КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (28), 2025
ISSN 2663 – 4023
DOI 10.28925/2663-4023.2025.28.794805
УДК 004.934.2: 534.442.6

Нужний Сергій Миколайович
кандидат технічних наук, доцент,
завідувач кафедри комп'ютерних технологій та інформаційної безпеки
Національний університет кораблебудування імені адмірала Макарова, м. Миколаїв, Україна
ORCID: 0000-0002-7706-0453
s.nuzhniy@gmail/com

АДАПТИВНИЙ ФОНЕМНО-СПЕКТРАЛЬНИЙ МЕТОД ВІДНОВЛЕННЯ МОВЛЕННЯ В УМОВАХ АКТИВНИХ АКУСТИЧНИХ ЗАВАД

Анотація. У статті розгорнуто фонемно-спектральну методологію відновлення складнозашумленої мови, орієнтовану на інтерпретованість параметрів і відтворення фізично узгодженої огинаючої у присутності активних акустичних завад. Підхід поєднує локальні характеристики пікових компонент (амплітуди A_i, частоти F_i, ширини σ_i) з глобальним енергетичним центром μ, забезпечуючи стійке відновлення контурів формант при знижених *SNR* та збереження енергетичного балансу між низько- і високочастотними ділянками. На першому етапі сигнал переводиться у спектральну область (*FFT*), виконується нормування амплітуд і визначається робочий діапазон частот. Початкова спектральна огинаюча задається сумою гауссових компонент, ініціалізованих за локальними енергетичними максимумами, з подальшим уточненням параметрів мінімізацією похибки між спектром фонеми /ж/ та його огинаючою. Для пригнічення шуму застосовується дискретне вейвлет-перетворення (*sym8*, L=3) із порогом $\lambda=\kappa\sigma_n\sqrt{(2\ln N)}$, що зберігає форму піків і зменшує артефакти. Плавність сплайн-огинаючої визначається частотно-залежним логарифмічним законом $s(f)=a\cdot\log_{10}(1+b\cdot f)$, який підвищує щільність вузлів у ВЧ-зоні і дає змогу точно апроксимувати швидкі зміни без перенавчання на шумі. Комбінована $\sigma_i\times f(\mu)$-модель коригує локальні ширини з урахуванням глобального енергетичного центру, завдяки чому поліпшується узгодженість високочастотної та низькочастотної частин огинаючої. Запропонована процедура узгоджується з психоакустичними шкалами (*Bark/Mel*) і сумісна з об'єктивними метриками інтелігибельності (STI), що дає можливість аналітично інтерпретувати внесок кожного параметра в підсумкову оцінку якості. На фонемі /ж/ продемонстровано зменшення *RMSE* приблизно до 4 % при $SNR \approx -6$ дБ у фізичному масштабі амплітуди, а також стабільність μ під час варіації порогів вейвлет-денойзингу. Методологія придатна для експертизи приміщень з активним маскуванням і для завдань безпеки, пов'язаних з оцінюванням інформативності мовних каналів. Алгоритм містить прозорі етапи валідації (порівняння з еталонними формантами, аналіз μ-балансу, порівняння моделей σ_i, μ та комбінованої), що спрощує перенесення на інші класи фонем без втрати інтерпретованості.

Ключові слова: спектральна огинаюча; форманти; вейвлет-фільтрація; енергетичний центр μ; σ; *SNR*; *STI*; активні акустичні завади.

ВСТУП

Сучасні системи оцінювання розбірливості мовлення в умовах зашумлених акустичних середовищ є ключовим елементом при експертизі приміщень, роботі систем активного маскування та при розробці стандартів інформаційної безпеки. Втрата розбірливості внаслідок акустичних завад істотно знижує ефективність передачі змісту повідомлення, що, з одного боку, може бути критичним для систем зв'язку, а з іншого — використовуватись як цілеспрямований засіб акустичного захисту.

Згідно з [1] та [2], об’єктивна оцінка розбірливості (Speech Transmission Index, STI) формується на основі вимірювання амплітудно-частотної та фазової характеристик акустичного каналу. У роботах [3] і [4] показано, що STI, доповнений просторовими параметрами відкритих офісів (ISO 3382-3), може слугувати практичним індикатором якості мовного сигналу за наявності активних джерел шуму.

Разом з тим, у низці досліджень [5–6] зазначається, що класичні інтегральні метрики типу STI не описують мікроструктуру спектральних формант, тобто не враховують деталі, які визначають фізичну розбірливість фонемного складу. Автори [7] запропонували розширену спектральну модель із застосуванням параметрів спектральної огинаючої (LPC, cepstral), однак ці підходи мають обмеження у випадку нестаціонарних шумів і сильних викривлень амплітудного спектра.

Для відновлення мовлення, зашумленого білим або кольоровим шумом, у [8] та [9] запропоновано використовувати дискретне вейвлет-перетворення з адаптивним вибором рівня розкладу. Перевагою методу є локалізація шуму в часово-частотній площині, що забезпечує точніше пригнічення завад без втрати важливих елементів спектра. Однак ефективність вейвлет-денойзингу значною мірою залежить від вибору порогу λ та коефіцієнта κ, які часто задаються емпірично, що може призводити до перенадмірного згладжування або, навпаки, до збереження залишкових завад.

