Одеський національний університет імені І. І. Мечникова
Факультет математики, фізики та інформаційних технологій
Кафедра оптимального керування і економічної кібернетики

Кваліфікаційна робота

на здобуття ступеня вищої освіти «магістр»

«Методи та алгоритми розпізнавання звукових образів»

«Methods and algorithms of sound image recognition»

Виконав(ла): здобувач денної форми навчання
спеціальності 113 Прикладна математика
Освітня програма «Прикладна математика»
Кулик Даніїл Вячеславович

Керівник: канд. техн. наук, доц. Мороз В.В. _____
Рецензент: доктор фіз.-мат. наук, доц. Кічмаренко О.Д.

Рекомендовано до захисту:

Протокол засідання кафедри

№ ___ від _____ 2024 р.

Завідувач кафедри

_____ _____

Захищено на засіданні ЕК № _____

Протокол № ___ від _____ 2024 р.

Оцінка _____ / _____ / _____

Голова ЕК

_____ _____

Одеса — 2024 р.

ЗМІСТ

ВСТУП

В роботі проводиться аналіз методів виділення звукових образів в аудіосигналах, які мають значні шумові компоненти. Проблема полягає в складності розпізнавання фонових компонент, які, в більшості випадків, носять нерегулярний характер та спотворюють корисний сигнал.

Для розв'язання поставленої задачі застосована емпірична модова декомпозиція EMD [1], яка дозволила розкласти сигнал на адаптивний набір внутрішніх модових функцій (IMFs). Для аналізу особливостей мод використовувалося перетворення Гільберта [2], яке дозволило отримати миттєві амплітуди та частоти для вилучення спектральних і часових ознак [5]. Фільтрація шумових компонентів у складних аудіосигналах здійснюється на основі адаптивного фільтру IMFs, який виділяє найбільш інформативні моди для подальшого аналізу та кластеризації їх ознак. Кластеризація цих ознак дозволяє розпізнати мовні сигнали від фонових перешкод, що сприяє подальшому зниженню шуму та очищенню сигналу."

Актуальність кваліфікаційної роботи полягає у дослідженні потенційного та ефективного методу видалення шумових компонент, та зменшенню обчислювальної складності з використанням спектрального аналізу Гільберта.

Мета роботи полягає в дослідженні та аналізу методів і алгоритмів, спрямованих на розпізнавання та видалення шуму в аудіосигналах різної складності і тривалості для подальшої обробки чистого сигналу.

Об'єктом дослідження є задача розпізнавання звукових образів в умовах значних шумових перешкод.

Предметом дослідження виступають методи та алгоритми розпізнавання голосових і фонових компонентів в аудіосигналах з подальшим видаленням шуму та очищенням сигналу.

Результати роботи були апробовані на XXII-й міжнародній науково-практичній конференції "Математичне та програмне забезпечення інтелектуальних систем (МПЗІС-2024)"20-22 листопада 2024 р., тези були опубліковані: Мороз В.В., Кулик Д.В., «Методи та алгоритми розпізнавання звукових

образів». Математичне та програмне забезпечення інтелектуальних систем (МПЗІС-2024): Тези доповідей XXII Міжнародної науково-практичної конференції, Дніпро, 20-22 листопада 2024 р. – Дніпро: ДНУ, 2024. – 316 с.

РОЗДІЛ 1

ОГЛЯД ПРОБЛЕМИ ОБРОБКИ ЗВУКОВИХ ОБРАЗІВ

1.1 Основі складності в розпізнаванні звукових образів

Із попередньої, дипломної роботи, було виявлено, що більшість відомих методів (віконне Фур'є перетворення (STFT) [6][7], дискретне вейвлет-перетворен-ня (DWT)[8]) добре працюють із музикою. STFT розбиває сигнал на малі часові вікна, в межах яких сигнал можна вважати стаціонарним і, в деякій мірі, періодичним. Це дозволяє аналізувати частотний склад сигналу в кожному вікні, що особливо корисно для розпізнавання музики, де частоти можуть змінюватися з часом. Але такі методи менш ефективні для сигналів, які є сильно нестаціонарними, нелінійними або містять швидкі, неперіодичні зміни (їхні характеристики постійно змінюються і не повторюються). У таких випадках часові вікна можуть бути занадто великими, щоб точно відобразити зміни, або занадто малими, що призводить до поганої частотної роздільної здатності. При дослідженні аудіосигналів були виявленні шумові компоненти, які заважали розпізнати схожі сигнали та змінювали відображення скейлограми.

Шум, або шумові компоненти, у контексті обробки сигналів — це небажані сигнали, які накладаються на корисну інформацію і можуть спотворювати її або ускладнювати її аналіз. Шум може мати різне походження: від електронних перешкод у обладнанні до зовнішніх звукових джерел, таких як транспортні засоби чи людська діяльність. Навіть якщо припустити, що людський голос можна виділити, аналізуючи лише певний діапазон частот (від 300 до 3000 Гц), різноманітний шум може перебувати в будь-якому частотному спектрі, перекриваючи голос та інші важливі елементи сигналу. Наприклад, частоти деяких музичних інструментів можуть повністю збігатися з частотами людського голосу та його характеристиками. Більше того, у ситуаціях, таких як розмова по телефону в аеропорту, фоновий шум

від реактивних двигунів літаків може повністю заглушити голос. Таким чином, розуміння та ефективне усунення шумових компонентів є критично важливим для точної обробки та аналізу сигналів.

Використання нейронних мереж та методів машинного навчання для розпізнавання та видалення шуму звучить привабливо, але на практиці це досить складне завдання. По-перше, такі підходи вимагають значних обчислювальних ресурсів — потужних процесорів або графічних карт, які здатні обробляти великі обсяги даних і виконувати складні обчислення. Процес навчання моделей може займати багато часу, іноді дні або навіть тижні, залежно від складності мережі та обсягу даних. По-друге, для ефективного навчання потрібні великі та різноманітні датасети. На жаль, в інтернеті не так багато доступних наборів даних, які містять записи голосу з різноманітними шумовими компонентами. Це ускладнює можливість створити модель, яка добре узагальнює та працює в реальних умовах. Усе це робить застосування нейронних мереж та машинного навчання менш практичним для цієї задачі, особливо коли ресурси та час обробки обмежені. Тому часто доводиться шукати альтернативні методи, які є менш ресурсомісткими, але все ще ефективними для розпізнавання та видалення шуму.

1.2 Відомі алгоритми та інструменти для видалення шуму

Перед аналізом чи модифікаціями методів для видалення шуму зі звукових сигналів, важливо ознайомитися з наявними продуктами та методами в цій галузі. Маємо наступне:

- Adobe Audition
- Audacity
- iZotope RX
- Waves NS1 Noise Suppressor
- Krisp
- RNNoise
- NVIDIA RTX Voice
- Різноманітні сайти із видаленням шуму онлайн

В більшості випадків програми застосовують плагіни чи інші інструменти, які були побудовані та натреновані на основі нейронних мереж. І, звісно, більшість компаній не розкривають ідеї своїх алгоритмів чи навчання.

Однак, окрім методів на основі машинного навчання та штучного інтелекту, існує потенціал у розробці алгоритмів на основі традиційних математичних методів. Такі алгоритми могли б одразу, без попереднього навчання, ефективно та швидко видаляти шум із сигналів. Використання математичних підходів може дозволити створювати рішення, які не потребують великих обчислювальних ресурсів та об'ємних навчальних датасетів.

Область видалення шуму зі звукових сигналів активно розвивається, особливо з появою методів машинного навчання та штучного інтелекту. Незважаючи на наявність багатьох інструментів, все ще існує великий потенціал для вдосконалення алгоритмів, покращення якості обробки та зменшення вимог до обчислювальних ресурсів. Це робить дослідження в цій сфері актуальними та перспективними.

РОЗДІЛ 2

МАТЕМАТИЧНІ МЕТОДИ РОЗВ'ЯЗУВАННЯ ЗАДАЧІ

Розв'язання задач класифікації, видалення шумових компонент та розпізнавання мовних сигналів вимагає застосування ефективних методів аналізу, що враховують як часову, так і частотну структуру сигналів. Часо-частотний аналіз забезпечує можливість дослідження нестаціонарних сигналів шляхом представлення їх локальних характеристик у двовимірному просторі часу та частоти, що є необхідним для виявлення прихованих закономірностей і значущих компонент. У цьому розділі розглянуті теоретичні основи часо-частотного аналізу, а також методи і засоби, що забезпечують його застосування для обробки складних звукових сигналів.

2.1 Hilbert-Huang transform

Згідно книжці Рам Білас Пачорі «Time-Frequency Analysis Techniques and their Applications»[1] Перетворення Гільберта-Хуанга (Hilbert-Huang transform, HHT) - це двоетапний підхід, який використовується для аналізу нелінійних і нестаціонарних сигналів. Перший етап - це емпіричне модове розкладання (EMD), яке розділяє сигнал на набір обмежених смугою компонентів, відомих як внутрішньо-модові функції (IMFs). Кожна з цих функцій характеризується певними частотними властивостями: найвищі частоти відповідають першим IMF, тоді як найнижчі - останнім. Другий етап передбачає застосування Гільбертового перетворення до отриманих мод для побудови аналітичних функцій (аналітичних IMF), на основі яких можна визначати миттєві амплітуди (AE) та частоти (IF). Сформований набір AE та IF для всіх IMF утворює часово-частотне представлення сигналу (TFR, Time-Frequency Representation), яке і є кінцевим результатом методу HHT.

2.2 Емпіричне модове розкладання

Емпіричне модове розкладання (Empirical Mode Decomposition, EMD), згідно тієї ж книжки, є адаптивним і даними-залежним методом аналізу, який не потребує припущень про стаціонарність чи лінійність сигналу. Цей метод дозволяє розкласти будь-який сигнал із часової області на кінцеву кількість внутрішніх модових функцій (IMF), що представляють основні компоненти розкладу, базуючись на локальних особливостях сигналу.

IMF, отримані за допомогою EMD, мають відповідати наступним умовам:

- **Умова 1:** У межах усього сигналу кількість екстремумів і кількість переходів через нуль повинні бути рівними або відрізнятися не більше ніж на одиницю.
- **Умова 2:** У будь-якій точці сигналу середнє значення оболонок, побудованих на основі локальних максимумів і мінімумів, має дорівнювати нулю.

На основі методу EMD будь-який нестаціонарний сигнал $x(t)$ може бути представлений у вигляді суми внутрішніх модових функцій (IMF) та залишкової компоненти $r_J(t)$:

$$x(t) = \sum_{j=1}^{J} IMF_j(t) + r_J(t), \tag{2.1}$$

де $IMF_j(t)$ – це j-та модова функція, а $r_J(t)$ – залишкова частина, яка є монотонною складовою сигналу.

2.2.1 Основні етапи методу EMD

Процедура емпіричного модового розкладання складається з наступних кроків:

1) **Визначення екстремумів сигналу** $x(t)$**:** знаходження локальних максимумів і мінімумів.
2) **Побудова оболонок:** створення верхньої $E_{\max}(t)$ та нижньої $E_{\min}(t)$ оболонок сигналу за допомогою кубічної сплайн-інтерполяції.

3) **Обчислення середнього значення оболонок:** середнє значення $A(t)$ знаходиться за формулою:

$$A(t) = \frac{E_{\min}(t) + E_{\max}(t)}{2}. \tag{2.2}$$

4) **Визначення залишкової складової:** залишкова складова $D(t)$ обчислюється за формулою:

$$D(t) = x(t) - A(t). \tag{2.3}$$

5) **Перевірка умов IMF:**

- Якщо $D(t)$ задовольняє умови IMF, ця функція вважається внутрішньою модою, і процедура повторюється для залишкової компоненти $x(t) - D(t)$.
- Якщо $D(t)$ не відповідає умовам, розрахунки продовжуються до досягнення допустимого стандартного відхилення σ_k:

$$\sigma_k = \sum_{t=0}^{B} \left[\frac{|D_{k-1}(t) - D_k(t)|^2}{D_{k-1}^2(t)} \right], \tag{2.4}$$

де B – тривалість сигналу, а $D_k(t)$ – залишкова компонента після k-ї ітерації.

Якщо залишковий сигнал $x(t) - D(t)$ є монотонною функцією або подальші IMF не можуть бути отримані, процес EMD завершується. Отримані внутрішні модові функції (IMF), такі як $IMF_1(t), IMF_2(t), \ldots, IMF_N(t)$, представляють компоненти сигналу в різних частотних діапазонах — від високочастотних до низькочастотних.

2.2.2 Приклад розкладання EMD

Розглянемо приклад, який надає Рам Білас Пачорі у своїй книжці. На Рисунку 2.1 показано синтетичний сигнал $x(n)$ із частотою дискретизації $f_s = 1000\,\text{Гц}$ і відповідні IMF, отримані за допомогою методу EMD. Вихідний

сигнал $x(n)$ описується рівнянням:

$$x(n) = \cos\left[2\pi\left(30 + 45\frac{n}{f_s}\right)\frac{n}{f_s}\right] + 0.85\cos\left[2\pi\left(100 + \frac{115n^2}{3f_s^2}\right)\frac{n}{f_s}\right], \quad (2.5)$$

де $n = 0, 1, 2, \ldots, 1000$.

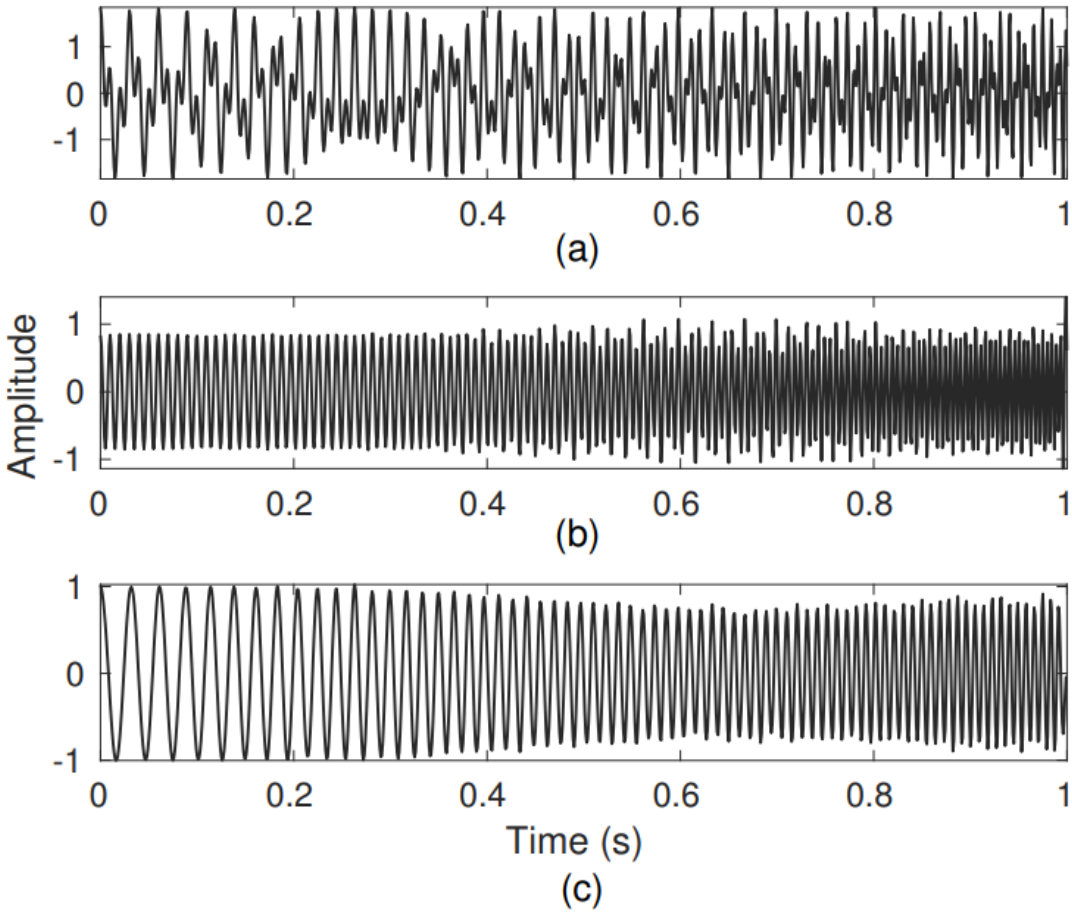

Рис. 2.1. На графіку (a) представлено вихідний сигнал, а на (b) та (c) — отримані IMF, що демонструють зміну частотного складу сигналу.

2.2.3 Часово-частотне представлення

Далі розглянемо блок-схему процесу HHT для отримання часово-частотного представлення (Time-Frequency Representation, TFR). TFR є сукупністю амплітудних (AE) та миттєвих частотних (IF) функцій, отриманих із IMF за допомогою перетворення Гільберта. Кожна IMF обробляється окремо, утворюючи часово-частотну картину сигналу.

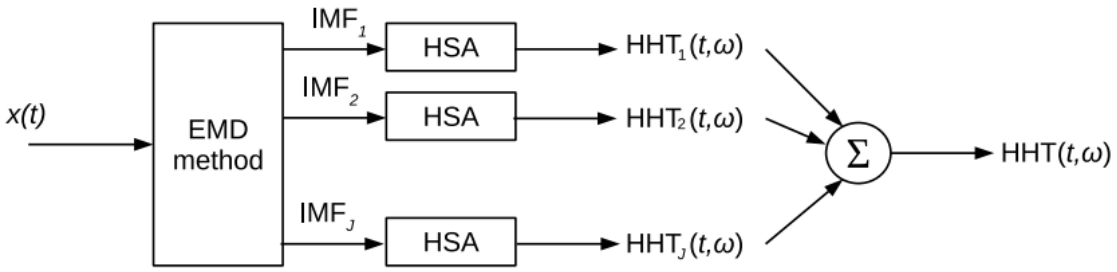

Рис. 2.2. Блок-схема процесу HHT для отримання часово-частотного представлення.

