійнішим у цих умовах та показує точніший результат, на відміну від VAD кодеку G.729.

Таблиця 2

Порівняльна характеристика роботи VAD в умовах шуму

Шум	SNR, дБ	FEC, %		MSC, %		OVER, %		NDS, %		Сума, %	
		G.729	новий VAD	G.729	новий VAD	G.729	новий VAD	G.729	новий VAD	G.729	новий VAD
Білий	0	5,37	3,34	7,92	0	7,86	5,91	0	0	21,15	9,25
	5	4,48	2,88	5,58	0	5,48	3,88	0	0	15,54	6,76
	10	3,82	2,15	0	0	3,67	2,61	0	0	7,49	4,76
Рожевий	0	5,29	3,66	1,22	0	4,88	3,56	1,83	0	13,22	7,22
	5	3,44	2,99	0	0	4,31	3,66	4,11	0	11,86	6,65
	10	3,44	3,17	0	0	6,75	1,44	0	1,22	10,19	5,83

Також слід зазначити, що в роботі VAD кодеку G.729 у разі впливу білого шуму зі встановленим SNR спостерігаються пропуски та зміщення в зони виявлення за помилкової класифікації мови (рис. 8*б*, *г*). У разі впливу рожевого шуму до зміщення зони виявлення додаються ділянки помилкового виявлення мовних сегментів (рис. 9*б*, *г*). Це пояснюється тим, що VAD кодеку G.729 використовує як критерій прийняття рішень ознаки спотворені адитивним шумом, що підвищує ймовірність помилкових рішень.

Висновки

Запропоновано новий VAD, оснований на аналізі поведінки кутів нахилу апроксимувальних прямих ВЗ. Низька обчислювальна складність запропонованого підходу та застосування поліпшеної оцінки дисперсії шуму дозволить підпросторовим методам ефективно ідентифікувати мовні сегменти в різних системах застосування без додаткових складних обчислень. Запропонований VAD працює з коригованим спектром ВЗ, така корекція дозволяє підвищити ефективність методу SSA [21], [22] та використана для реалізації запропонованого VAD. Більше того, як критерій прийняття рішень запропоновано використання адаптивного до SNR порогу. Використання запропонованого підходу збільшує надійність виявлення мовлення та зменшує відсоток помилкових рішень VAD за низького SNR (до 0 дБ).

Як напрямок подальших досліджень викликає зацікавленість дослідження впливу акустичних

шумів, шум літака F-16 на ефективність запропонованого підходу. До того ж, цікаво розглянути застосування запропонованого VAD розпізнавання мовних команд в системах управління військового призначення та для ідентифікації диктора.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] L. R. Rabiner, and R. W. Schafer, *Theory and Applications of Digital Speech Processing*, Pearson Education, 2011, 1060 p.

[2] Y. Hu, and P. Loizou, "Subjective Comparison of Speech Enhancement Algorithms," in *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, 2006, vol. 1, pp. I-I. https://doi.org/10.1109/ICASSP.2006.1659980 .

[3] N. Golyandina, and A. Zhigljavsky, *Singular spectrum analysis for time series*. London: Springer, 2013, 120 p.

[4] V. Vasylyshyn, "Adaptive Complex Singular Spectrum Analysis with Application to Modern Superresolution Methods," *Data-Centric Business and Applications*. Cham, 2020. pp. 35-54. https://doi.org/10.1007/978-3-030-43070-2_3.

[5] R. Wang, "Karhunen-Loève transform and principal component analysis," *In Introduction to Orthogonal Transforms: With Applications in Data Processing and Analysis.* Cambridge*:* Cambridge University Press, 2012, pp. 412-460. https://doi.org/10.1017/cbo9781139015158.011 .

[6] J. Ramírez, "Efficient voice activity detection algorithms using long-term speech information," *Speech Communication*, vol. 42, no. 3-4. pp. 271-287, April. 2004. https://doi.org/10.1016/j.specom.2003.10.002 .