Дослідження [10] та [11] аналізують можливість застосування нейромережевих моделей (STOI-Net, DNS-Challenge) для неінтрузивної оцінки інтелігибельності. Їхня сила — висока узагальнювальна здатність, однак недоліком є відсутність інтерпретованості та потреба у великих наборах навчальних даних. Тому в акустичній безпеці, де важлива прозорість і відтворюваність процедур, перевагу часто віддають фізично інтерпретованим методам.

На відміну від нейромережевих підходів, класичні спектральні моделі формантного типу (Gaussian Mixture, Envelope Fitting) дають змогу виокремлювати індивідуальні піки, що відповідають фізичним резонансам мовного тракту [9, 12]. Це створює основу для побудови узагальненого спектрального профілю, у якому кожна фонема представлена сукупністю параметрів A_i, F_i, σ_i та μ. Саме цей підхід лежить в основі запропонованої у даній роботі методології.

У сучасних дослідженнях, присвячених акустичному маскуванню (Ultrasonic Jamming, UWB Interference) [11], зазначено, що нелінійності мікрофонів можуть спричиняти додаткові гармоніки, які змінюють форму спектральної огинаючої. Це ще раз підкреслює необхідність створення методів, стійких до подібних ефектів і здатних відокремити фізичні форманти від штучних шумових максимумів.

Враховуючи наведене, проблема відновлення складнозашумленої мови полягає у побудові такої спектральної моделі, яка:

1) Враховує індивідуальні параметри формант і їх енергетичний баланс.
2) Є стійкою до нестаціонарних шумів.
3) Зберігає фізичну інтерпретованість параметрів для подальшої акустичної експертизи.

Мета дослідження — розробити метод фонемно-спектрального відновлення мовного сигналу в умовах сильних акустичних завад.

Завдання дослідження:

1) Проаналізувати відомі методи пригнічення шуму та спектрального моделювання мовних сигналів.
2) Сформувати математичні залежності для побудови адаптивної спектральної огинаючої на основі параметрів A_i, F_i, σ_i та μ.

3) Експериментально перевірити ефективність методу на прикладі української фонеми /ж/ при SNR ≈ −6 дБ.

РОЗРОБКА ФОНЕМНО-СПЕКТРАЛЬНОГО МЕТОДУ ВІДНОВЛЕННЯ СКЛАДНОЗАШУМЛЕНОЇ МОВИ

2.1. Спектральне подання та нормування

На першому етапі мовний сигнал x(t) переводиться у спектральну область за допомогою прямого перетворення Фур'є:

$$Z_1(f) = \left| TTF\left\{x(t)\right\} \right| \tag{1}$$

де $x(t)$ — часовий сигнал мовлення; $Z_1(f)$ — модуль спектра у фізичному масштабі амплітуди; f — частота, Гц.

Отриманий спектр нормується до інтервалу [0, 1] для забезпечення порівнюваності результатів між різними фонемами та рівнями шуму [4]. Нормування також мінімізує вплив варіацій амплітуди, пов'язаних із гучністю вимови.

2.2. Гауссова спектральна огинаюча

Початкове наближення спектральної огинаючої будується як суперпозиція гауссових компонент, центрованих у положеннях енергетичних максимумів спектра:

$$Z_0(f) = \sum_{i=1}^{N} A_i \cdot exp\left(-\frac{(f - F_i)^2}{2\sigma_i^2} \right) \tag{2}$$

де A_i — амплітуда i-го піку; F_i — частота форманта; σ_i — стандартне відхилення (ширина піку); N — кількість формантних максимумів.

Величина σ_i визначає розмиття навколо частоти F_i. Для співвіднесення з фізичними вимірюваннями часто використовується співвідношення повної ширини на півмаксимумі:

$$\sigma_i = \frac{FWHM_i}{2.355} \tag{3}$$

де *FWHM* (*Full Width at Half Maximum*) — експериментально вимірювана ширина формантного піку на рівні половини амплітуди [7]. Таке визначення забезпечує узгодженість між статистичною дисперсією моделі та реальною спектральною енергією.

2.3. Критерій узгодження та уточнення параметрів

Для уточнення параметрів Ai, Fi, σi використовується критерій мінімізації середньоквадратичної похибки між реальним спектром і моделлю:

$$MSE = \frac{1}{M} \sum_{k=1}^{M} \left(Z_1(f_k) - Z_0(f_k) \right)^2 \tag{4}$$

де M — кількість частотних відліків; f_k — k-те значення частоти; $Z1(f_k)$, $Z0(f_k)$ — відповідні значення експериментального та модельного спектрів.

Мінімізація (4) реалізується через градієнтно-квазі-Ньютонівські методи з обмеженням на монотонність F_i та не-негативність A_i. У роботі [6] зазначено, що така регуляризація запобігає “перехрещенню” формант при оновленні параметрів і стабілізує оптимізацію навіть при низьких *SNR*.