2.3 Варіаційне модове розкладання

Варіаційне модове розкладання (Variational Mode Decomposition, VMD)[9] є сучасним підходом, який долає обмеження методу EMD, зокрема його емпіричну природу, ефект змішування мод та слабкість при роботі з шумом. Метод VMD базується на концепції фільтрації Вінера, що забезпечує високу стійкість до шуму. Основна ідея VMD полягає у розкладанні дійсного сигналу $x(t)$ на набір вузькосмугових компонент (NBC), центрованих навколо відповідних центральних частот.

2.3.1 Математична формалізація

Метод VMD формулюється як задача варіаційної оптимізації для визначення ширини смуг NBC. Математична постановка задачі включає три етапи:

1) **Застосування перетворення Гільберта** до кожної компоненти $x_l(t)$ для отримання її одностороннього частотного спектра.
2) **Зміщення частотного спектра** кожної компоненти з використанням властивостей модуляції.
3) **Оцінка ширини смуги пропускання** для кожної компоненти на основі гладкості у просторі Гауса H^1.

Математична постановка задачі обмеженої варіаційної оптимізації вира-

жається як:

$$\min_{\{x_l\},\{\omega_l\}} \sum_l \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * x_l(t) \right] e^{-j\omega_l t} \right\|_2^2, \tag{2.6}$$

де $\sum_l x_l(t) = x(t)$.

З використанням множників Лагранжа та штрафного коефіцієнта α, задача перетворюється у нестрогу оптимізаційну проблему:

$$\begin{aligned} \mathcal{L}(\{x_l\},\{\omega_l\},\lambda) = \alpha \sum_l \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * x_l(t) \right] e^{-j\omega_l t} \right\|_2^2 \\ + \left\| x(t) - \sum_l x_l(t) \right\|_2^2 + \left\langle \lambda(t), x(t) - \sum_l x_l(t) \right\rangle. \end{aligned} \tag{2.7}$$

Центральні частоти для оновлених компонент розраховуються за формулою:

$$\omega_l^{n+1} = \frac{\int_0^\infty \omega |X_l(\omega)|^2 d\omega}{\int_0^\infty |X_l(\omega)|^2 d\omega}. \tag{2.8}$$

2.3.2 Вибір параметрів методу VMD

Реалізація VMD вимагає налаштування наступних параметрів:

- **Кількість компонент (NBC):** визначає кількість вузькосмугових мод, які необхідно отримати.
- **Штрафний коефіцієнт α:** контролює жорсткість ширини смуги для кожної компоненти.
- **Початкові частоти (ω_{0n}):** задаються як піки у спектрі Фур'є для розподілу компонент.
- **Критерій збіжності (tol):** задає допустиме відхилення між ітераціями.
- **Кількість ітерацій:** визначає максимальну кількість кроків алгоритму.

Ці параметри впливають на точність і збіжність методу та потребують оптимізації залежно від типу сигналу.

2.3.3 Приклад аналізу мовних сигналів

Розглянемо приклад застосування варіаційного модового розкладання до мовного сигналу з бази даних **CMU ARCTIC** із частотою дискретизації 16 kHz, який наведений у [1]. Для реалізації методу VMD були використані наступні параметри:

- **Штрафний коефіцієнт:** $\alpha = 1000$, що контролює жорсткість ширини смуг пропускання компонент.
- **Кількість компонент:** $K = 5$, яка визначає кількість вузькосмугових мод, необхідних для розкладу сигналу.
- **Критерій збіжності:** $tol = 5 \times 10^{-6}$, який встановлює точність завершення ітераційного процесу.
- **Початкові центральні частоти:** ω_{0n} були обрані на основі піків у спектрі Фур'є мовного сигналу.

На рисунку 2.3 представлено початковий мовний сигнал (графік (a)) та відповідні вузькосмугові компоненти (графіки (b)-(f)), отримані за допомогою методу VMD (із використання вікна Хеммінга).

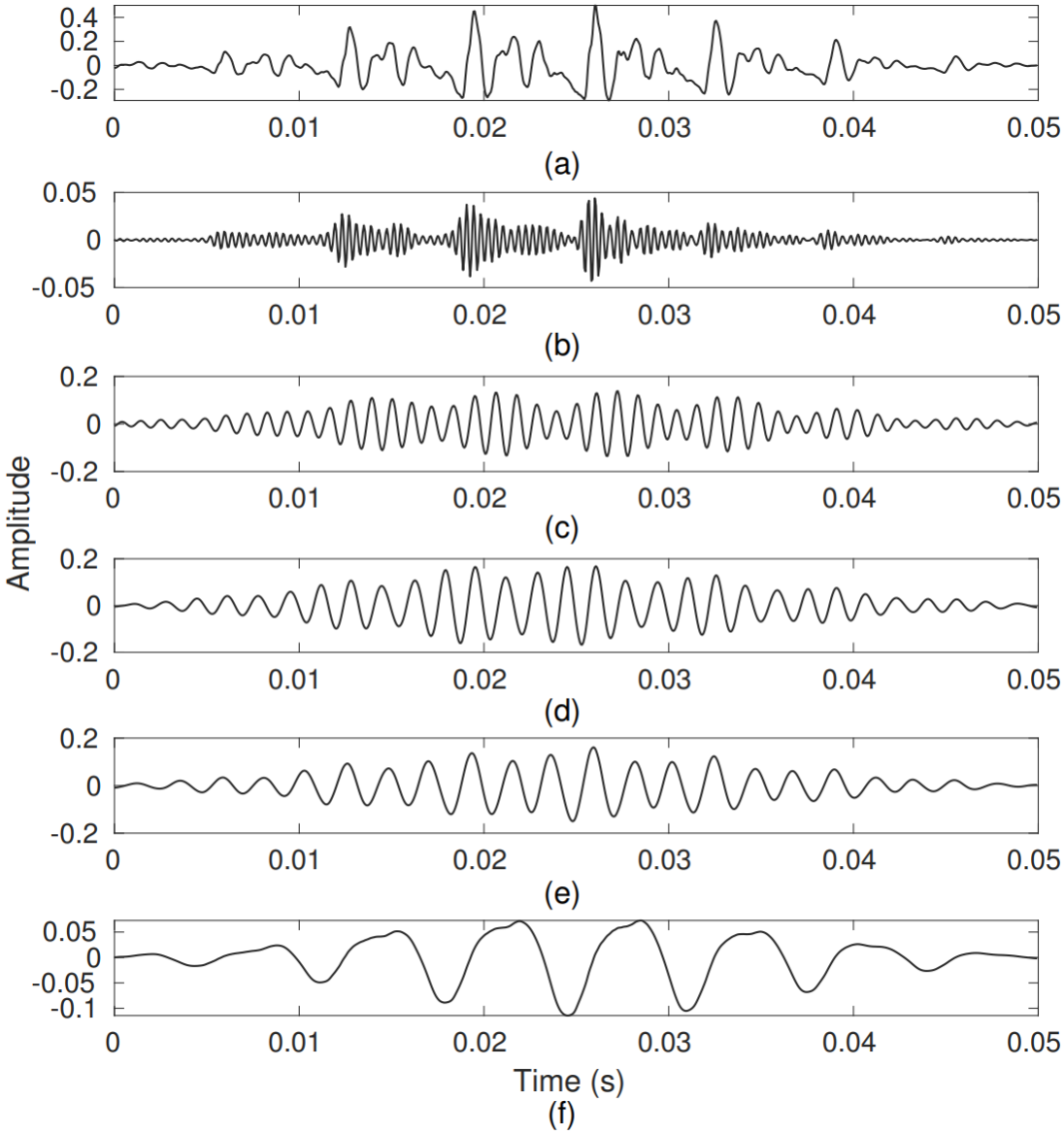

Рис. 2.3. Блок-схема процесу HHT для отримання часово-частотного представлення з книги Рам Білас Пачорі[1].

2.4 Спектральне знешумлювання на основі перетворення Гільберта-Хуанга в поєднанні з F-тестом

Аналізуючи існуючи варіанти методів на основі перетворення Гільберта-Хуанга була знайдена стаття «Spectral denoising based on Hilbert–Huang transform combined with F-test» [10]. В ній використовується HHT та EMD для спектрального аналізу в галузі традиційної китайської медицини.

Логіку роботи методу та застосування F-тесту можна побачити на рисунку 2.4, взятому з оригінальної статті.

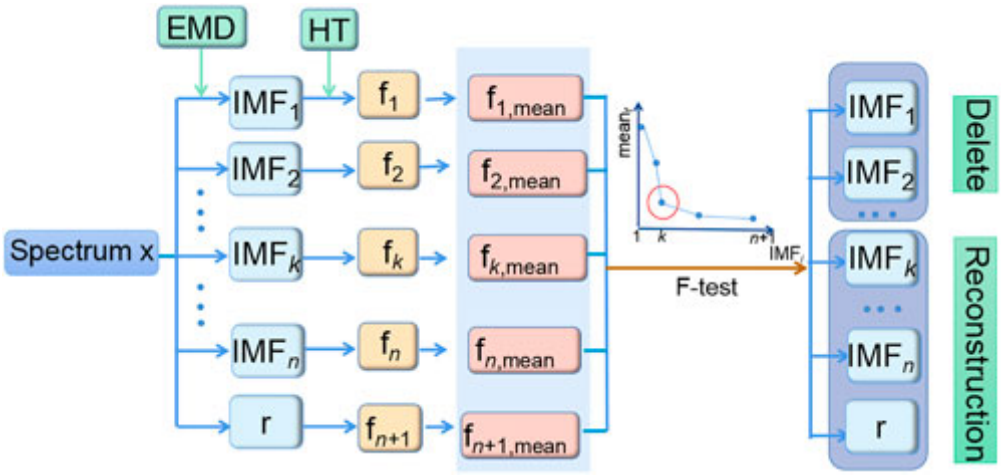

Рис. 2.4. Блок-схема методу спектрального знешумлювання на основі перетворення Гільберта-Хуанга в поєднанні з F-тестом.

Метод очищення сигналу базується на використанні емпіричного модального розкладу (EMD) та перетворення Гільберта (HT). EMD розкладає початковий спектр x на внутрішні моди (IMFs) з різними частотами, а також залишок r. Компоненти з високою частотою вважаються шумом, тоді як низькочастотні моди відповідають корисному сигналу.

1. **Розкладання сигналу**: Сигнал $x(t)$ розкладається на n внутрішніх мод (IMFs) і залишок r за допомогою EMD. Високочастотні моди відповідають шуму, а низькочастотні — основному сигналу:

$$x(t) = \sum_{i=1}^{n} IMF_i(t) + r(t),$$

де $IMF_i(t)$ — внутрішні моди, а $r(t)$ — залишок.

2. **Перетворення Гільберта (HT)**: До кожної моди $IMF_i(t)$ і залишку $r(t)$ застосовується перетворення Гільберта:

$$H(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau,$$

де P — головне значення інтегралу. Це дозволяє обчислити миттєві частоти

за формулами:

$$a(t) = \sqrt{x^2(t) + H^2(t)},$$

$$\theta(t) = \arctan\left(\frac{H(t)}{x(t)}\right),$$

$$\omega(t) = \frac{d\theta(t)}{dt},$$

де $a(t)$ — миттєва амплітуда, $\theta(t)$ — миттєва фаза, а $\omega(t)$ — миттєва частота.

3. **F-тест для визначення шуму**: Для визначення, які моди є шумом, розраховується F-критерій між сусідніми стандартними відхиленнями частот:

$$F_k = \frac{(\mathrm{SD}_k)^2}{(\mathrm{SD}_{k+1})^2},$$

де SD_k і SD_{k+1} — стандартні відхилення частот f_k та f_{k+1}. Значущість значення F_k перевіряється за допомогою F-тесту з рівнем довіри 99.95% (або $p = 0.0005$). Цей рівень означає, що відмінності між модами значущі, якщо ймовірність помилки менша за 0.05%. Ступені свободи для F-тесту встановлені як 2 і 4.

4. **Видалення шуму**: Якщо F-критерій показує значну різницю між сусідніми модами, точка k визначається як поріг. Моди з індексами $1, 2, \ldots, k$ вважаються шумом і видаляються. Корисний сигнал відновлюється шляхом підсумовування решти мод $(IMF_{k+1}, IMF_{k+2}, \ldots, IMF_n)$ і залишку r:

$$x_{\text{denoised}}(t) = \sum_{i=k+1}^{n} IMF_i(t) + r(t).$$

Таким чином, метод дозволяє видаляти шумові частоти, зберігаючи важливі частоти сигналу, що відповідають основному сигналу.

2.5 Мел-частотні кепстральні коефіцієнти

Мел-частотні кепстральні коефіцієнти (Mel-Frequency Cepstral Coefficients, MFCC) є одним із найпоширеніших методів вилучення ознак у задачах обробки мовних сигналів, розпізнавання мови та класифікації аудіо. Вони базуються на людському сприйнятті звуку і дозволяють ефективно

представляти спектральні характеристики сигналу в компактній формі.

2.5.1 Основи мел-шкали

Мел-шкала— це нелінійна частотна шкала, яка краще відображає сприйняття частоти людським вухом. Згідно з експериментальними дослідженнями [3], співвідношення між частотою в герцах f і мел-частотою $m(f)$ визначається формулою:

$$m(f) = 2595 \cdot \log_{10}\left(1 + \frac{f}{700}\right). \tag{2.9}$$

Зворотне перетворення з мел-шкали до частотної:

$$f(m) = 700 \cdot \left(10^{\frac{m}{2595}} - 1\right). \tag{2.10}$$

Це перетворення відображає те, що людське вухо має високу роздільну здатність на низьких частотах і нижчу на високих.

2.5.2 Алгоритм обчислення MFCC

Процес обчислення мел-частотних кепстральних коефіцієнтів складається з наступних етапів [4]:

1) **Підготовка сигналу**:
   - **Фреймування**: Сигнал розбивається на короткі фрейми тривалістю від 20 до 40 мс із перекриттям (зазвичай 50%). Це необхідно, оскільки мовні сигнали можна вважати квазістаціонарними на таких інтервалах.
   - **Віконування**: До кожного фрейму застосовується віконна функція (наприклад, вікно Хеммінга) для зменшення спектральних витоків.
2) **Обчислення спектру**:
   - **Дискретне перетворення Фур'є (DWT)**: Для кожного фрейму обчислюється DWT, що дає комплексний спектр $X(k)$.

- **Амплітудний спектр**: Визначається як модуль спектра $|X(k)|$.

3) **Застосування мел-фільтрового банку**:

- **Побудова фільтрів**: На мел-шкалі визначається набір трикутних фільтрів, які перекриваються і покривають весь діапазон частот, що цікавить.
- **Фільтрація**: Амплітудний спектр множиться на кожен з фільтрів, після чого енергії під фільтрами сумуються:

$$S_m = \sum_{k=0}^{N-1} |X(k)|^2 H_m(k), \tag{2.11}$$

де $H_m(k)$ — частотна характеристика m-го фільтра, N — кількість точок DWT.

4) **Логарифмування енергій**:

$$\hat{S}_m = \ln(S_m). \tag{2.12}$$

5) **Дискретне косинусне перетворення (DCT)**:

- **Обчислення кепстральних коефіцієнтів**:

$$c_n = \sum_{m=1}^{M} \hat{S}_m \cos\left[\frac{\pi n}{M}\left(m - \frac{1}{2}\right)\right], \quad n = 1,2,\ldots,L, \tag{2.13}$$

де M — кількість мел-фільтрів, L — кількість бажаних коефіцієнтів.

6) **Додаткові коефіцієнти**:

- **Дельта-коефіцієнти**: Для врахування динаміки сигналу обчислюються перші та другі похідні кепстральних коефіцієнтів.

2.5.3 Математичні вирази

Дискретне перетворення Фур'є (DWT):

$$X(k) = \sum_{n=0}^{N-1} x(n)w(n)e^{-j2\pi kn/N}, \tag{2.14}$$

де $x(n)$ — фреймований сигнал, $w(n)$ — віконна функція.

Енергія в мел-фільтрах:

$$S_m = \sum_{k=0}^{N-1} |X(k)|^2 H_m(k). \tag{2.15}$$

Кепстральні коефіцієнти:

$$c_n = \sum_{m=1}^{M} \ln(S_m) \cos\left[\frac{\pi n}{M}\left(m - \frac{1}{2}\right)\right]. \tag{2.16}$$

2.5.4 Дельта та дельта-дельта коефіцієнти

Для моделювання динамічних властивостей сигналу використовуються дельта-коефіцієнти (перші похідні) та дельта-дельта коефіцієнти (другі похідні):

$$\Delta c_n = \frac{\sum_{l=1}^{L} l(c_{n+l} - c_{n-l})}{2\sum_{l=1}^{L} l^2}, \tag{2.17}$$

де L — довжина вікна для обчислення похідних.

2.6 Особливості для класифікації мови/музики на основі спектру Гільберта

Інша ідея для видалення шумових компонент була представлена у статті «Hilbert Spectrum Based Features for Speech/Music Classification» [5]. У статті описується метод класифікації аудіосигналів на мовні та музичні з використанням перетворення Гільберта-Хуанга (HHT) та емпіричної декомпозиції мод (EMD). Запропонована методика передбачає розклад сигналу

на амплітудно-частотні компоненти (IMFs), з яких отримуються нові кепстральні ознаки, засновані на миттєвих амплітудах (IA) і частотах (IF). Ці ознаки використовуються для класифікації сигналів за допомогою машинного навчання. Для перевірки ефективності методу проведено експерименти на трьох базах даних (S&S, GTZAN та MUSAN).