[7] M. Sankar, and S. Arun, "Speech Sound Classification and Estimation of Optimal Order of LPC Using Neural Network," in *The 2nd International Conference on Vision, Image and Signal Processing*. ACM, 2018. https://doi.org/10.1145/3271553.3271611 .

[8] S. Ozaydin, "Design of a Voice Activity Detection Algorithm based on Logarithmic Signal Energy," in *International Conference on Electrical and Computing Technologies and Applications*. Ras Al Khaimah, United Arab Emirates, 2022, pp. 19-22. https://doi.org/10.1109/ICECTA57148.2022.9990492 .

[9] R. Çolak, and R. Akdeniz, "A Novel Voice Activity Detection for Multi-Channel Noise Reduction," *IEEE Access*, vol.9. pp. 91017-91026, June. 2021. URL: https://doi.org/10.1109/ACCESS.2021.3086364 .

[10] K. Yang, L. Zhu, and W. Shan, "Design of an ultra-low Power MFCC Feature Extraction Circuit with Embedded Speech Activity Detector," in *International Conference on Integrated Circuits, Technologies and Applications*. IEEE, 2021 pp. 82-83. URL: https://doi.org/10.1109/ICTA53157.2021.9661980 .

[11] A.Samanta, I.Hatai, and A. Mal, "A Reconfigurable Gaussian Base Normalization Deep Neural Network Design for an Energy-Efficient Voice Activity Detector," in *2nd International Conference on Communication, Computing and Industry 4.0*: conference paper. Bangalore, 2021, pp. 1-6. https://doi.org/10.1109/C2I454156.2021.9689307 .

[12] S. Abdullah, M. Zamani, and A. Demosthenous, "A Discrete wavelet transform-based voice activity detection and noise classification with sub-band selection," in *International Symposium on Circuits and Systems*: conference paper. IEEE, 2021, pp. 1-5. https://doi.org/10.1109/iscas51556.2021.9401647 .

[13] V. Neo, S. Weiss, S. McKnight, A. Hogg, and P. Naylor, "Polynomial Eigenvalue Decomposition-Based Target Speaker Voice Activity Detection in the Presence of Competing Talkers," in *International Workshop on Acoustic Signal Enhancement*: conference paper. IEEE, 2022, pp. 1-5. https://doi.org/10.1109/IWAENC53105.2022.9914796 .

[14] J. Ghasemi, A. Afzalian, and M.Mollaei, "A Combined Voice Activity Detector Based On Singular Value Decomposition and Fourier Transform," *Signal Processing*: An International Journal, vol. 4 (1). pp. 54-61, 2010.

[15] Y. Dongwen, "Robust Voice Activity Detection Based on Noise Eigenspace," *Acoustical Science and Technology*, vol. 28, no. 6. pp. 413-423, June. 2007. https://doi.org/10.1250/ast.28.413 .

[16] H. Song, S. Ban, and H. Kim, "Voice activity detection using singular value decomposition-based filter," in *Interspeech*: conference paper. ISCA, 2009, pp. 2223-2226. https://doi.org/10.21437/Interspeech.2009-632 .

[17] D. Kim, and J. Chang, "A subspace approach based on embedded prewhitening for voice activity detection," *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. EL304-EL310, Nov. 2011. https://doi.org/10.1121/1.3638927 .

[18] V. Vasylyshyn, "DOA estimation based on proximity of the roots of several polynomials of superresolution methods," *Advanced Information Systems*, vol. 4, no. 3, pp. 80-84, March. 2020. https://doi.org/10.20998/2522-9052.2020.3.10 .

[19] P. Stoica, and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36-47, July, 2004. https://doi.org/10.1109/MSP.2004.1311138 .

[20] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*,. vol. 19, no. 6, pp. 716-723, December. 1974. https://doi.org/10.1109/TAC.1974.1100705 .