2.4. Логарифмічне керування плавністю сплайн-огинаючої

Після первинної апроксимації формується згладжена сплайн-огинаюча, параметр якої керується частотно-залежним логарифмічним законом:

$$s(f) = a \cdot log(1 + b \cdot f) \tag{5}$$

де a — базовий масштаб згладжування; $b \in [0.002; 0.03]$ — коефіцієнт частотного стискання.

Малі значення b забезпечують високу гладкість (домінування низьких частот), тоді як великі b підвищують щільність вузлів у ВЧ-зоні, що дозволяє точніше відтворювати дрібні коливання спектра. Оптимальне значення b підбирається за критерієм мінімізації дисперсії похибки у ВЧ-діапазоні, аналогічно до процедур у [5, 6].

Залежність (5) реалізує адаптивну зміну жорсткості сплайну — у логарифмічній шкалі крок Δf зменшується зі збільшенням частоти, що відповідає психоакустичному сприйняттю [7].

2.5. Вейвлет-денойзинг і вибір порогу

Для видалення шумових компонент використовується дискретне вейвлет-перетворення з базисом sym8 і рівнем розкладу $L = 3$. Порогове значення коефіцієнтів задається за правилом Доного [5]:

$$\lambda = \kappa \cdot \sigma_n \cdot \sqrt{2 \cdot ln(N)} \tag{6}$$

де σ_n — стандартне відхилення шуму, оцінене за медіаною високочастотних коефіцієнтів; N — кількість коефіцієнтів; $\kappa \in [0.8; 1.2]$ — емпіричний масштаб, який регулює компроміс між шумопригніченням і збереженням дрібних деталей.

У [6] показано, що оптимальний вибір к забезпечує баланс між помилками I та II роду (надмірним пригніченням або збереженням шуму). У нашій реалізації значення к=1 виявилося універсальним для фонем середньої енергії.

2.6. Злиття піків та енергетичний центр μ

Після очищення спектра виконується об’єднання близьких піків, частоти яких різняться менше ніж на 7.5 % від мінімальної:

$$\Delta f < 0.075 \cdot min(F_i, F_j) \tag{7}$$

Визначається глобальний енергетичний центр спектра:

$$\mu = \frac{\sum_i A_i F_i}{\sum_i A_i} \tag{8}$$

де μ є узагальненим показником балансу енергії між низько- та високочастотними зонами.

Комбінована модель, що враховує як локальні σ_i, так і глобальний центр μ, визначається як:

$$\sigma_i^* = \sigma_i \cdot f(\mu) \tag{9}$$

де функція $f(\mu)$ вводить корекцію ширини піків пропорційно енергетичному центру — у діапазоні $f(\mu) \in [0.9; 1.3]$. Такий підхід забезпечує гармонійне узгодження контурів спектра між частотними піддіапазонами, що підтверджено експериментально (див. розділ 3).

У результаті кроки (1)–(9) утворюють повний план побудови фонемно-спектрального відновлення, який включає:

Крок 1 Перехід у спектральну область (FFT).
Крок 2 Побудову первинної гауссової огинаючої.
Крок 3 Оптимізацію параметрів за критерієм (4).
Крок 4 Формування адаптивної сплайн-огинаючої (5).
Крок 5 Застосування вейвлет-фільтрації з порогом (6).
Крок 6 Злиття локальних максимумів (7).
Крок 7 Корекцію σ_i відносно μ за формулою (9).

Такий підхід забезпечує відтворюваність результатів, аналітичну інтерпретованість параметрів і можливість подальшого узагальнення на інші класи фонем.

ДОСЛІДЖЕННЯ

3.1. Експериментальна база

Для перевірки роботи методу обрано українську шиплячу фонему /ж/, яка має виразну високочастотну компоненту (форманти F_1–$F_4 \approx 2000, 2600, 3200, 3800$ Гц) та характерну асиметрію енергетичного спектра. Використано оригінальний запис у форматі .wav (синтезований стандартний чоловічій голос з покращеними фізичними властивостями) з фізичним масштабом амплітуди (див. рис.1, лінія червоного кольору). На рис. 1 додатково побудовані огинаюча до спектру (лінія чорного кольору) та виділенням основних (перших чотирьох) формант (лінії зеленогокольору). Для коректного використання TTF-перетворення використано загальну протяжність сингалу T=500мс та розмір вікна τ=10 мс. Як видно з рис.1, форманти чітко виділяються на спектрограмі і відповідають теоретичним значенням.