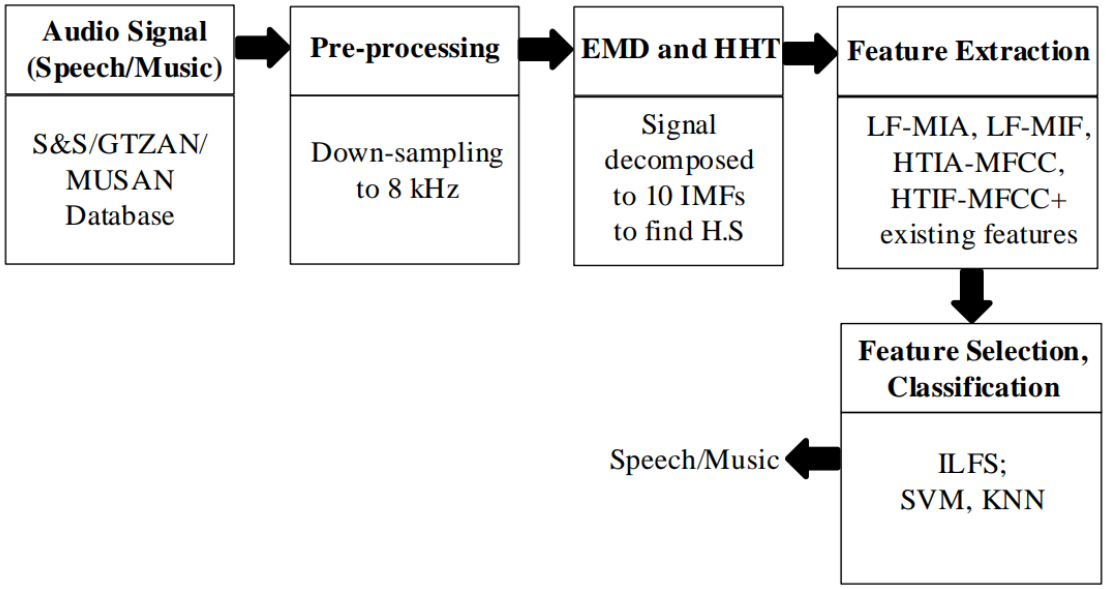

Рис. 2.5. Процес отримання результату на основі спектру Гільберта

2.6.1 Отримання ознак HTIA-MFCC та HTIF-MFCC

У процесі аналізу статті особливу увагу було приділено таким ознакам, як HTIA-MFCC (Hilbert-Huang Transform-Instantaneous Amplitude Mel-Frequency Cepstral Coefficients) та HTIF-MFCC (Hilbert-Huang Transform-Instantaneous Frequency Mel-Frequency Cepstral Coefficients), які показали найкращі результати при класифікації. Ці ознаки поєднують переваги перетворення Гільберта-Хуанга (HHT), яке дозволяє аналізувати миттєві амплітуди (IA) та частоти (IF), із мел-фільтровим банком (Mel Filter Bank), що підвищує роздільну здатність у низькочастотній області сигналу.

Методика отримання цих ознак передбачає наступні етапи:

1) **Розклад сигналу** за допомогою емпіричної модової декомпозиції (EMD) для вилучення 10 внутрішніх модових функцій (IMFs).
2) **Демодуляція (HT):** застосування перетворення Гільберта для обчислення миттєвих амплітуд (IA) та частот (IF) для кожної з модових

функцій.

3) **Використання вікон і сегментація:** IA/IF поділяються на вікна тривалістю 30 мс із перекриттям 50%.
4) **Частотний аналіз:** до кожного вікна застосовується швидке перетворення Фур'є (FFT), після чого спектр сигналу передається на *Mel-фільтри*.
5) **Логарифмування і DCT:** на вихід Mel-фільтрів обчислюються логарифм та дискретне косинусне перетворення (DCT) для отримання 39 коефіцієнтів для кожної IMF.

Кінцевим результатом є отримання двох наборів ознак:

- **HTIA-MFCC:** базуються на миттєвих амплітудах модових функцій.
- **HTIF-MFCC:** враховують миттєві частоти.

Згідно з результатами дослідження, ці ознаки дозволяють зберігати гармонійну структуру сигналу та забезпечують підвищену роздільну здатність у низькочастотній області. Вони використовуються для побудови векторів ознак, які можуть бути застосовані у задачах класифікації аудіосигналів. Методика отримання ознак представлена на Рисунку 2.6 у вигляді блок-схеми.

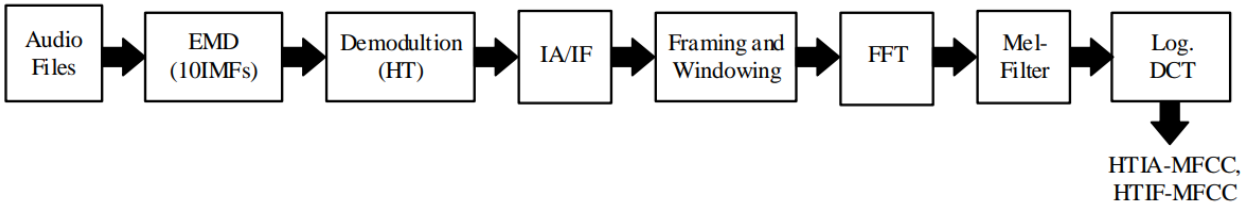

Рис. 2.6. Блок-схема особливостей HTIA-MFCC/HTIF-MFCC

РОЗДІЛ 3

АНАЛІЗ ТА МОДИФІКАЦІЯ МЕТОДІВ РОЗПІЗНАВАННЯ ТА ВИДАЛЕННЯ ШУМОВИХ КОМПОНЕНТІВ В АУДІОСИГНАЛАХ

3.1 Тестові звукові образи

Для аналізу та подальшої модифікації методів було використано аудіозаписи з відкритих джерел, що містять голоси людей, будь то розмова по телефону, або монолог, що містять велику кількість шумових компонентів та сторонніх, небажаних звуків. Для тесту та перевірки моделі використовувалося чотири аудіозаписи:

- Розмова по телефону клієнта банку та працівника в зашумленому офісі, із сторонніми голосами на фоні, дзвінками телефонів та поганою якістю запису розмови по телефону.
- Запис репортажу двох людей у гелікоптері. Тут за шум виступає звук роботи несучого гвинта, повітря та якість запису розмови на мікрофон, яка є приглушеною.
- Диктор зачитує текст із оповіді, у той час як на фоні злітає літак. Звуки реактивних двигунів значно гучніші, ніж монолог людини.
- Диктор зачитує текст із оповіді, а на фоні чутно клацання клавіатури. Звуки клацання тихіші, ніж голос.

Ціль роботи методу полягає в тому, щоб усі сторонні голоси, дзвінки, клацання, гупотіння та інші шумові компоненти були видалені з аудіо, при цьому зберігаючи якість та гучність основних голосів. Також важливою компонентою є час виконання коду, адже в пріоритеті саме отримання якісного результату за коротких проміжок часу.

3.2 Класичні математичні методи для розпізнавання та видалення шумових компонентів

Перш ніж перейти до складних часо-частотних методів розглянемо роботу класичних математичних методів, які в теорії можуть швидко виявити та видалити шум із аудіосигналу. Для оцінки якості знешумлювання сигналу будемо користуватися метрикою Signal-to-noise ratio (SNR) для порівняння рівня чистого сигналу із рівнем шуму. Також слід прослуховувати отримані сигнали, адже SNR не завжди вказує на гарний результат.

Також згідно статті «Hilbert Spectrum Based Features for Speech/Music Classification» [5] було використано частоту дискретизації у 8000 гЦ для зменшення часу обробки при збережені достатньої кількості корисної інформації про звуковий образ.

3.2.1 High-Pass Filtering

Високочастотне фільтрування (High-Pass Filtering) є одним із розповсюджених методів для обмеження шумових компонентів. Було спробовано фільтром пропустити частоти, що перевищують заданий поріг, та пригнітити низькочастотні компоненти сигналу. Згідно теорії - це повинно видалити низькочастотний шум, такий як гул або реверберація, з одночасним збереженням корисних високочастотних компонентів (мовлення).

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
High-Pass Filtering	-0.70	2.99	-0.47	7.98	-5.36	1.99	-0.69	2.99

Табл. 3.1. Результати видалення шуму методом High-Pass Filtering.

Результат фільтрації не прийнятний. SNR вийшов від'ємним. Це вказує на те, що шум гучніший, ніж голос. Багато шумових компонентів залишилося і якість голосу також стала гірше. Поглянемо на отримані графіки

першого сигналу для візуалізаційного розуміння роботи методу. Виводити графіки всіх аудіосигналів немає сенсу, адже розглянутий метод працює не ідеально, а отримати розуміння можна і з одного прикладу, щоб не показувати марну інформацію.

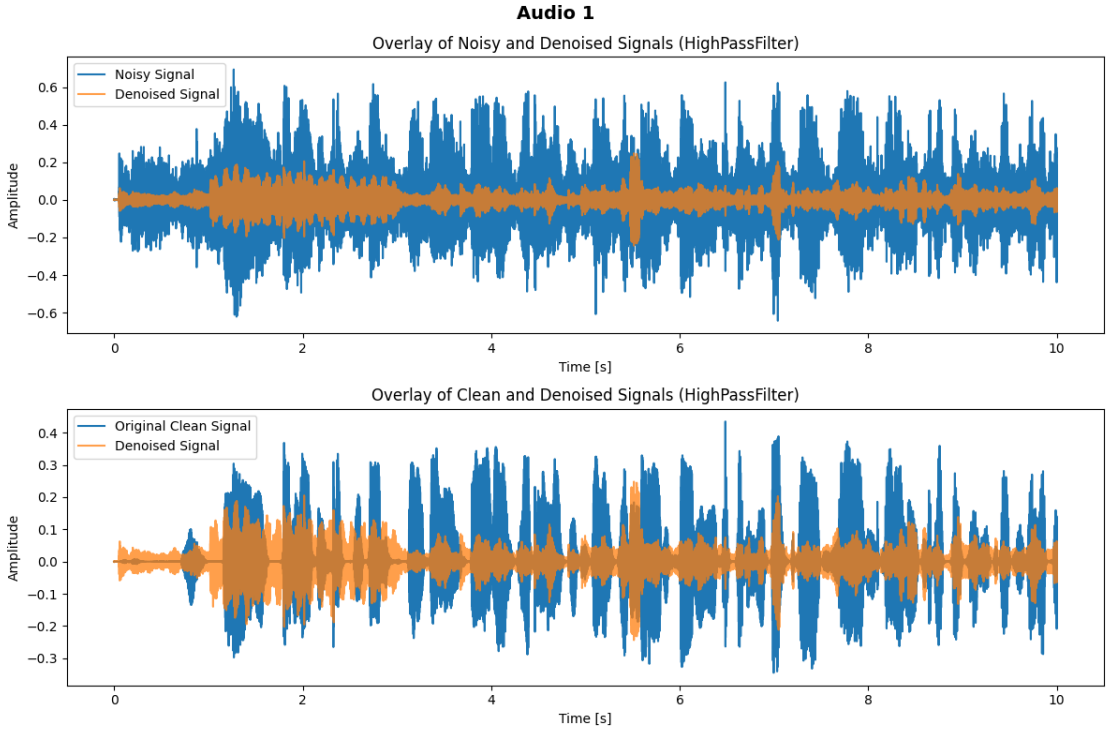

Рис. 3.1. Графіки порівняння для High-Pass Filtering.

3.2.2 PCEN

PCEN (Per-Channel Energy Normalization) є методом нормалізації енергії по каналах, який покращує співвідношення сигнал/шум шляхом нелінійного перетворення спектрограми. Цей метод адаптивно пригнічує постійний фоновий шум та підсилює транзієнтні події в аудіосигналі.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
PCEN	-1.08	3204.19	-2.13	5934.62	-2.37	1317.48	-0.95	2981.04

Табл. 3.2. Результати видалення шуму методом PCEN.

Результат фільтрації також не прийнятний. Багато шумових компонентів

залишилося і якість голосу стала гірше. Переглянемо графіки третього сигналу.

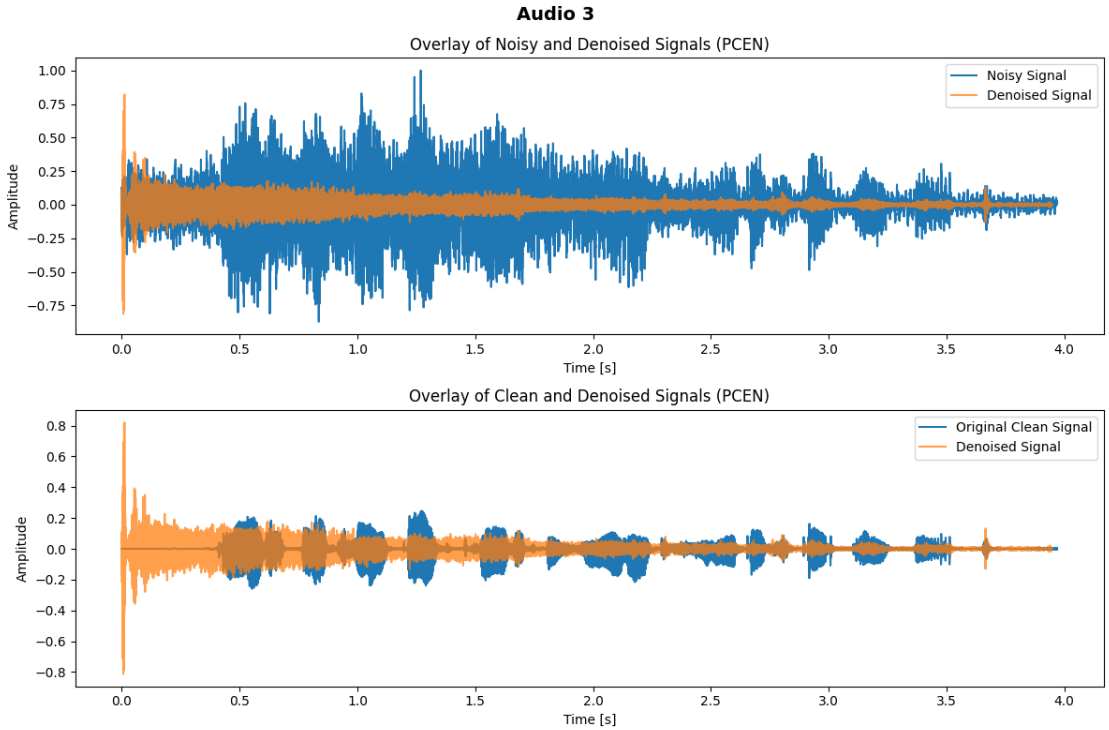

Рис. 3.2. Графіки порівняння для PCEN.

3.2.3 Spectral Gating

Спектральні ворота (Spectral Gating)[11] передбачає перетворення сигналу в частотну область за допомогою швидкого перетворення Фур'є (FFT), аналіз амплітудних спектрів та пригнічення частотних компонентів, що знаходяться нижче певного порогу. Поріг встановлюється на основі оцінки рівня фонового шуму (перші відліки сигналу). Ефективний для сигналів зі стаціонарним або квазістаціонарним шумом, де корисний сигнал має відмінні спектральні характеристики від шуму.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Spectral Gating	2.38	160.58	1.96	256.16	1.69	46.88	2.39	135.64

Табл. 3.3. Результати видалення шуму методом Spectral Gating.

Отримали значно кращі результати, ніж при попередніх методах. Прослуховуючи аудіо можна відмітити, що фоновий шум, такий як звук роботи несучого гвинта, вітер та навіть звук реактивного двигуна доволі непогано спробувало видалити. Хоча і не без недоліків. Такі шумові компоненти, як дзвін телефонів, фонові голоси, або високі частоти двигунів літака все ще залишилися. Голос людей на передньому плані став тихішим. Для аналізу результатів поглянемо на отриману фільтрацію «важких» сигналів (1-й та 3-й) (оскільки вони мають або складні шумові компоненти, або розмову людей на другому плані), а також на досить непогану фільтрацію 2-го сигналу.

Перший аудіосигнал

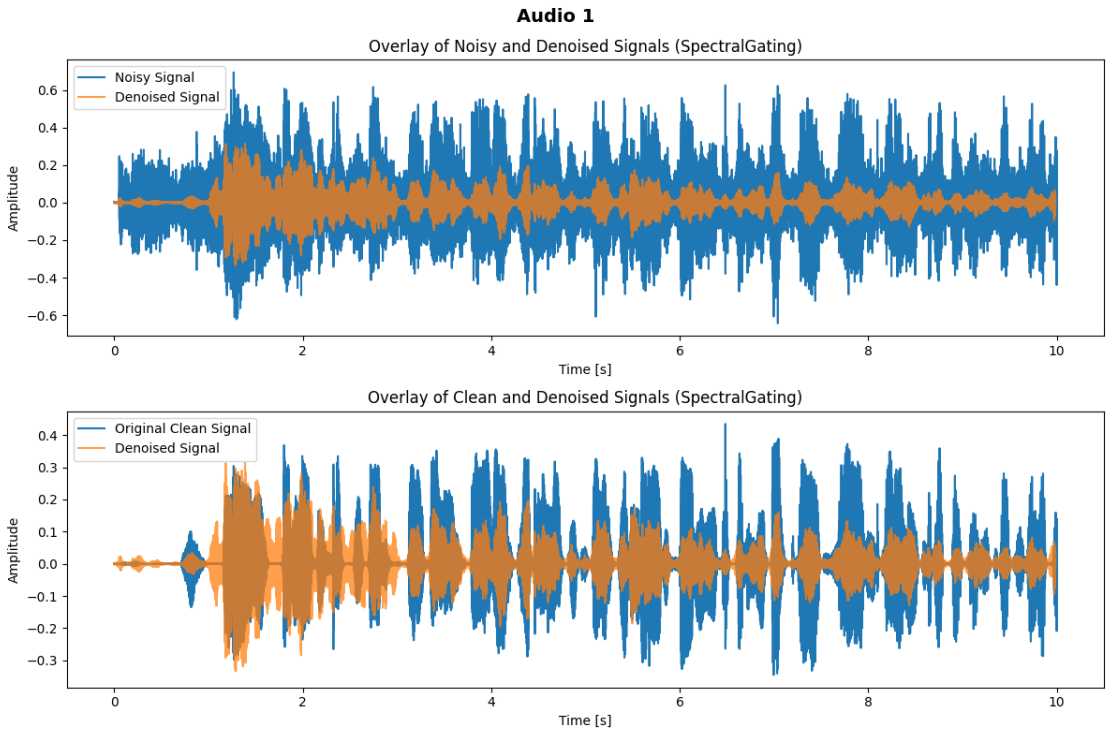

Рис. 3.3. Графіки порівняння 1-го аудіосигналу для Spectral Gating.