[21] V. Vasylyshyn, O. Koval, and K. Vasylyshyn, "Speech Enhancement Using Modified SSA," in *IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics*: conference paper. IEEE, 2021, pp. 203-206. https://doi.org/10.1109/UkrMiCo52950.2021.9716635 .

[22] В. И. Василишин, «Предварительная обработка сигналов с использованием метода SSA в задачах спектрального анализа,» *Прикладная радиоэлектроника,* № 13(1), с. 43-50, 2014.

[23] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transaction on Speech and Audio Processing*, vol. 9, no. 5, pp. 504-512, July, 2001. https://doi.org/10.1109/89.928915 .

[24] *A noisy speech corpus for evaluation of speech enhancement algorithms NOIZEUS*. [Electronic resource]. Available: https://ecs.utdallas.edu/loizou/speech/noizeus. Accessed: 06.06.2023 .

[25] *G.729 Voice Activity Detection MATHWORKS*. [[Electronic resource]. Available: https://www.mathworks.com/help/dsp/ug/g-729-voice-activity-detection.html . Accessed: 06.06.2023.

[26] D. Freeman, G. Cosier, C. Southcott, and I. Boyd, "The voice activity detector for the Pan-European digital cellular mobile telephone service," *International Conference on Acoustics, Speech, and Signal Processing*, IEEE, 1989, vol. 1, pp. 369-372. https://doi.org/10.1109/ICASSP.1989.266442 .

Рекомендована кафедрою інформаційних радіоелектронних технологій і систем ВНТУ

Стаття надійшла до редакції 27.06.2023

Коваль Олексій Васильович — ад'юнкт кафедри радіоелектронних систем пунктів управління Повітряних Сил, e-mail: hnups@ukr.net .

Харківський національний університет Повітряних Сил ім. І. Кожедуба, Харків

O. V. Koval[1]

Detection of Voice Activity Based on the Angle of the Slope of the Approximating Line of the Eigenvalues

[1]Ivan Kozhedub National Air Force University, Kharkiv

The article discusses a method for detecting voice activity with the aim of improving the effectiveness of noise reduction methods in the conditions of low signal-to-noise ratio. The presence of acoustic disturbances limits the use of VAD (Voice Activity Detection) and degrades the performance. Special attention in the study is given to VAD methods that work in the interest of noise reduction systems, for estimating noise in noisy speech signals. The high efficiency of subspace-based noise reduction methods, based on the Karhunen–Loève transform, has prompted the search for a simple and reliable VAD for them. The method proposed in the article for voice activity detection does not require additional transformations of the noisy speech and facilitates the detection of voice activity in subspace-based noise reduction methods.

The proposed VAD utilizes the slope angle of the approximating line of the adjusted eigenvalues as the classification feature for speech frame classification during voice activity detection. The implementation of this approach involves an adjustable eigenvalue spectrum. By subtracting the noise variance from the eigenvalues of the input data covariance matrix, the reduction of noise energy in the observation is achieved. The use of the improved estimation of the noise variance takes into account the presence of additive noise components in the signal space. An adaptive threshold based on the input signal-to-noise ratio is proposed as the decision criterion in the study. A comparative analysis of the performance of the proposed VAD under the influence of color noise was conducted compared to the G.729 VAD codec. The implementation of the VAD models was done in MATLAB and evaluated using objective parameters for assessing erroneous decisions in noisy conditions. The presented simulation results indicate the effectiveness of the proposed method at low signal-to-noise ratios (down to 0 dB). The proposed method for voice activity detection increases speech detection accuracy and reduces the number of VAD erroneous decisions. The conducted research can be used to improve noise suppression systems.

Keywords: voice activity detector, speech signal, eigenvalues, noise reduction.

Koval Oleksii V. — Post-Graduate Student of the Chair of Radioelectronic Systems of Control Points of the Air Force, e-mail: hnups@ukr.net