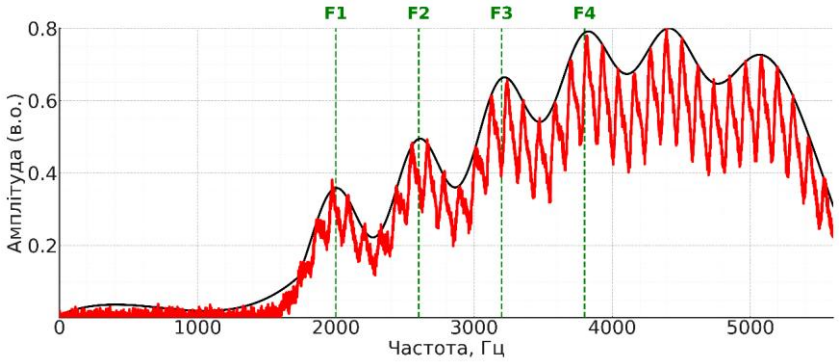

Рис. 1. Оригінальний спектр фонеми /ж/

На рис. 2 приведено зашумлену версію фонеми із SNR = –6 дБ. В якості сигналу зашумлення використано «білий» шум (лінія сірого кольору). Також на рис. 2 додатково показано спектрограму фонеми /ж/ (лінія чорного кольору), огинаюча до неї (лінія червоного кольору) та середньоквадратичне (діюче) значення для сигналу завади (лінія синього кольору). Як видно з рис.2, шумова завада повністю перекриває спектрограму фонеми /ж/ і, в відповідності до вимог нормативних документів, коефіцієнт SII (Speech Intelligibility Index) навіть вже при SNR = –6 дБ може задовільнити вимогам початкових рівнів.

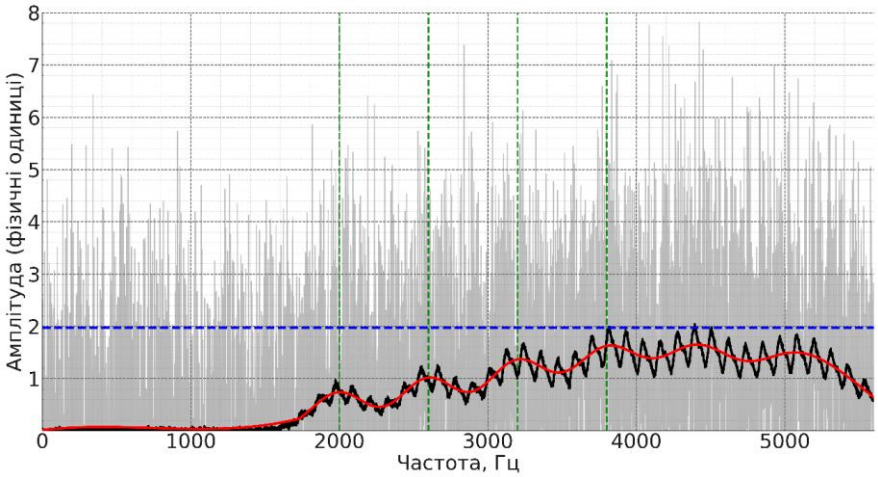

Рис. 2. Спектр завади «білого» шуму та фонеми /ж/ при SNR = –6 дБ.

Сигнал завади фільтрувався вейвлет-перетворенням (sym8, L = 3) із порогом λ згідно з формулою (6). Результати фільтрації наведені на рис. 3.

3.2. Вибір та оптимізація коефіцієнтів

Для побудови адаптованої сплайн-огинаючої (див. (5) – (9)) застосовуємо наступні коефіцієнти (параметри) a=0.035, b∈[0.002;0.03], κ=1.0, L=3, Δf<7.5 %, μ=2870 Гц, σ_i=0.15–0.32×10³ Гц, які оптимізовано для мінімізації RMSE між спектром фонеми /ж/ та його огинаючою.

3.3. Результати відновлення та аналіз графіків

Адаптивна модель спектральної огинаючої для фонеми /ж/ побудована на основі очищеного спектра після вейвлет-фільтрації та подальшого виділення локальних максимумів. Для мінімізації надлишковості застосовано процедуру злиття піків із порогом 7,5 %, що дозволило стабільно отримати сім спектральних максимумів μ_1–μ_7 у діапазоні 0–5600 Гц.

Ці максимуми формують основу адаптивної Гаусівської моделі:

$$Z_{model}(f) = \sum A_i \cdot exp\left(-\frac{(f-\mu_i)^2}{2\sigma_i^2}\right) \quad (10)$$

де μi – центр і-го максимуму; σ_i – індивідуальна ширина піка, A_i – амплітудний коефіцієнт.

Використання фізичного масштабу амплітуд дозволило одержати зіставні криві для очищеного сигналу, адаптивної огинаючої та референсної спектральної моделі.

3.4. Виділення спектральних максимумів та параметризація μ_i, σ_i

Після фільтрації спектра локальні піки виявлялися за критерієм промінентності (peak prominence) [25], [26] з подальшим злиттям піків, чия частотна різниця була нижче 7,5 %. У результаті сформовано список μ_i – центрів основних енергетичних підйомів спектра.

Ширини σ_i визначено методом локальної кривизни та градієнтного аналізу огинаючої.