Другий аудіосигнал

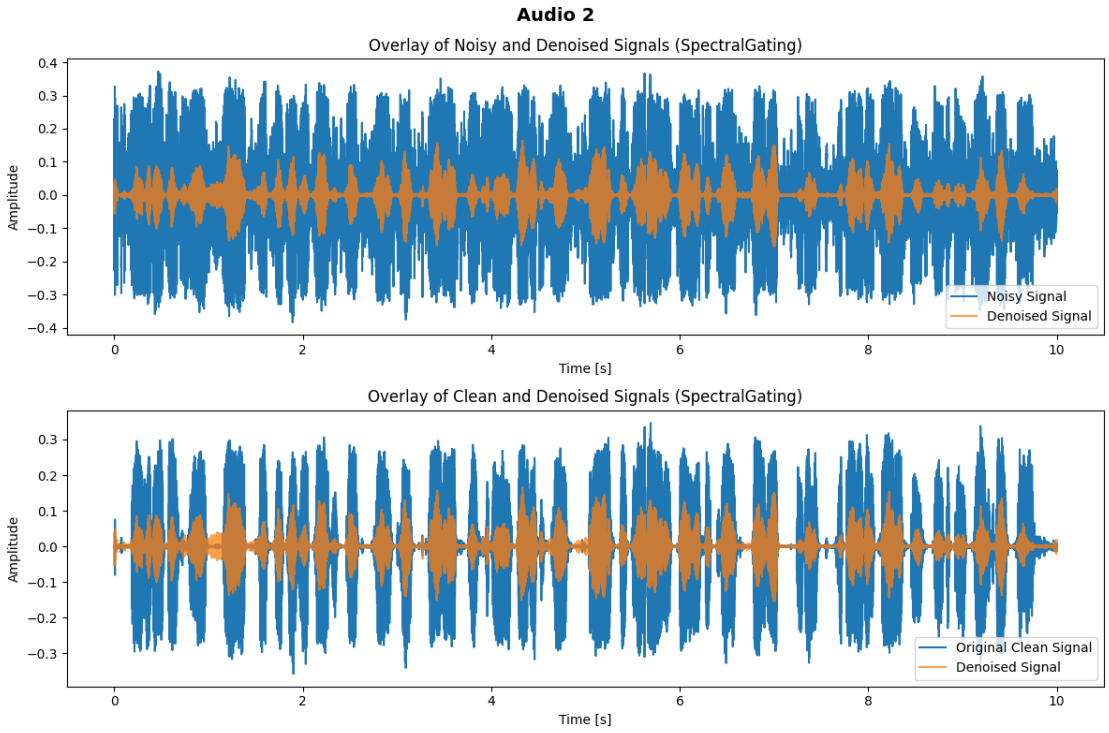

Рис. 3.4. Графіки порівняння 2-го аудіосигналу для Spectral Gating.

Третій аудіосигнал

Рис. 3.5. Графіки порівняння 3-го аудіосигналу для Spectral Gating.

Говорячи про останній, 4-й сигнал то метод хоча і приглушив клацання

клавіатури, але воно все одно чутно на фоні спокійного голосу диктора. Графік виглядає наступним чином.

Четвертий аудіосигнал

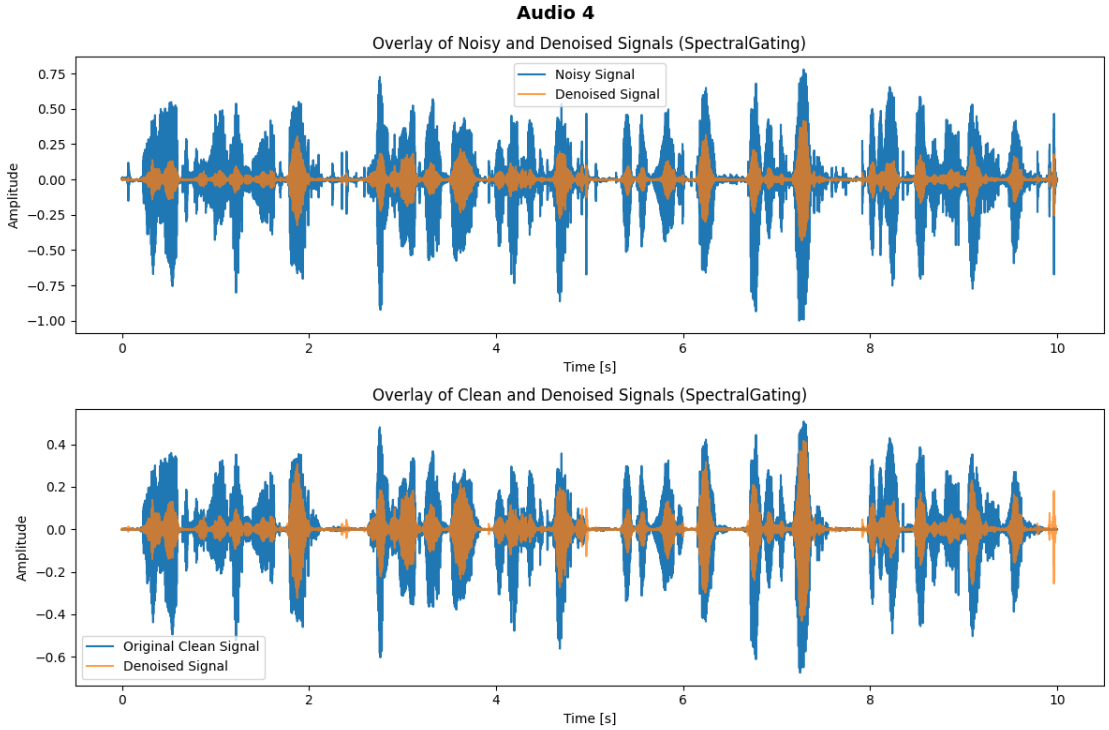

Рис. 3.6. Графіки порівняння 4-го аудіосигналу для Spectral Gating.

3.2.4 Wiener Filtering

Фільтр Вінера (Wiener Filtering) є оптимальним лінійним фільтром, який мінімізує середньоквадратичну помилку між оціненим та істинним сигналом. Він враховує статистичні властивості сигналу та шуму для побудови фільтра, який максимально відновлює початковий сигнал.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Wiener Filtering	-0.65	8.98	4.51	22.95	-9.51	2.00	5.22	7.00

Табл. 3.4. Результати видалення шуму методом Wiener Filtering.

Дивлячись на результати по SNR для 2 і 4 аудіо можна припустити, що цей метод працює краще, ніж попередні. Натомість це не зовсім так. Якщо

прослухати отримані сигнали, то різниці майже не почути. Вони все ще мають шум, і не дуже зрозуміло, що видалило. Для прикладу подивимось на графіки отриманих результатів для 2-го аудіозапису.

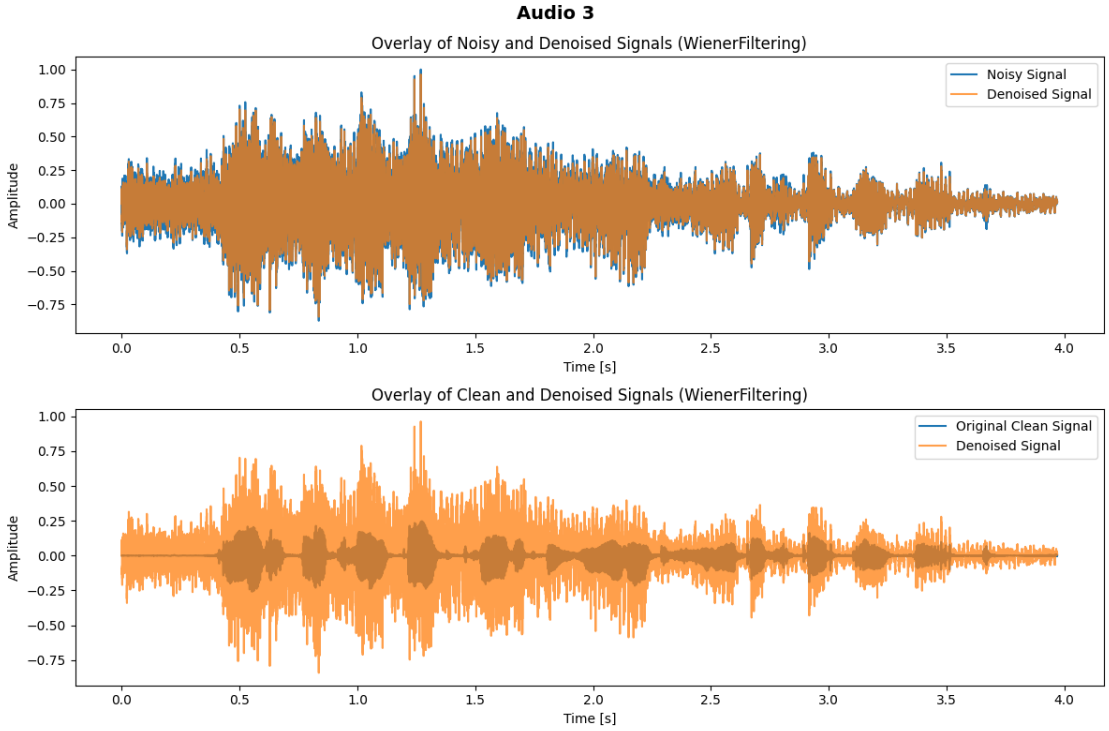

Рис. 3.7. Графіки порівняння для Wiener Filtering.

З графіків бачимо, що різниці майже нема, а SNR став краще. Це пов'язано із тим, що зберігається шум на частотах сигналу. Тобто якщо шум і сигнал перекриваються в частотній області, Wiener фільтр не зможе повністю відокремити їх. Отриманий результат може призводити до поліпшення SNR за рахунок згладжування або зменшення шуму в деяких областях, але шум залишається в тих частотах, які сприймаються людським вухом як ключові для аудіо. Також в Wiener фільтрі може бути переусереднення. Фільтр ґрунтується на припущенні про стаціонарність шуму і сигналу, і може переусереднювати дані, втрачаючи важливі деталі. Це створює ілюзію «поліпшення» метрики SNR, але насправді залишається шум або додається «розмитість» до сигналу. І по графікам дійсно видно, що для короткочасного шумового компоненту метод намагається видалити шум, але не справляється до кінця. Отже SNR не завжди вказує на якість результату, хочу і в цьому випадку шум став трохи меншим, згідно графікам, ніж був.

3.2.5 Wavelet Denoising

Вейвлетне знешумлювання (Wavelet Denoising) включає розкладання сигналу на вейвлетні коефіцієнти, застосування порогового значення для видалення шумових компонентів та реконструкцію сигналу з очищених коефіцієнтів. Вейвлети дозволяють ефективно аналізувати сигнали, які мають нестаціонарні характеристики.

Метод	Аудіо 1		Аудіо 2		Аудіо 3		Аудіо 4	
	SNR (дБ)	Час (мс)	SNR (дБ)	Час (мс)	SNR (дБ)	Час (мс)	SNR (дБ)	Час (мс)
	10 с		10 с		3 с		10 с	
Wavelet Denoising	4.12	2.99	6.34	4.99	-3.04	1.97	8.09	3.99

Табл. 3.5. Результати видалення шуму методом Wavelet Denoising.

Знов отримали доволі непогані результати, але при прослуховуванні сигналів з'являється помітне потріскування після знешумлювання. Також поглянемо на отримані графіки для 2-го аудіозапису.

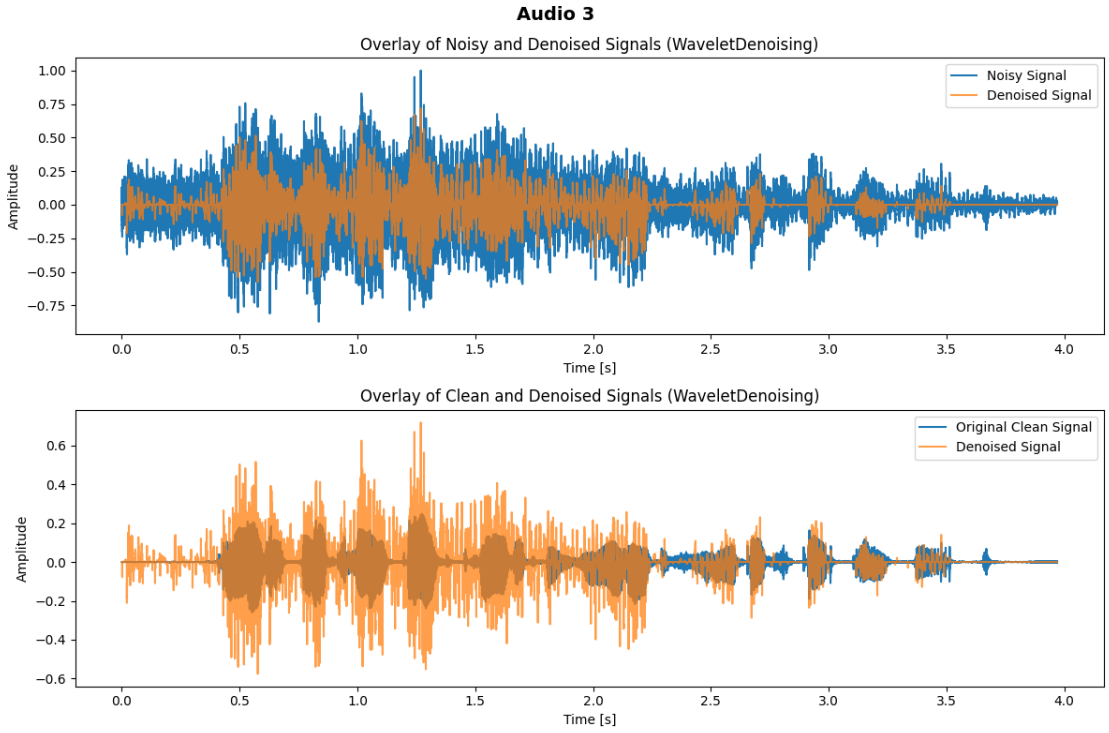

Рис. 3.8. Графіки порівняння для Wavelet Denoising.

Метод намагається видалити шум, але все ще залишається багато зайвого в отриманому аудіо.

3.2.6 Median Filtering

Медіанне фільтрування є нелінійним методом, де кожна точка сигналу замінюється на медіанне значення сусідніх точок у заданому вікні. Це ефективно для видалення імпульсного (сплескового) шуму без розмиття корисних деталей сигналу.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Median Filtering	-0.73	4.98	4.40	9.21	-9.38	1.99	5.14	6.00

Табл. 3.6. Результати видалення шуму методом Median Filtering.

Ситуація аналогічна фільтру Вінера. Знову 2 та 4 сигналі мають кращі результати, у той час як для 1 і 3 значення від'ємні. Цього разу поглянемо на отримані графіки 4-го аудіо.

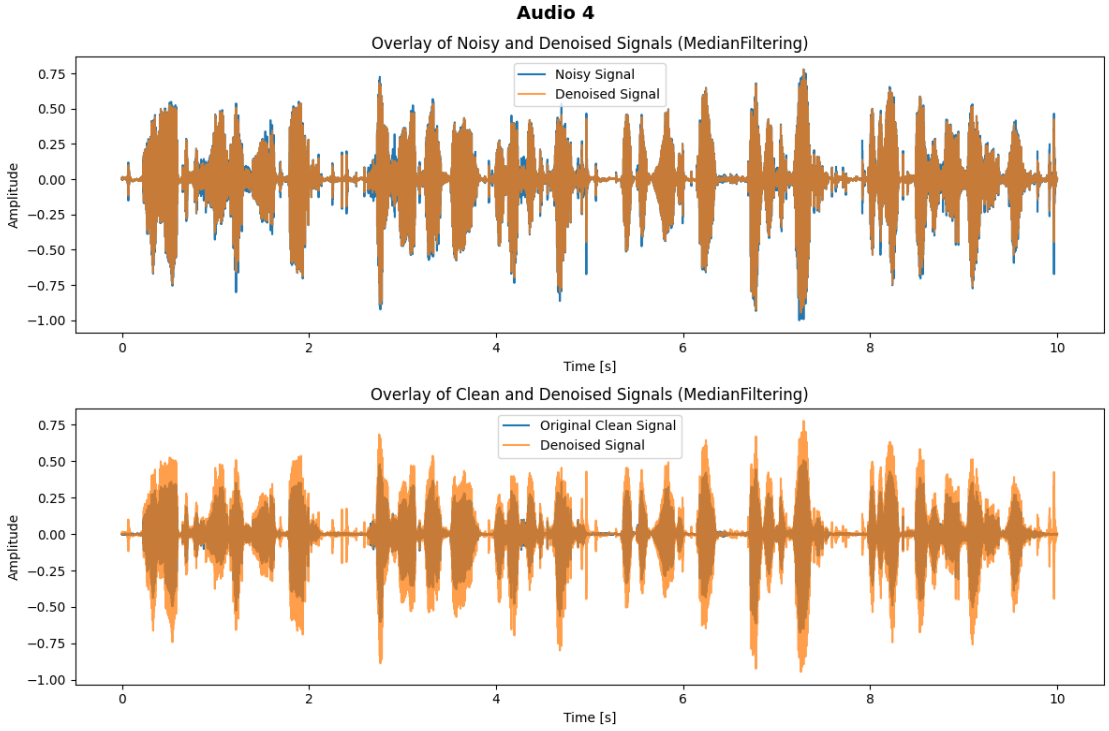

Рис. 3.9. Графіки порівняння для Median Filtering.

Повторюється ситуація, що для короткочасних потужних шумовий компонент, як то оберт несучого гвинта чи натискання клавіші, метод намагається видалити шум, але доволі слабо.

3.2.7 Spectral Subtraction

Метод Спектрального віднімання (Spectral Subtraction) полягає в оцінці спектра шуму та його відніманні від спектра зашумленого сигналу. Після цього виконується зворотне перетворення в часову область для отримання очищеного сигналу. Зазвичай використовується для зменшення рівня стаціонарного або квазістаціонарного шуму в аудіосигналах, таких як гул або шипіння, де спектр шуму можна точно оцінити.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Spectral Subtraction	4.13	60.84	9.22	167.55	-3.80	31.91	8.70	54.85

Табл. 3.7. Результати видалення шуму методом Spectral Subtraction.