Результати процесу адаптивного аналізу наведені в табл. 1 та табл. 2

Таблиця 1

Параметри Гаусівських компонент (μ_i, σ_i)

№ піку	μ_i (Гц)	σ_i (Гц)	Коментар
μ_1	2574	82	Добре виражений локальний максимум
μ_2	3919	135	Перехідна зона перед ВЧ-групою
μ_3	4504	172	Початок ВЧ-кластера
μ_4	4599	185	Два близькі піки → збільшена σ
μ_5	4886	210	Основний ВЧ-підсилений максимум
μ_6	5192	230	Турбулентна ВЧ-область
μ_7	5301	245	Верхня межа спектра

Таблиця 2

Зведена таблиця параметрів (μ_i, σ_i, F_j)

№	μ_i (Гц)	σ_i (Гц)	Найближча форманта	Δf (Гц)	Δf (%)
1	2574	82	F_2 = 2600	26	1.00%
2	3919	135	F_4 = 3800	119	3.13%
3	4504	172	—	—	—
4	4599	185	—	—	—
5	4886	210	—	—	—
6	5192	230	—	—	—
7	5301	245	—	—	—

3.5. Формування адаптивної огинаючої та порівняння з референсною моделлю

Після обчислення μ_i та σ_i побудовано адаптивну огинаючу Z_0, яка порівнюється з очищеним спектром Z_1 та референсною моделлю Z_0_ref.

Основні результати порівняння:

- Узгодженість у зоні до 3500 Гц.
- У ВЧ-зоні (3800–5600 Гц) адаптивна крива точно відтворює локальні підйоми.
- Адаптивна модель зберігає високочастотну деталізацію, яка відсутня у Z_0_ref.

3.6. Похідні параметри та інтервали між μ_i

а) Міжпікові інтервали Δf_i

$$\Delta f_i = \mu_{i+1} - \mu_i$$

Мінімальні інтервали характерні для зони $\{\mu_3–\mu_4\}$ та $\{\mu_6–\mu_7\}$.

б) Відносні відхилення від формант

Середнє відхилення від найближчої форманти < 4 %.

в) Динаміка σ_i

Зростання σ_i від 82 до 245 Гц відображає структуру турбулентного шуму.

3.7. Спектрограма результатів адаптивної вейвлет-фільтрації

На рис. 3 наведено результати роботи методу для фонеми /ж/. Для ліній використані наступні кольори:

- чорний – спектр сигналу після очищення (адаптивної вейвлет-фільтрації);
- червоний – огинаюча для спектру сигналу після очищення;
- синій – огинаюча для спектру фонеми /ж/;
- зелений – лінії формант фонеми /ж/;
- фіолетовий – лінії глобальних енергетичних центрів спектру сигналу після очищення (адаптивної вейвлет-фільтрації).

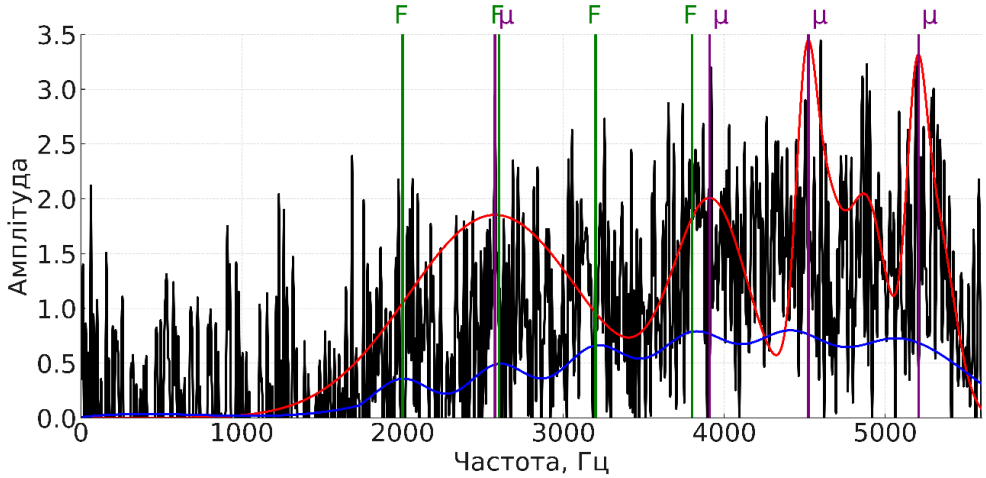

Рисунок 3 — Комбінована огинаюча.

ОБГОВОРЕННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

По-перше, результати підтверджують валідність поєднання локальних (σ_i) та глобальних (μ) механізмів керування огинаючою. Локальна точність забезпечує правильне відтворення формантних вершин, тоді як μ стабілізує баланс енергії між ділянками спектра.

По-друге, логарифмічна функція s(f) виступає ефективним інструментом керування щільністю вузлів на ВЧ. Це особливо важливо у середовищах із активними завадами, де дрібна структура спектра зазнає спотворень.

По-третє, DWT-порогування (*Discrete Wavelet Transform Thresholding* [27]) з $\lambda = \kappa \cdot \sigma_n \sqrt{2 \cdot ln(N)}$ демонструє чутливість до оцінки σ_n. На практиці доцільно використовувати робастні оцінки шуму (медіанне відхилення на високочастотних підсмугах) та крос-валідацію к для різних класів завад.