Отримали непогані результати. Якщо прослухати та поглянути на результати, що спектральне віднімання дійсно непогано працює на стаціонарному шумі. Аудіо 2, із записом розмови в гелікоптері, показало найкращий результат у знешумлюванні.

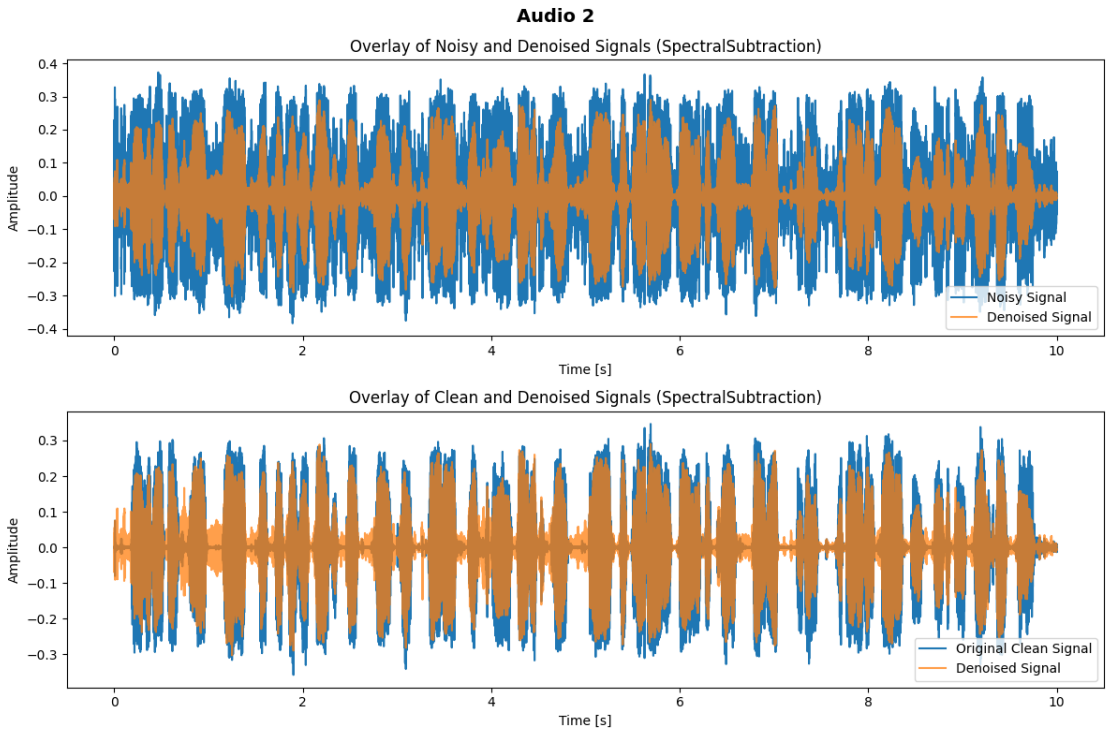

Рис. 3.10. Графіки порівняння для Spectral Subtraction.

В частотній області метод намагається зберегти голос, у той час як на паузах розмов знижується гучність. Прослуховуючи сигнали також можна почути значне видалення шуму.

Хоча для аудіо 1 (із розмовами людей в офісі), в якого додатній SNR із значенням 4.13, а також для аудіо 3 (монолог із фоновим звуком двигуна літака) з від'ємним SNR у -3.80, ситуація значно інша. Для 3-го аудіо спектральне віднімання не справляється із гучним протяжним шумом двигунів. А у першого - на фоні розмови, що для методу не розпізнається як шум. Для прикладу поглянемо на отримані графіки третього аудіосигналу.

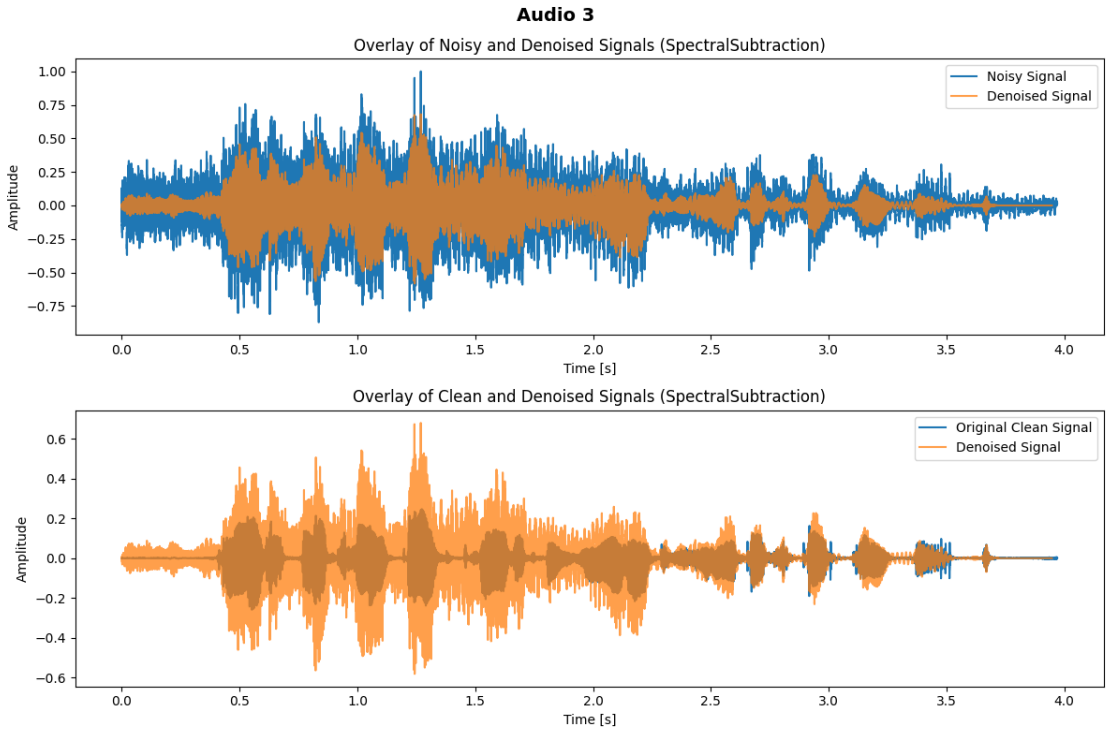

Рис. 3.11. Графіки порівняння для Spectral Subtraction.

3.2.8 Таблиця результатів та висновки

Метод	Аудіо 1		Аудіо 2		Аудіо 3		Аудіо 4	
	SNR (дБ)	Час (мс)	SNR (дБ)	Час (мс)	SNR (дБ)	Час (мс)	SNR (дБ)	Час (мс)
	10 с		10 с		3 с		10 с	
High-Pass Filtering	-0.70	2.99	-0.47	7.98	-5.36	1.99	-0.69	2.99
PCEN	-1.08	3204.19	-2.13	5934.62	-2.37	1317.48	-0.95	2981.04
Spectral Gating	2.38	160.58	1.96	256.16	1.69	46.88	2.39	135.64
Wiener Filtering	-0.65	8.98	4.51	22.95	-9.51	2.00	5.22	7.00
Wavelet Denoising	4.12	2.99	6.34	4.99	-3.04	1.97	8.09	3.99
Median Filtering	-0.73	4.98	4.40	9.21	-9.38	1.99	5.14	6.00
Spectral Subtraction	4.13	60.84	9.22	167.55	-3.80	31.91	8.70	54.85

Табл. 3.8. Результати класичних методів для видалення шуму.

Після огляду класичних методів можна впевнено сказати, що хоч вони і намагаються видалити шумові компоненти, але не справляються із задачею до кінця. Згідно отриманим результатам по SNR та експертної оцінки при прослуховуванні, найкращим результатом є Spectral Gating, який працює для усіх чотирьох аудіо з різноманітними шумовими компонентами. Слід також зазначити, що проаналізовані методи працюють досить швидко, окрім PCEN, який працює Розглянуті методи можуть бути використані у більш складних методах для отримання кращих результатів.

3.3 Перетворення Гільберта-Хуанга для розпізнавання та видалення шумових компонентів

Відомим методом для часо-частотного аналізу є перетворення Гільберта-Хуанга (HHT) в поєднанні з Empirical Mode Decomposition (EMD). Вхідний аудіосигнал потрібно розкласти на внутрішні моди (intrinsic mode functions

(IMFs)) саме за допомогою EMD, а далі виконати HHT для кожної IMF. Але є велика проблема, яка не дозволяє виконати такий метод на аудіосигналах - це час його роботи. Приклади в інтернеті зазвичай базуються на простих графіках функцій, таких як Синусоїда, Експонента, Парабола, Логарифмічна функція і тому подібне. На таких простих прикладах метод працює відносно швидко. А на побутових аудіосигналах вже починаються проблеми. Сьогодні більшість аудіосигналів мають частоту дискретизації 44100 Гц. З такою частотою та довжиною сигналу в 10 секунд метод буде працювати нескінченність. Для прикладу, з частотою дискретизації у 8000 Гц та довжиною у 3 секунди алгоритм працює приблизно 340 мілісекунд, а вже збільшуючи довжину до 4 секунд - працює більше 2-х годин без отримання результатів. Така швидкість обробки є неприпустимою для задач розпізнавання із видаленням шуму, адже одна із ідей використання математичних методів, а не нейронних мереж, полягала в швидкості обробки сигналу. Натомість, якщо метод досить швидко працює на невеликих сигналах, то чому б не спробувати зробити віконну функцію, наприклад Hamming Window[12], для обробки маленьких вікон аудіосигналу для збільшення швидкості.

Слід зазначити, що в процесі тестів звичайного методу HHT та EMD було спробовано використати Варіаційну модову декомпозицію (Variational mode decomposition, VMD), адже згідно теорії метод варіаційної модової декомпозиції (VMD) долає обмеження методу EMD. В цілому, VMD демонструє вищу швидкість роботи порівняно з EMD, проте його універсальність обмежена. Метод вимагає налаштування шести параметрів, значення яких залежать від конкретного аудіосигналу, що ускладнює створення адаптивного алгоритму для автоматичного визначення цих параметрів.

3.4 Спектральне знешумлювання на основі перетворення Гільберта-Хуанга в поєднанні з F-тестом

Розглянувши та проаналізувавши теоретичну частину, що була вказана у статті[10], було модифіковано метод для використання із віконною функцією.

Це дозволить працювати не тільки з простими сигналами, а і з побутовими аудіосигналами, як то мовлення чи музика. Розмір вікна залежить від довжини аудіо та задається вручну. Решта кроків залишається згідно статті.

Поглянемо на таблицю результатів.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Window EMD F-test	-0.01	49931.53	-0.06	25108.25	-6.43	157517.94	-0.01	43515.68

Табл. 3.9. Результати видалення шуму методом HHT в поєднанні з F-тестом.

Отримані результати для поточної задачі видалення шумових компонентів із мовлення людини показали незадовільний результат. Метод працює довго, а отриманий аудіофайл взагалі спотворився та не має чітких звуків. Метод знешумлення призвів до значного зменшення амплітуди сигналу, що може свідчити про втрату енергії корисного сигналу. Натомість як для аналізу та знешумлювання простих графіків функцій, або, наприклад, отриманих сигналів з електрокардіографії - то метод працює чудово та відносно швидко, адже сигнали зазвичай короткі. Розглянемо те, як було знешумлено перший аудіозапис.

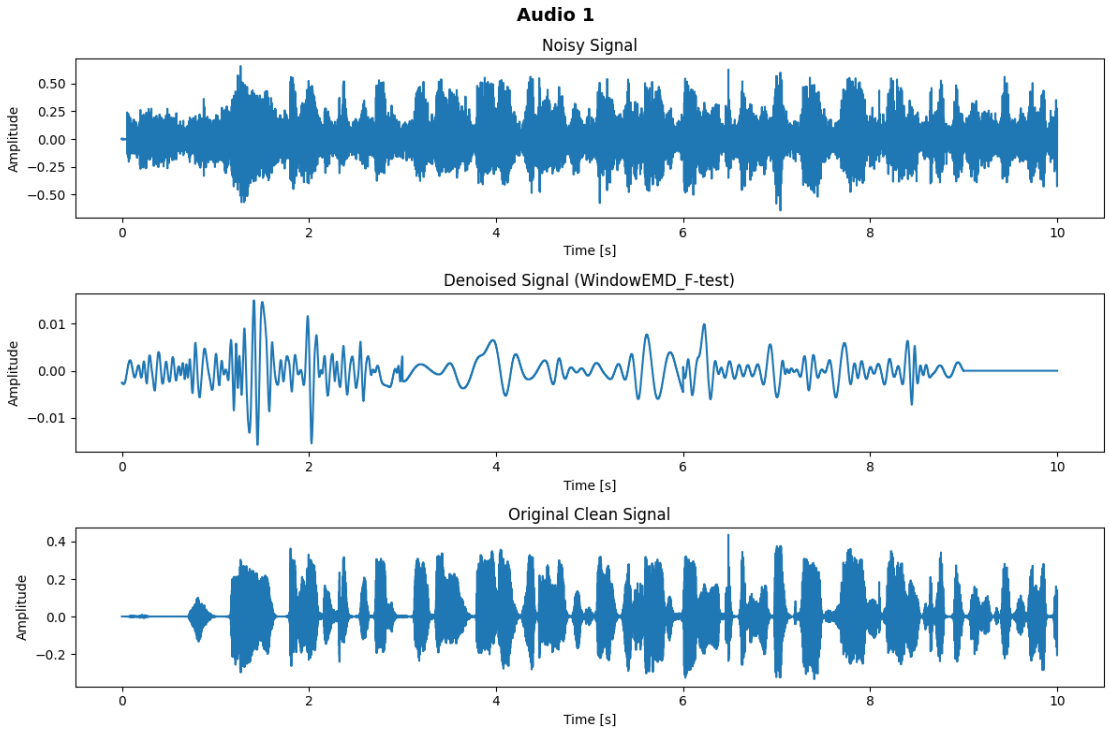

Рис. 3.12. Графіки аудіосигналів для Window EMD F-test.

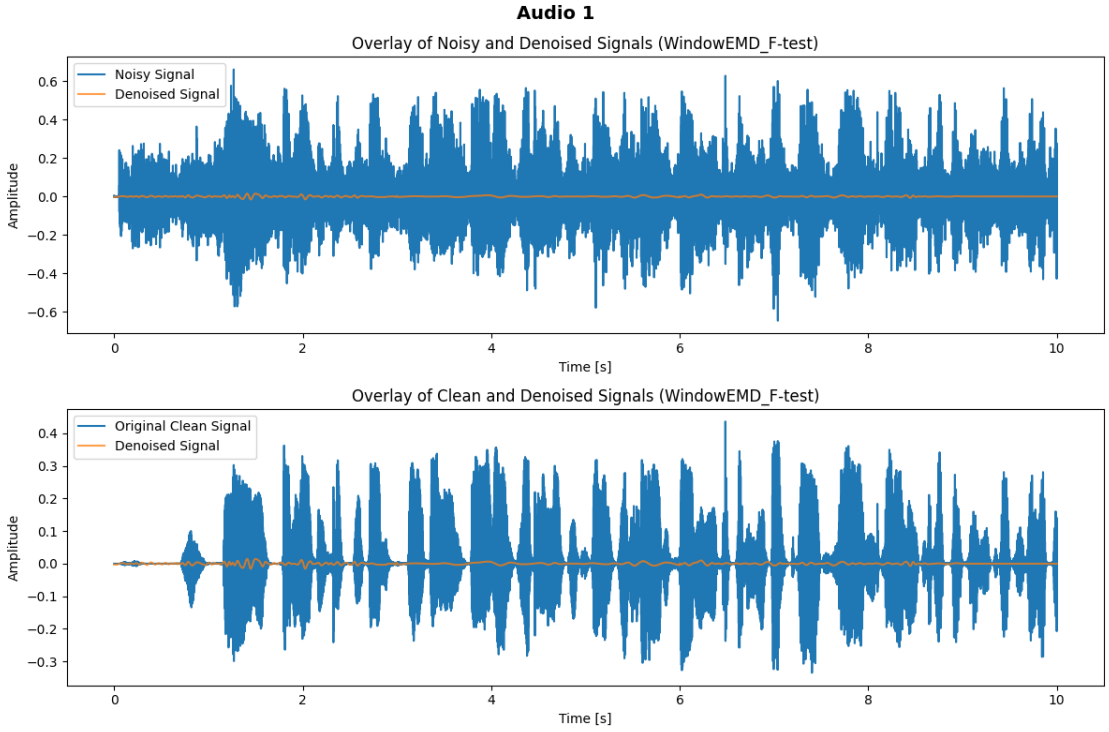

Рис. 3.13. Графіки порівняння для Window EMD F-test.

Також на результат в модифікованому методі впливає розмір вікна. На маленькому та середньому вікнах різниці між шумним та знешумленним сигналом майже не чутна, а на великому вікні втрачаються корисні

компоненти.

Отже поточний метод непогано справляється із задачею по видаленню шуму та стабілізації сигналу на графіках функцій, у той час як для повсякденних сигналів із записом голосу та інших корисних компонентів потребує подальшої модифікації.

3.5 Спектральне знешумлювання на основі перетворення Гільберта-Хуанга в поєднанні з енергетичним фільтром

Переглянувши роботу спектрального знешумлювання на основі перетворення Гільберта-Хуанга в поєднанні з F-тестом виникла ідея спробувати модифікувати F-test. Замість використання миттєвої частоти як критерію відмінності мод, можна спробувати оцінити потужність або енергію кожної моди, а потім застосовувати F-тест для порівняння дисперсій енергій між модами, який дозволив би відсіяти шумові компоненти та дістати корисні звуки. Для цього спробуємо дістати енергію кожної IMF як суму квадратів амплітуд (або миттєвих значень), що відображає потужність сигналу в кожній моді, і на їх основі обчислюємо F-статистику. F-статистика тепер заснована на співвідношенні енергій між двома сусідніми модами. Це порівняння допомагає виявити різкі стрибки або падіння енергії, які можуть вказувати на шум. Далі, як і раніше, F-тест проводиться для порівняння енергій, і на основі порогового значення відбувається відсіювання шумних мод.