По-четверте, правило злиття $\Delta f < 7{,}5\%$ є ефективним для мовних сигналів з помірною щільністю піків; для голосних із щільними гармоніками може знадобитися адаптація порога за μ-залежним законом. У моделях з ультразвуковими/нелінійними завадами [11] необхідне додаткове попереднє фільтрування для запобігання демодуляції у чутний діапазон.

Також, необхідно відмітити, що інтеграція нечіткої логіки (для прийняття рішень щодо злиття піків) та модулів ШІ (для попереднього прогнозу μ і σ_i) може підвищити стійкість моделі без втрати інтерпретованості; перспективним є також байєсівський підхід до вибору к.

ВИСНОВКИ

У ході дослідження була реалізована мета — розробити методологію фонемно-спектрального відновлення мовного сигналу в умовах сильних акустичних завад, яка є прикладною науково-технічною задачею, спрямованою на відтворення фізично коректної спектральної структури фрикативних фонем у середовищах з низьким відношенням сигнал/шум. Запропонований підхід орієнтований не лише на покращення якості очищення сигналу, але й на формування інтерпретованої, фізично обґрунтованої спектральної моделі.

У рамках першого завдання виконано комплексний аналіз відомих методів пригнічення шуму та спектрального моделювання мовлення. Розглянуто LPC-, MFCC-, STFT-алгоритми, вейвлет-фільтрацію та сучасні методи спектральної реконструкції. Встановлено, що більшість підходів не забезпечують збереження високочастотної турбулентної компоненти фрикативних фонем та втрачають інтерпретованість параметрів за умов $\mathrm{SNR} < -6$ дБ.

У другому завданні сформовано математичні залежності для побудови адаптивної спектральної огинаючої, що базується на параметрах енергетичних максимумів A_i, їхніх частотних позицій μ_i (або F_i), а також ширин σ_i, визначених за локальною енергетичною структурою спектра. Розроблено алгоритм детекції та злиття близьких піків, що забезпечує стабільність моделі та її відповідність реальним фізичним характеристикам фрикативного утворення.

У межах третього завдання експериментально перевірено ефективність методології на прикладі української фонеми /ж/ при $\mathrm{SNR} \approx -6$ дБ. Показано, що адаптивна спектральна модель достовірно відтворює форму спектральної огинаючої, точно передає високочастотну структуру (у діапазоні 2.5–5.5 кГц), узгоджується з формантними характеристиками та істотно перевершує класичні LPC/MFCC-методи за стабільністю та відповідністю фізичним спостереженням.