Перевіримо модифікований метод на звуковик образах.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Window EMD Energy F-test	-0.24 dB	11204.14	0.73	12719.33	-3.47	5218.19	0.67	11984.08

Табл. 3.10. Результати видалення шуму методом Window EMD Energy F-test.

Отримані результати стали набагато кращими. За експертною оцінкою знешумленні звукові образи намагаються видалити шумові компоненти краще. Для прикладу розмови на фоні в офісі трохи були приглушені. Також добре видалили шум від гвинта гелікоптера. Але високочастотні шуми, як дзвінок телефону все ще залишився. Час роботи методів також став меншим. Слід зазначити, що в цьому методі короткі вікна показують кращі результати, ніж великі. Було підібрано невелике вікно, яке давало гарні результати. Поглянемо на графіки першого сигналу.

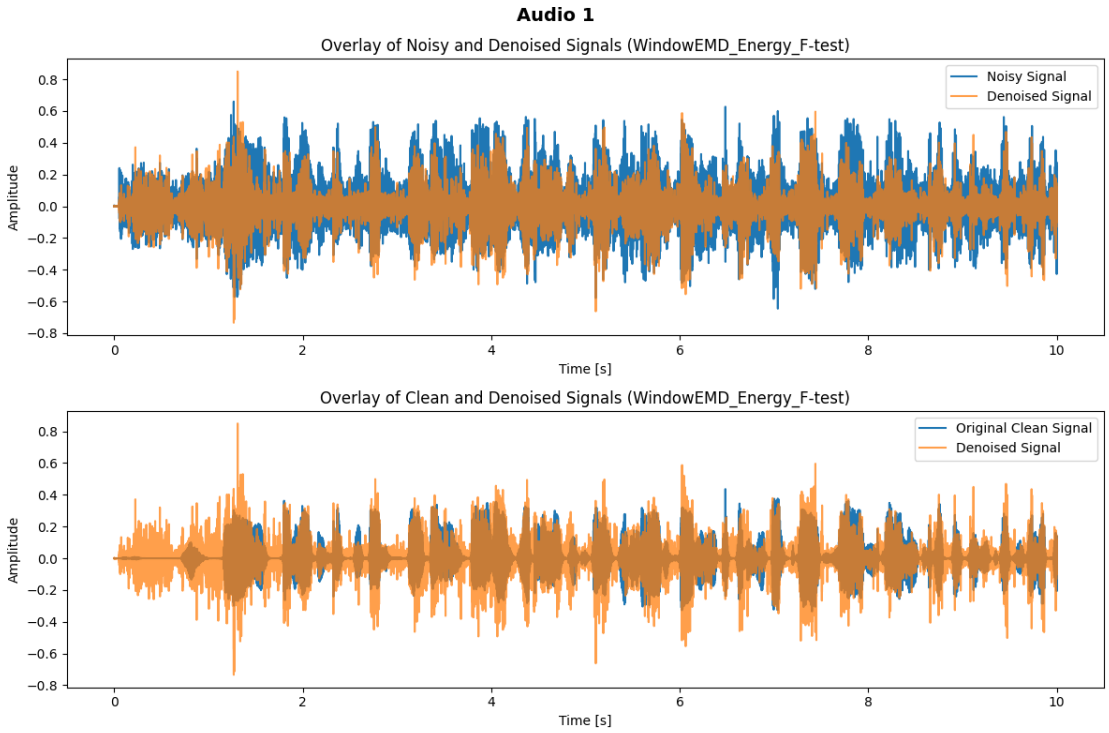

Рис. 3.14. Графіки порівняння для Window EMD Energy F-test.

3.6 Перетворення Гільберта-Хуанга з розпізнаванням особливостей сигналу

Попередній метод хоча і не показує результатів по SNR вище за 20 дБ, але дає розуміння того, що лише одна характеристика сигналу не дасть задовільний результат. Деякі шумові компоненти були видалені, але добра частина все ще залишається у вихідному аудіосигналі.

Після аналізу попередніх алгоритмів та отриманих результатів можна зробити висновок, що для детального розпізнавання шуму із подальшим

видаленням і збереженням високої якості голосу необхідно користуватися не одним методом фільтрації, а спробувати дістати декілька особливостей та класифікувати шумові компоненти на їх основі. Для цього методу скористаємося ідеями, запозиченими із роботи «Hilbert Spectrum Based Features for Speech/Music Classification» [5]. Розпізнавати голос чи шум будемо за допомогою K-Nearest Neighbors (KNN). Основна ідея KNN полягає в тому, щоб робити передбачення на основі найближчих сусідів, тобто найбільш схожих зразків у масиві отриманих особливостей. Згідно статті обмеження кількості IMFs до 10 буде достатнім для отримання потрібних особливостей та зменшуючи час обробки.

План роботи алгоритму наступний:

1) Попередня підготовка даних.
   Завантажуємо сигнал, задаємо частоту дискретизації у 8000 гЦ та розбиваємо його на короткі сегменти з фіксованою тривалістю.
2) Розкладання сигналу.
   Кожен сегмент сигналу розкладається на Intrinsic Mode Functions (IMFs) за допомогою емпіричного модового розкладання (EMD). Ці компоненти є основою для подальшого аналізу.
3) Аналіз компонентів за допомогою HHT.
   До кожної компоненти (IMF) застосовується трансформація Гільберта для оцінки миттєвих характеристик, таких як частота та амплітуда. Це дозволяє отримати більш точну інформацію про локальні коливання сигналу.
4) Обчислення характеристик компонентів.
   Виконується розрахунок енергії, спектральної ентропії, полосового фільтра (від 200 гЦ до 3400 гЦ, де зазвичай знаходиться голос людини) для кожної компоненти. Ці характеристики допомагають розрізнити корисні компоненти сигналу від шуму.
5) Розпізнавання корисних компонентів.
   Шумові компоненти визначаються за допомогою кластеризації (наприклад, K-means) або аналізу характеристик. Компоненти з високою нерегулярністю або низькою енергією відкидаються.
6) Фільтрація та очищення компонентів.
   До вибраних компонентів застосовуються додаткові методи очищення,

такі як смугові фільтри або вейвлетне перетворення, для зменшення залишкових шумів.

7) Реконструкція сигналу.
   Очищені компоненти об'єднуються у цілісний сигнал. Застосовується перекриття сегментів для забезпечення плавності відновленого сигналу.
8) Оцінка результатів.
   Результат оцінюється через аналіз співвідношення сигнал/шум (SNR), порівнюючи очищений сигнал з еталонним.

Запустимо новий алгоритм на звукових образах.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Window EMD KNN 3 features	0.36 dB	7372.29	6.40	8322.75	-5.28	3333.09	4.53	8641.90

Табл. 3.11. Результати видалення шуму методом Window EMD та HHT з 3-ма особливостями.

Згідно SNR не можна сказати, що результати стали кращими. Але прослуховуючи аудіофайли можна зазначити, що прогрес присутній. Після застосування алгоритму голоси переднього плану зберігають свої акустичні характеристики та розбірливість, що свідчить про зменшення впливу приглушення корисного сигналу під час обробки.

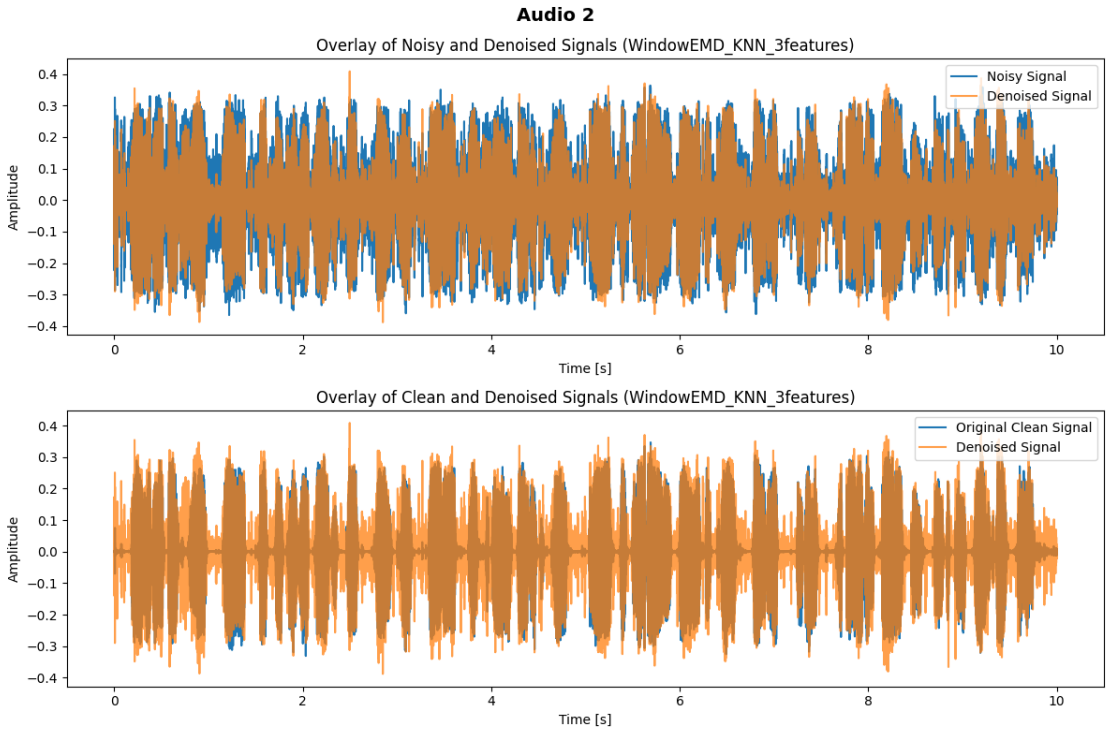

Рис. 3.15. Графіки порівняння для Window EMD та HHT з 3-ма особливостями.

3.7 EMD та HHT з розпізнаванням особливостей сигналу та ручною фільтрацією

У попередньому методі були відсутні фільтри для особливостей, що в свою чергу передавало до кластеризації усі отримані IMFs. Тому для подальшої модифікації були додані емпірично визначені порогові значення, які б відсіювали очевидні шумові компоненти та передавали б кращі особливості сигналу для розпізнавання. В теорії такий підхід повинен краще знаходити шумові компоненти серед переднього плану звукових образів. Також було вирішено в пошук особливостей сигналу додати Мел-частотні кепстральні коефіцієнти (MFCC, Mel-frequency Cepstral Coefficients).

Перевіримо, як фільтрація вплине на результати.

Метод	Аудіо 1		Аудіо 2		Аудіо 3		Аудіо 4	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Window EMD KNN 3 features	-2.39 dB	4171.90	-4.07	4350.45	-7.19	1365.36	1.49	4875.07

Табл. 3.12. Результати розпізнаванням шуму методом Window EMD та HHT з фільтрацією.

Отримані результати не прийнятні як по SNR, так і за експертною оцінкою. Вихідні сигнали були спотворені збільшеною гучністю шуму та різкими високочастотними звуковими піками. Хоча для 4-го вихідного звукового образу дійсно не чутно клацання клавіатури, але якість цього аудіо зовсім не прийнятна. Використання ручного налаштування фільтрів для розпізнавання особливостей сигналу є малоефективним через декілька причин. По-перше, особливості сигналу є високо-нестабільними та залежать від його природи, що робить створення універсальних фільтрів майже неможливим. По-друге, у випадку розкладання сигналу на Intrinsic Mode Functions (IMFs), кожна компонента має свої унікальні частотні та амплітудні характеристики, які можуть суттєво відрізнятися навіть у межах одного сигналу. Це означає, що для кожної IMF потрібно окремо налаштовувати фільтр, що значно ускладнює процес обробки. Більше того, таке ручне налаштування фільтрів призводить до втрати автоматизації процесу та підвищує ризик суб'єктивності під час вибору параметрів. У підсумку, ручні фільтри не лише не адаптивні, але й не забезпечують стабільності та універсальності в обробці різних сигналів і їх компонент. Для покращення відсіювання шумових компонент краще розробити автоматизовані фільтри, які зможуть підлаштовуватися під характеристику сигналу.

3.8 Модифікований метод віконного EMD

Для покращення роботи віконного методу перетворення Гільберта-Хуанга (HHT) в поєднанні з Empirical Mode Decomposition (EMD) та розпізнаванням особливостей сигналу було зроблено адаптивний вибір особливостей. Та-

кож згідно вищезазначеної статті [5] додамо вилучення нових особливостей на основі Гільбертового спектрального аналізу IMF, а саме: перетворення Гільберта-Хуанга миттєвої амплітуди Мел-частотних коефіцієнтів (HTIA-MFCC) та перетворення Гільберта-Хуанга миттєвої частоти Мел-частотних коефіцієнтів (HTIF-MFCC), щоб скористатися перевагами обох наборів фільтрів HT і Мел на основі EMD. План модифікованого алгоритму виглядає наступним чином:

1) Підготовка аудіосигналу.
   Завантажуємо сигнал, задаємо частоту дискретизації у 8000 гЦ. Сигнал розбивається на вікна фіксованого розміру з певним ступенем перекриття. Розмір вікна та ступінь перекриття обираються для забезпечення балансу між часовою роздільною здатністю та якістю відновлення сигналу. Перекриття вікон дозволяє уникнути артефактів на межах та забезпечує плавне відновлення сигналу без розривів.
2) Застосування віконної функції.
   До кожного вікна застосовується віконна функція (наприклад, вікно Хеммінга) для зменшення ефектів на краях вікна.
3) Емпіричне модальне розкладання (EMD).
   Для кожного вікна виконується EMD, що дозволяє розкласти сигнал на набір внутрішніх модальних функцій (IMFs), впорядкованих за частотним вмістом.
4) Розрахунок ознак для кожного IMF.
   Для кожного IMF обчислюються статистичні та енергетичні ознаки.
   Енергетичний критерій: обчислюється відношення енергії в діапазоні мовлення (наприклад, 300–3400 Гц) до загальної енергії IMF.
   Статистичні ознаки: ексцес (kurtosis), асиметрія (skewness), ентропія.
5) Адаптивний вибір значущих IMF.
   Далі нормалізуються обчислені ознаки для забезпечення їх співвідносності та розраховується комбінований показник для кожного IMF на основі зваженого сумування нормалізованих ознак. Після встановлюється поріг на основі процентиля комбінованих показників (в даному методі 89-й процентиль). Значення процентиля є ключовим параметром у методі, що впливає на баланс між збереженням мовного сигналу та видаленням шуму. Наприкінці відбираються IMF,

комбінований показник яких перевищує встановлений поріг.

6) Виділення ознак.
   З отриманих IMFs після фільтрації витягуємо наступні ознаки:
   - *HTIA-MFCC*: Мел-частотні кепстральні коефіцієнти, обчислені з огинаючої аналітичного сигналу IMF.
   - *HTIF-MFCC*: Мел-частотні кепстральні коефіцієнти, обчислені з миттєвої частоти IMF.
   - *Спектральні ознаки*:
     - Спектральний центроїд.
     - Ширина спектральної смуги.
     - Частота нульових перетинів.
   - *MFSE*: Мел-частотна спектральна ентропія.
   - *Вейвлетні ознаки*:
     - Енергії коефіцієнтів вейвлетного розкладання на різних рівнях.
7) Кластеризація IMF.
   Перед кластеризацією отримані ознаки нормалізуються (наприклад, StandardScaler) і застосовується метод зниження розмірності (наприклад, Kernel PCA) для покращення якості кластеризації. Після виконується неконтрольована кластеризація (наприклад, методом k-середніх із $n\_components$=2 (розділення голосу та шуму)) для групування IMF на основі їх ознак.
8) Визначення мовного кластеру.
   Кластер, що відповідає мовному сигналу, визначається на основі енергетичного критерію. Припускається, що кластер з найбільшою середньою енергією відповідає мовленню.
9) Реконструкція очищеного сигналу.
   Класифіковані IMF, що належать до мовного кластеру, сумуються для отримання очищеного вікна. Далі ці вікна об'єднуються з урахуванням перекриття.
10) Постобробка сигналу.
    Амплітуда очищеного сигналу коригується для відповідності піковому значенню вхідного сигналу для відображення накладень на графіках.

Слід зазначити, що модифіковані методи в більшості адаптовані під аудіосигнали різної довжини та складності шумових компонентів. Але для отримання кращого результату такі параметри як поріг процентиля, розмір та крок вікна можуть бути змінені для кращого розпізнавання компонентів сигналу із наступним знешумлюванням різноманітних звукових образів.

Запустимо новий алгоритм на звукових образах.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Modified Window EMD & HHT	0.67 dB	7862.28	1.33	8586.03	-0.68	3333.09	3.49	13758.68

Табл. 3.13. Результати розпізнавання та видалення шуму методом Modified Window EMD & HHT.

Результати SNR все ще залишаються не великими. Натомість модифікований метод за експертною оцінкою непогано видалив розмови на другому плані першого аудіозапису. Але все ще залишається проблема із швидкими високочастотними шумовими компонентами. Особливо це чутно, коли голос переднього плану починає говорити. Вочевидь це пов'язано із тим, що метод не може до кінця виділити шумові компоненти посеред голосу. Поглянемо на графіки розпізнавання та видалення шуму першого та другого звукових образів.