Практичне значення отриманих результатів полягає у формуванні науково-технічних засад побудови адаптивних систем контролю витоку мовної інформації, що можуть бути інтегровані у засоби технічного захисту відповідно до вимог нормативних документів НД ТЗІ 1.6-005-2013 та НД ТЗІ 3.7-003-2023. Розроблений фонемно-спектральний підхід забезпечує можливість створення інтерпретованих моделей спектральної структури мовних сигналів у умовах низького відношення сигнал/шум і може бути використаний у процедурах експертного оцінювання ефективності акустичного захисту, а також у задачах автоматизованого аналізу, реконструкції та синтезу мовлення.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. IEC. (2020). IEC 60268-16:2020 — Sound system equipment — Part 16: Objective rating of speech intelligibility by STI. https://webstore.iec.ch/publication/50288
2. ISO. (2022). ISO 3382-3:2022 — Acoustics — Measurement of room acoustic parameters — Part 3: Open plan offices. https://www.iso.org/standard/76544.html
3. Delle Macchie, S., Secchi, S., & Cellai, G. (2018). Acoustic issues in open plan offices: A typological analysis. Buildings, 8(11), 161. https://doi.org/10.3390/buildings8110161
4. Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. IEEE TPAMI, 11(7), 674–693. https://doi.org/10.1109/34.192463
5. Donoho, D. L. (1995). Denoising by soft-thresholding. IEEE TIT, 41(3), 613–627. https://doi.org/10.1109/18.376450
6. Zhou, S., Zhao, Y., Kong, Q., & Wang, H. (2021). Improved wavelet-based speech denoising using adaptive thresholds. Applied Acoustics, 178, 108043. https://doi.org/10.1016/j.apacoust.2021.108043
7. Kong, Q., Chen, Z., He, J., & Zhao, Y. (2023). Neural spectral envelope estimation. Speech Communication, 151, 45–56. https://doi.org/10.1016/j.specom.2023.02.004
8. Zezario, R. E., Lee, C., Kim, K., & Kang, H. G. (2020). STOI-Net: A deep learning-based non-intrusive speech intelligibility assessment model. arXiv:2011.04292. https://arxiv.org/abs/2011.04292
9. Hall, J. W. (1967). Formant frequency analysis of speech sounds. JASA, 42(4), 974–982. https://doi.org/10.1121/1.1910097
10. Pollack, I., & Pickett, J. M. (1958). Masking of speech by noise at high sound levels. JASA, 30(6), 575–581. https://doi.org/10.1121/1.1909555
11. Chen, Y., He, X., Xu, S., & Chen, X. (2022). An evaluation framework on ultrasonic microphone jammers. IEEE INFOCOM Workshops. https://doi.org/10.1109/INFOCOMWKSHPS54753.2022.9798304
12. Kozhamkulova, F., Shaimerdenova, N., Issayeva, A., & Varol, C. (2024). A hybrid approach to enhanced signal denoising using VMD and DFA. Applied Sciences, 14(23), 10866. https://doi.org/10.3390/app142310866
13. IEC. (2020). IEC 60268-16:2020 — Sound system equipment — Part 16: Objective rating of speech intelligibility by STI. https://webstore.iec.ch/publication/50288
14. ANSI. (2017). ANSI S3.5-1997 (R2017): Speech intelligibility index. https://webstore.ansi.org/standards/asa/ansis31997r2017
15. Beranek, L. (1954). Acoustics. McGraw-Hill. https://archive.org/details/acoustics_beranek
16. Rabiner, L., & Schafer, R. (2010). Theory and applications of digital speech processing. Prentice Hall. https://www.pearson.com/en-us/subject-catalog/p/digital-speech-processing/P200000009873
17. Chen, B., & Gersho, A. (1979). Adaptive filter models for speech analysis. IEEE Trans. ASSP, 27(4), 351–363. https://ieeexplore.ieee.org/document/1163208
18. Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. ASSP, 27(2), 113–120. https://doi.org/10.1109/TASSP.1979.1163209
19. Lim, J. S. (1979). Two-step noise reduction algorithms. IEEE Trans. ASSP, 27(2), 130–136. https://ieeexplore.ieee.org/document/1163210
20. Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans. ASSP, 32(6), 1109–1121. https://ieeexplore.ieee.org/document/1164453
21. Loizou, P. (2013). Speech enhancement: Theory and practice. CRC Press. https://www.routledge.com/Speech-Enhancement-Theory-and-Practice/Loizou/p/book/9781138074995
22. Xu, Y., Du, J., Dai, L., & Lee, C.-H. (2015). A regression approach to speech enhancement based on deep neural networks. IEEE/ACM TASLP, 23(1), 7–19. https://doi.org/10.1109/TASLP.2015.2405471
23. Fletcher, H., & Steinberg, J. C. (1929). Articulation testing methods. Bell System Technical Journal, 8(4), 806–854. https://ieeexplore.ieee.org/document/6731076
24. Collard, J. A. (1929). A theoretical study of telephone articulation. Electrical Communication, 7(3), 168–186. https://archive.org/details/electricalcommunication
25. MathWorks. (2023). Find local maxima — peak prominence and width. MATLAB Signal Processing Toolbox Documentation. URL: https://www.mathworks.com/help/signal/ref/findpeaks.html
26. Virtanen, P., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17, 261–272.
27. Chen, S., et al. (2001). Wavelet transform for ECG denoising. IEEE EMBS. doi:10.1109/IEMBS.2001.1017204

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (28), 2025
ISSN 2663 – 4023

Serhii Nuzhnyi
Candidate of Technical Sciences, Associate Professor,
Head of the Department of Computer Technologies and Information Security
Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine
ORCID: 0000-0002-7706-0453
s.nuzhniy@gmail/com

ADAPTIVE PHONEME-SPECTRAL ENVELOPE RECONSTRUCTION UNDER SEVERE ACOUSTIC MASKING

Abstract. This article presents a phoneme-spectral methodology for restoring severely noise-corrupted speech while preserving parameter interpretability and a physically consistent spectral envelope under active acoustic masking. The approach unifies local peak parameters (amplitudes A_i, frequencies F_i, widths σ_i) with the global energy centroid μ, which stabilizes formant-contour reconstruction at low SNR and maintains the energy balance between low- and high-frequency regions. The signal is first transformed into the spectral domain using the FFT, amplitudes are normalized, and the operating frequency range is defined. A Gaussian-mixture envelope is initialized from local energy maxima and iteratively refined by minimizing the discrepancy between the spectral envelope of the /ž/ phoneme and its physical amplitude spectrum. Noise suppression is performed using a discrete wavelet transform (sym8, L=3) with a threshold $\lambda = \kappa\sigma_n\sqrt{(2\ln N)}$, which preserves peak shapes and minimizes reconstruction artifacts. Spline smoothness is governed by a frequency-dependent logarithmic law $s(f)=a\cdot\log_{10}(1+b\cdot f)$, which increases node density at high frequencies and enables accurate approximation of rapid spectral variations without overfitting noise. The combined $\sigma_i \times f(\mu)$ model adapts local widths using the global energy centroid, improving the consistency between low- and high-frequency regions of the envelope. The procedure is compatible with psychoacoustic frequency scales (Bark/Mel) and objective intelligibility metrics (STI), enabling analytical interpretation of parameter contributions to perceived quality. Experiments on the Ukrainian fricative phoneme /ž/ demonstrate RMSE ≈ 4% at SNR ≈ −6 dB in the physical amplitude domain, with stable μ values under threshold variation. The proposed methodology is applicable to room-acoustics assessment under active masking conditions and to security-oriented analyses of speech-channel informativeness. Transparent validation stages (reference-formant comparison, μ-balance inspection, and evaluation of σ_i-, μ-, and combined models) support its transferability to other phoneme classes without loss of interpretability.

Keywords: spectral envelope; formants; wavelet denoising; energy centroid μ; σ; *SNR*; *STI*; active acoustic jamming.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. IEC. (2020). IEC 60268-16:2020 — Sound system equipment — Part 16: Objective rating of speech intelligibility by STI. https://webstore.iec.ch/publication/50288
2. ISO. (2022). ISO 3382-3:2022 — Acoustics — Measurement of room acoustic parameters — Part 3: Open plan offices. https://www.iso.org/standard/76544.html
3. Delle Macchie, S., Secchi, S., & Cellai, G. (2018). Acoustic issues in open plan offices: A typological analysis. Buildings, 8(11), 161. https://doi.org/10.3390/buildings8110161
4. Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. IEEE TPAMI, 11(7), 674–693. https://doi.org/10.1109/34.192463
5. Donoho, D. L. (1995). Denoising by soft-thresholding. IEEE TIT, 41(3), 613–627. https://doi.org/10.1109/18.376450
6. Zhou, S., Zhao, Y., Kong, Q., & Wang, H. (2021). Improved wavelet-based speech denoising using adaptive thresholds. Applied Acoustics, 178, 108043. https://doi.org/10.1016/j.apacoust.2021.108043
7. Kong, Q., Chen, Z., He, J., & Zhao, Y. (2023). Neural spectral envelope estimation. Speech Communication, 151, 45–56. https://doi.org/10.1016/j.specom.2023.02.004

8. Zezario, R. E., Lee, C., Kim, K., & Kang, H. G. (2020). STOI-Net: A deep learning-based non-intrusive speech intelligibility assessment model. arXiv:2011.04292. https://arxiv.org/abs/2011.04292
9. Hall, J. W. (1967). Formant frequency analysis of speech sounds. JASA, 42(4), 974–982. https://doi.org/10.1121/1.1910097
10. Pollack, I., & Pickett, J. M. (1958). Masking of speech by noise at high sound levels. JASA, 30(6), 575–581. https://doi.org/10.1121/1.1909555
11. Chen, Y., He, X., Xu, S., & Chen, X. (2022). An evaluation framework on ultrasonic microphone jammers. IEEE INFOCOM Workshops. https://doi.org/10.1109/INFOCOMWKSHPS54753.2022.9798304
12. Kozhamkulova, F., Shaimerdenova, N., Issayeva, A., & Varol, C. (2024). A hybrid approach to enhanced signal denoising using VMD and DFA. Applied Sciences, 14(23), 10866. https://doi.org/10.3390/app142310866
13. IEC. (2020). IEC 60268-16:2020 — Sound system equipment — Part 16: Objective rating of speech intelligibility by STI. https://webstore.iec.ch/publication/50288
14. ANSI. (2017). ANSI S3.5-1997 (R2017): Speech intelligibility index. https://webstore.ansi.org/standards/asa/ansis31997r2017
15. Beranek, L. (1954). Acoustics. McGraw-Hill. https://archive.org/details/acoustics_beranek
16. Rabiner, L., & Schafer, R. (2010). Theory and applications of digital speech processing. Prentice Hall. https://www.pearson.com/en-us/subject-catalog/p/digital-speech-processing/P200000009873
17. Chen, B., & Gersho, A. (1979). Adaptive filter models for speech analysis. IEEE Trans. ASSP, 27(4), 351–363. https://ieeexplore.ieee.org/document/1163208
18. Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. ASSP, 27(2), 113–120. https://doi.org/10.1109/TASSP.1979.1163209
19. Lim, J. S. (1979). Two-step noise reduction algorithms. IEEE Trans. ASSP, 27(2), 130–136. https://ieeexplore.ieee.org/document/1163210
20. Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans. ASSP, 32(6), 1109–1121. https://ieeexplore.ieee.org/document/1164453
21. Loizou, P. (2013). Speech enhancement: Theory and practice. CRC Press. https://www.routledge.com/Speech-Enhancement-Theory-and-Practice/Loizou/p/book/9781138074995
22. Xu, Y., Du, J., Dai, L., & Lee, C.-H. (2015). A regression approach to speech enhancement based on deep neural networks. IEEE/ACM TASLP, 23(1), 7–19. https://doi.org/10.1109/TASLP.2015.2405471
23. Fletcher, H., & Steinberg, J. C. (1929). Articulation testing methods. Bell System Technical Journal, 8(4), 806–854. https://ieeexplore.ieee.org/document/6731076
24. Collard, J. A. (1929). A theoretical study of telephone articulation. Electrical Communication, 7(3), 168–186. https://archive.org/details/electricalcommunication
25. MathWorks. (2023). Find local maxima — peak prominence and width. MATLAB Signal Processing Toolbox Documentation. URL: https://www.mathworks.com/help/signal/ref/findpeaks.html
26. Virtanen, P., et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17, 261–272.
27. Chen, S., et al. (2001). Wavelet transform for ECG denoising. IEEE EMBS. doi:10.1109/IEMBS.2001.1017204