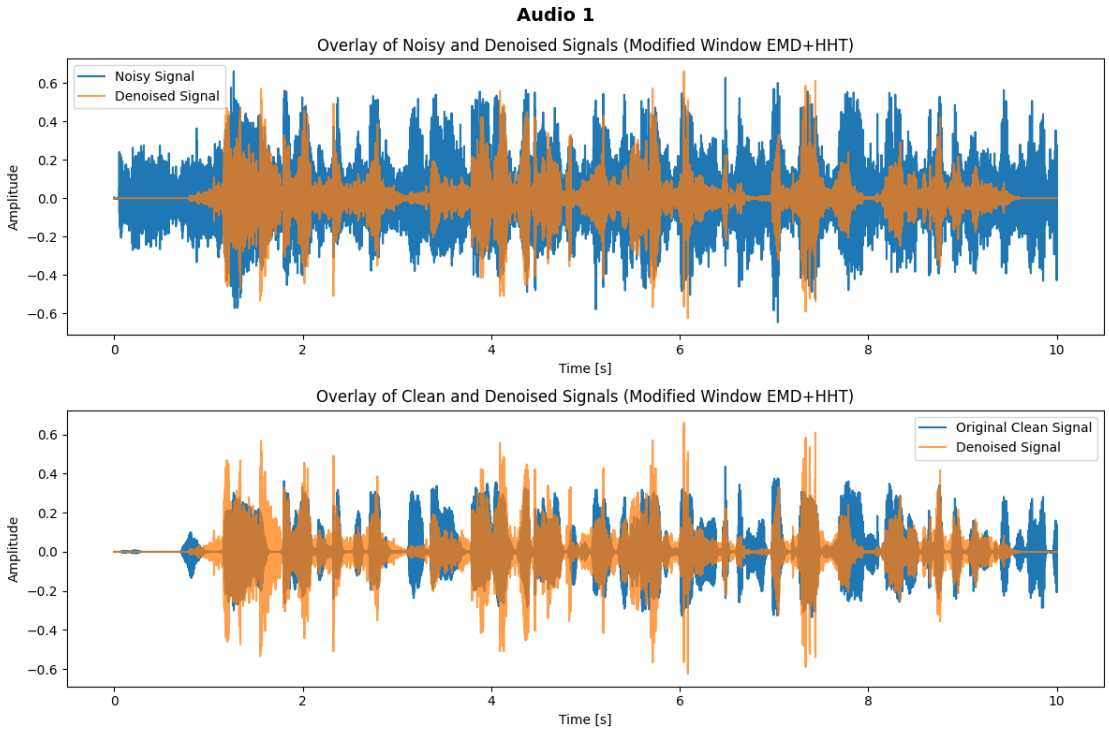

Рис. 3.16. Графіки порівняння 1-го аудіо для Modified Window EMD & HHT.

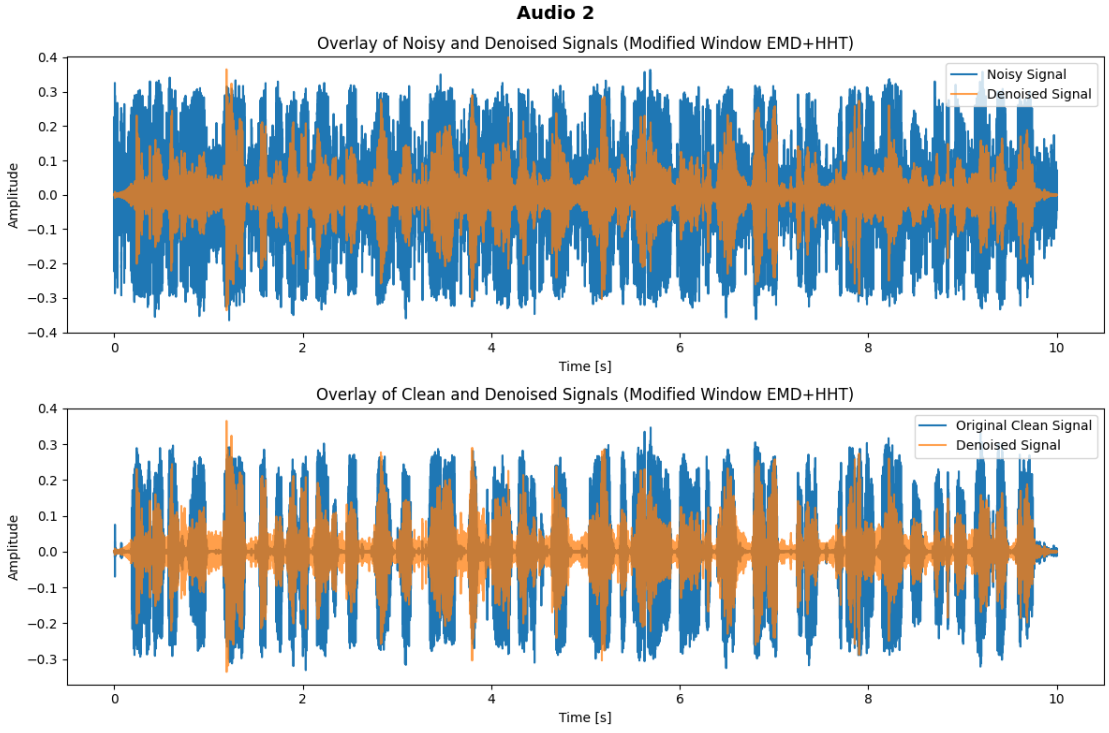

Рис. 3.17. Графіки порівняння 2-го аудіо для Modified Window EMD & HHT.

3.9 Модифікований метод віконного EMD із Spectral Gating

Попередній модифікований метод доволі непогано розпізнавав та видаляв складні фонові компоненти, такі як голоси людей другого плану. В класичних методах Spectral Gating доволі непогано намагався видалити шумові компоненти, але все ж таки залишалося багато зайвого. Спробуємо об'єднати модифікований метод Window EMD & HHT та застосувати Spectral Gating на кожній отриманій IMF для кращого розпізнавання та знешумлювання.

План моделі залишається таким самим, лише після пункту 3) Емпіричне модальне розкладання (EMD) будемо застосовувати спектральний гейтінг до кожного IMF для виявлення та видалення шуму. І далі, як у попередньому методі, розрахунок ознак для кожного знешумленого IMF (пункт 4)).

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Modified Window EMD & HHT + Spectral Gating	1.40 dB	12829.85	1.34	9112.78	0.84	3798.61	0.90	14036.23

Табл. 3.14. Результати розпізнавання та видалення шуму методом Modified Window EMD & HHT з фільтрацією Spectral Gating.

Результати SNR все ще невеликі, проте всі додатні. Також за експертною оцінкою метод значно краще почав видаляти шумові компоненти різної складності. Хоча і є свої мінуси - через використання Spectral Gating гучність сигналів стала меншою. Поглянемо на візуалізацію результатів виявлення та знешумлювання для усіх 4-х звукових образів.

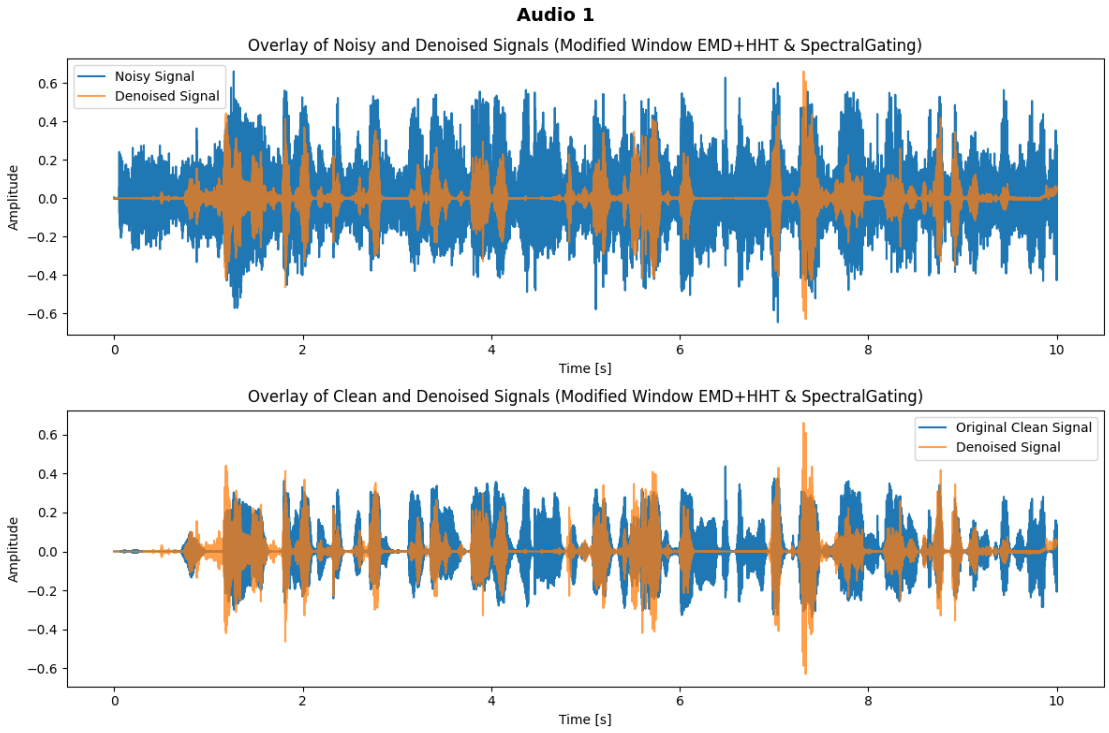

Рис. 3.18. Графіки порівняння 1-го аудіо для Modified Window EMD & HHT з фільтрацією Spectral Gating.

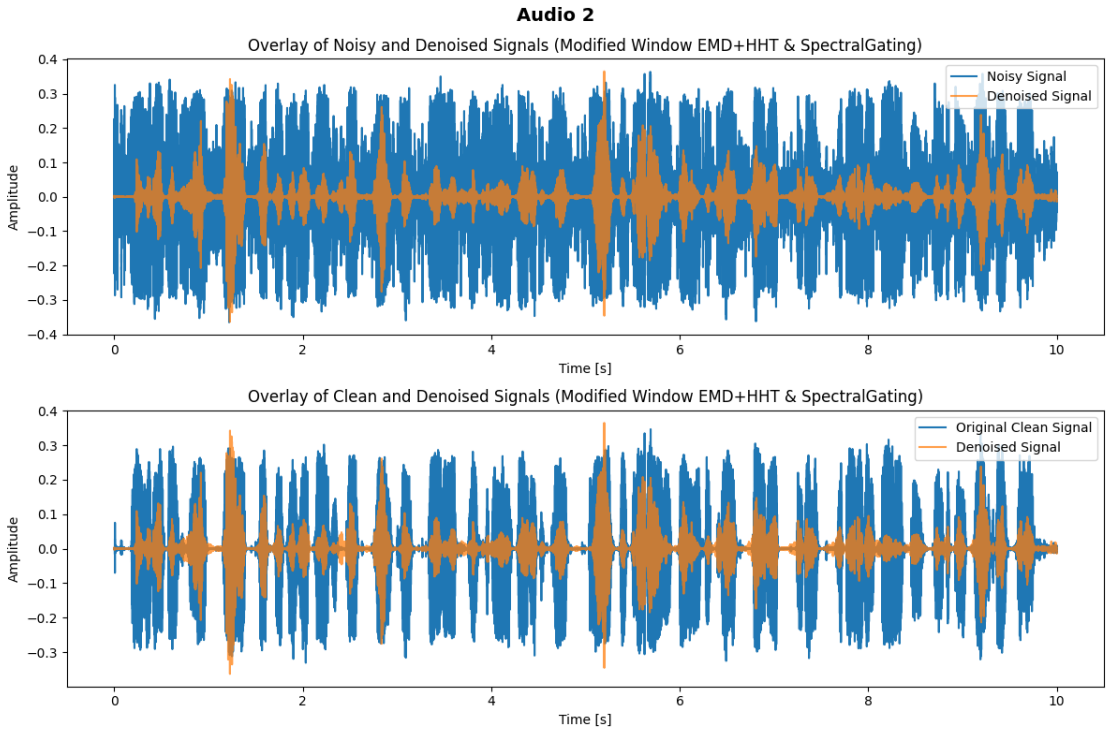

Рис. 3.19. Графіки порівняння 2-го аудіо для Modified Window EMD & HHT з фільтрацією Spectral Gating.

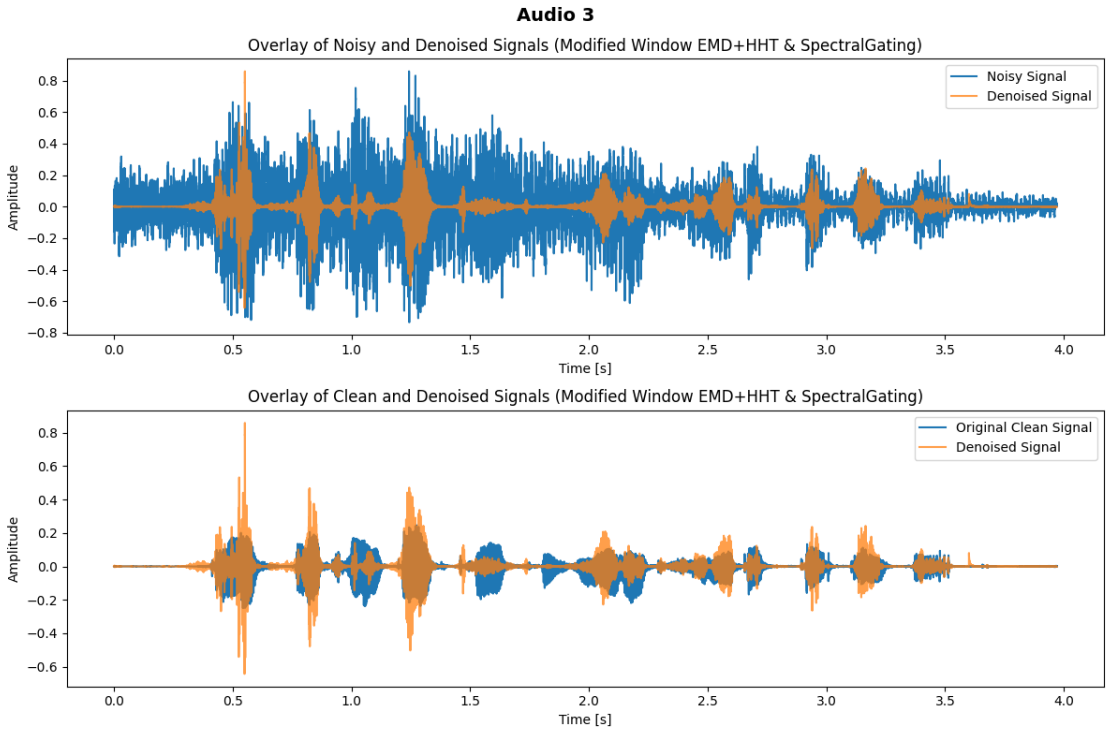

Рис. 3.20. Графіки порівняння 3-го аудіо для Modified Window EMD & HHT з фільтрацією Spectral Gating.

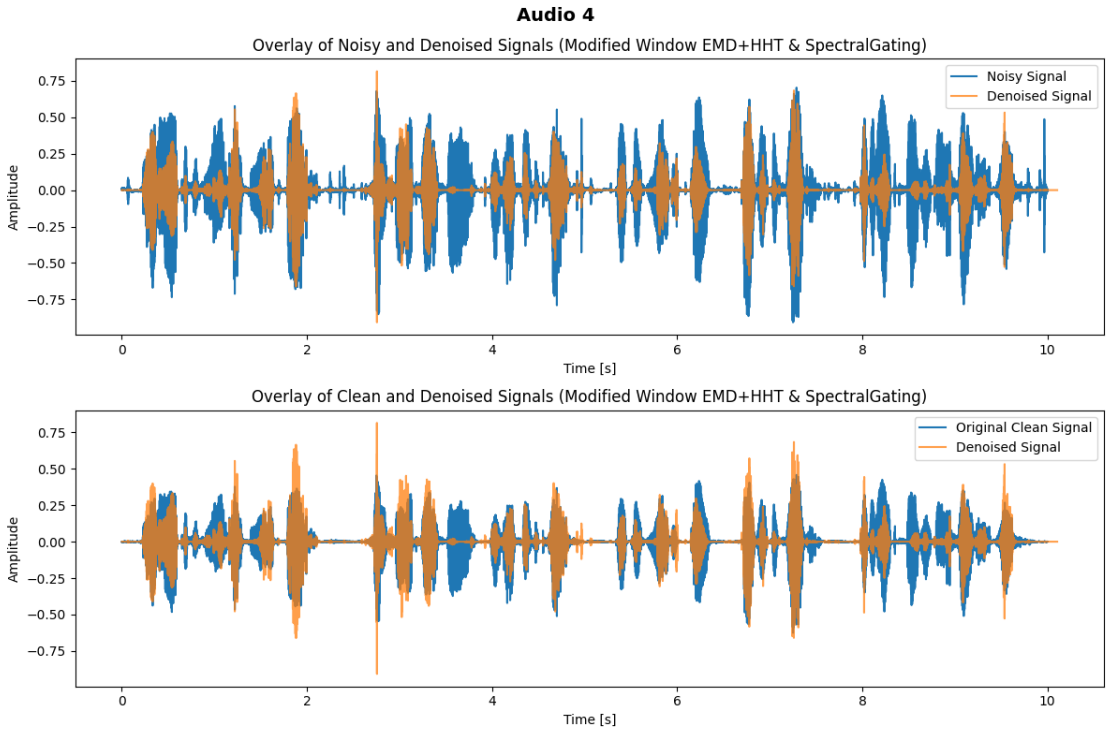

Рис. 3.21. Графіки порівняння 4-го аудіо для Modified Window EMD & HHT з фільтрацією Spectral Gating.

3.10 Розпізнавання та знешумлювання за допомогою TensorFlow(CNN)

Для розуміння та аналізу отриманих результатів була побудована нейронна мережа на основі TensorFlow(CNN), як описано у статті «A Deep Dive into Audio Denoising with TensorFlow(CNN)»[13]. TensorFlow — це відкритий програмний фреймворк для створення, тренування та впровадження моделей машинного навчання. Одним із популярних його застосувань є побудова згорткових нейронних мереж (Convolutional Neural Networks, CNN), які широко застосовуються в задачах аналізу зображень та сигналів завдяки можливості автоматично виявляти та виділяти важливі характеристики вхідних даних, такі як патерни чи структури.

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
TensorFlow(CNN)	3.82	-	7.42	-	2.22	-	7.95	-

Табл. 3.15. Результати розпізнавання та видалення шуму методом TensorFlow(CNN).

Отримані SNR хоча і не перевищують значення у 20 дБ, але все одно мають кращий результат за всі попередні методи. Натомість при прослуховуванні можна почути, що така проста модель може пропускати шумові компоненти, а іноді розмазувати голос на передньому плані. Для прикладу голос на аудіозаписі 3 зовсім не вдається розібрати, а довжина аудіо була некоректно записана через обмеженість моделі. Час роботи методу на основі нейронних мереж вже рахується не в мілісекундах чи секундах, а в годинах. Навчання моделі зайняло приблизно 2 години на невеликій кількості навчальних даних. Також був знайдено ще один недолік при навчанні нейронних мереж - це різниці внутрішніх методів при різних версіях. Нещодавно TensorFlow(CNN) оновився і перестав підтримувати старі файли з ваговими коефіцієнтами для використання найкращої навченої моделі. Графіки знешумлювання методом TensorFlow(CNN) виглядають наступним чином:

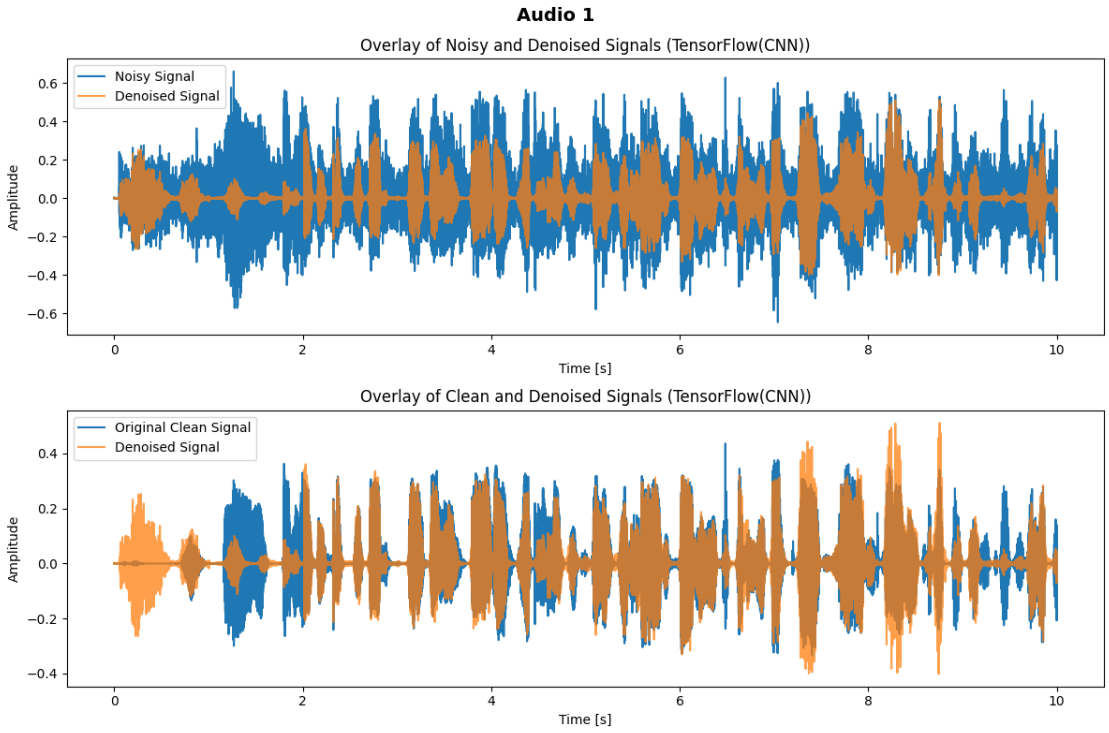

Рис. 3.22. Графіки порівняння 1-го аудіо для TensorFlow(CNN).

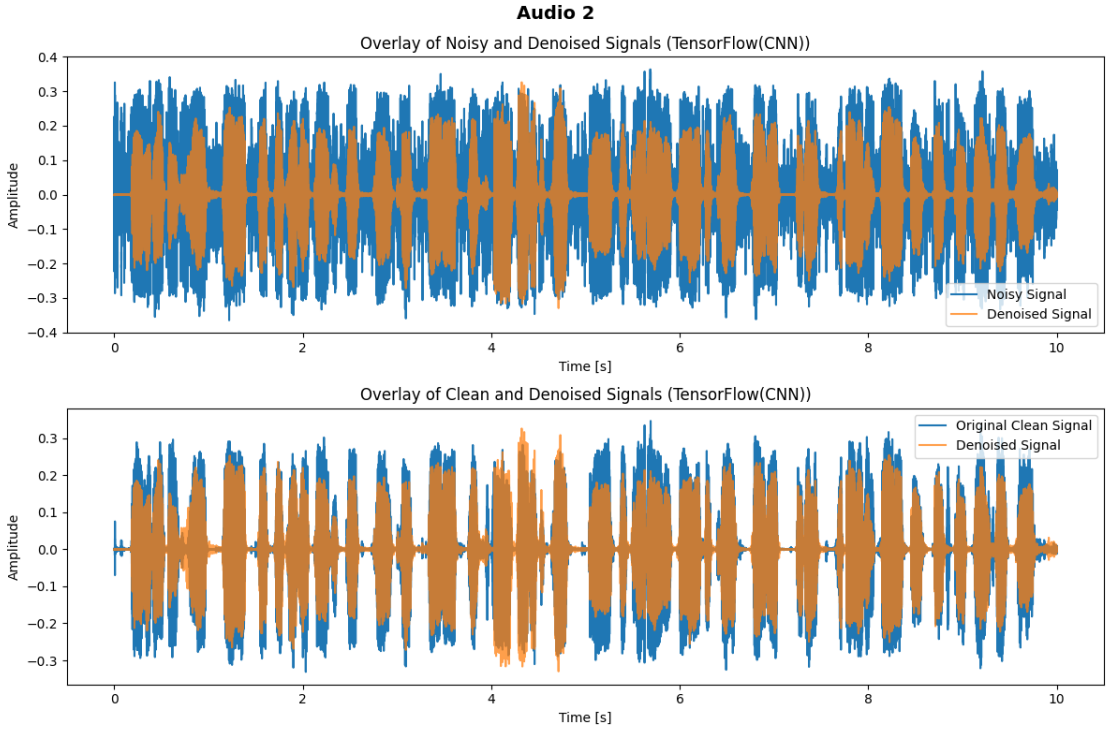

Рис. 3.23. Графіки порівняння 2-го аудіо для TensorFlow(CNN).

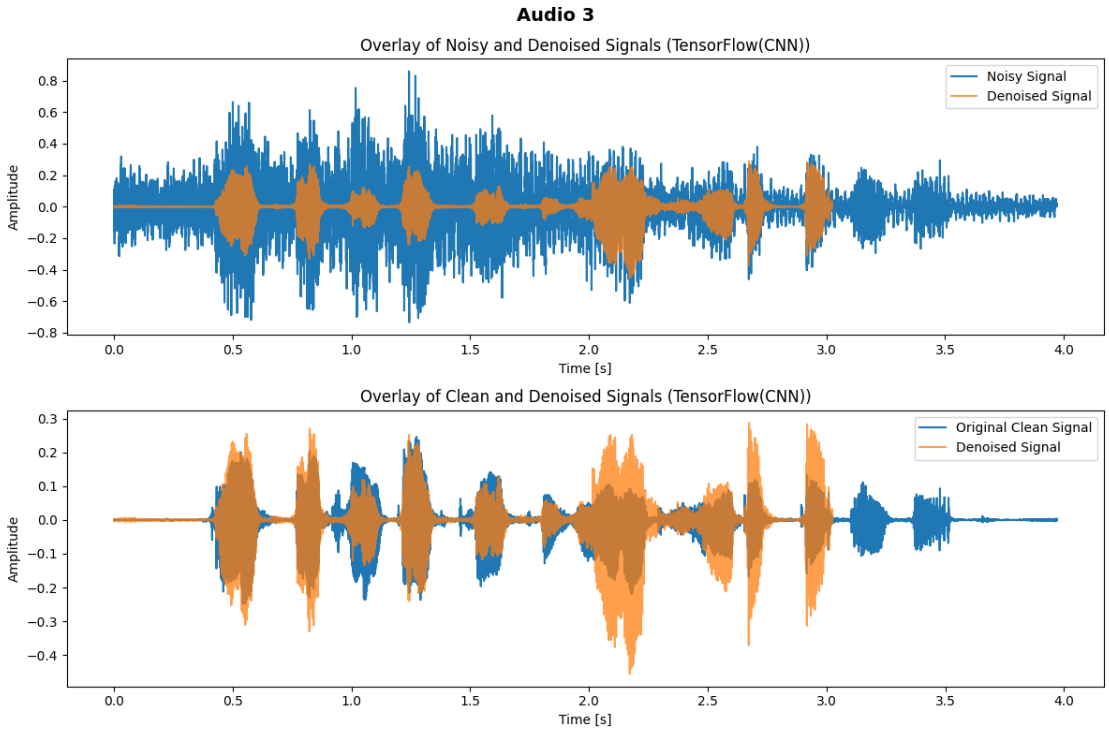

Рис. 3.24. Графіки порівняння 3-го аудіо для TensorFlow(CNN).

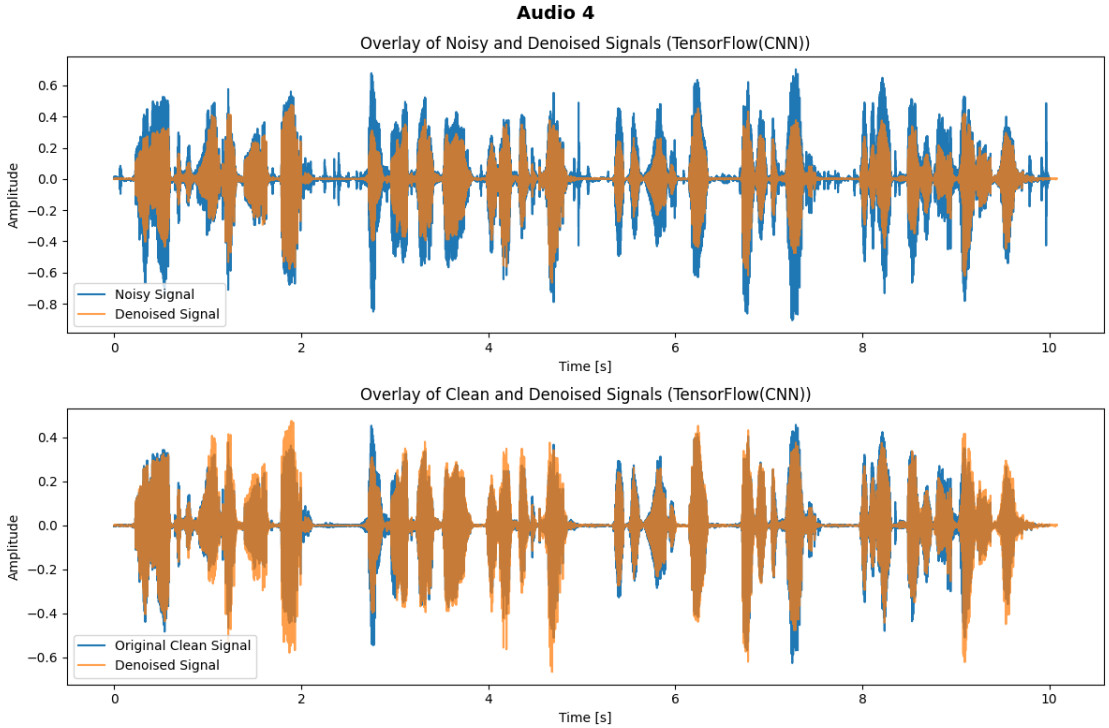

Рис. 3.25. Графіки порівняння 4-го аудіо для TensorFlow(CNN).

3.11 Таблиця отриманих результатів

Метод	**Аудіо 1**		**Аудіо 2**		**Аудіо 3**		**Аудіо 4**	
	SNR (дБ)	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**	**SNR (дБ)**	**Час (мс)**
	10 с		10 с		3 с		10 с	
Window EMD F-test	-0.01	49931.53	-0.06	25108.25	-6.43	157517.94	-0.01	43515.68
Window EMD Energy F-test	-0.24 dB	11204.14	0.73	12719.33	-3.47	5218.19	0.67	11984.08
Window EMD KNN 3 features	0.36 dB	7372.29	6.40	8322.75	-5.28	3333.09	4.53	8641.90
Window EMD KNN 3 features	-2.39 dB	4171.90	-4.07	4350.45	-7.19	1365.36	1.49	4875.07
Modified Window EMD & HHT	0.67 dB	7862.28	1.33	8586.03	-0.68	3333.09	3.49	13758.68
Modified Window EMD & HHT + Spectral Gating	1.40 dB	12829.85	1.34	9112.78	0.84	3798.61	0.90	14036.23
TensorFlow(CNN)	3.82	-	7.42	-	2.22	-	7.95	-

Табл. 3.16. Результати розпізнавання та видалення шумових компонент модифікованими методами.

РОЗДІЛ 4

АНАЛІЗ ТА ПІДТВЕРДЖЕННЯ ЕФЕКТИВНОСТІ МОДИФІКОВАНИХ МЕТОДІВ РОЗПІЗНАВАННЯ ШУМОВИХ КОМПОНЕНТІВ І ЇХ ПОДАЛЬШОГО ВИДАЛЕННЯ

З огляду на отримані результати можна зазначити, що модифікований метод віконного EMD (як звичайний, так і з Spectral Gating) дає досить непоганий результат для видалення шумових компонентів із аудіосигналів при збереженні голосу переднього плану. Модифіковані методи показали гарне знешумлення при видаленні сторонніх голосів на другому плані, видаленню шуму від несучого гвинта гелікоптера та, навіть, тривалого звуку реактивних двигунів літака, який майже перекривав голос переднього плану.

Згідно результатів SNR, експертної оцінки та побудованих графіків серед розглянутих класичних методів найкращий результат дає Spectral Gating. Із часо-частотних методів, включно з їх модифікацією, чудовий результат надає метод віконного EMD із Spectral Gating, який видалив остаточні шумові компоненти, які залишилися після класичного методу. Отримані графіки вже точніше збігаються з оригінальними, чистими звуковими образами, видаляючи та зменшуючи шум там, де потрібно. Проте метод також має свої недоліки. А саме - зменшення енергії корисного сигналу після обробки. Голоси переднього плану стають тихіше та трохи втрачають яскравість, порівняно із оригіналом.

Використання таких методів як спектральне знешумлювання на основі перетворення Гільберта-Хуанга в поєднанні з F-тестом та його модифікація із енергією не мають сенсу. Для поточної задачі із знешумлюванням аудіосигналів вони справляються погано. А класичні методи можуть непогано видаляти шум лише на простих та не тривалих сигналах. В застосунках для обробки аудіо існують плагіни на основі цих методів, які можуть показати задовільний результат на музичних записах у короткому вікні для, наприклад, видалення тихого та нетривалого компоненту шуму десь на фоні.

В такому випадку доцільно використати простий плагін, ніж проганяти складний метод зі своїми мінусами. Але для побутових звуків потрібні складніші методи, які будуть поєднувати як класичні, так і часо-частотні переваги.

Разом із цим, важливо врахувати вплив багатокомпонентного шуму, який містить як постійні, так і періодичні елементи. Наприклад, у випадку з аудіозаписом 3 тривалий шум двигунів літака суттєво перекривав голос людини, що ускладнювало виділення мовного сигналу. Це свідчить про необхідність розробки адаптивних методів, здатних працювати зі стабільними низькочастотними компонентами шуму, які важко відокремити стандартними алгоритмами.

На противагу цьому, шум несучого гвинта гелікоптера (аудіо 2) демонструє значно кращі результати при обробці. Завдяки регулярності та чіткій частотній структурі такого шуму алгоритми успішно відокремлюють його від мовного сигналу. Однак короткі імпульсні шуми, як-от клацання клавіатури (аудіо 4), залишаються складними для обробки. Їхня нерегулярність, низька інтенсивність і коротка тривалість у часі призводять до того, що вони часто втрачаються серед домінуючого мовного сигналу.

Ще одним важливим викликом є розпізнавання голосів переднього плану від фонових. Як показує приклад із аудіозаписом 1, методи стикаються з труднощами у відокремленні голосу-цілі від фонових голосів, особливо коли обидва мають схожу гучність і частотний спектр. Для вирішення цієї задачі слід спробувати використати методи, що враховують просторове розташування джерел звуку (зокрема, їхнє позиціонування в акустичному середовищі), а також аналізують часово-динамічні характеристики мовлення.

Щодо використання нейронних мереж - вони мають свої виклики. Їх основна проблема — це обмеженість даних, особливо для специфічних звуків, які важко знайти у відкритому доступі. Навчання таких моделей потребує значних обчислювальних ресурсів і часу, що вимагає потужного обладнання. Однак вже навчені моделі зазвичай швидко обробляють нові аудіосигнали і показують хороші результати. Водночас існує ризик, що в окремих, нестандартних (граничних) випадках модель може дати некоректний результат через недостатню кількість або різноманітність даних у навчальній вибірці.

ВИСНОВКИ

У даній роботі здійснено аналіз, дослідження та вдосконалення методів розпізнавання шумових компонентів і голосових сигналів із подальшим вилученням шуму. Особливу увагу приділено обробці аудіозаписів із різними типами шумових перешкод, включаючи тривалі фонові шуми, короткотривалі імпульсні звуки та багатокомпонентні шумові структури. Запропоновані методи апробовано на звукових даних із різноманітними джерелами та шумовими характеристиками, що дозволило оцінити їх ефективність у контексті різних акустичних сценаріїв.

Після аналізу отриманих результатів можна підсумувати, що класичні методи видалення шумових компонент не підходять для обробки побутових нелінійних і нестаціонарних звукових образів. Для вирішення цієї задачі були застосовані часо-частотні методи обробки сигналів, які дозволили більш точно виділяти характерні особливості сигналу, зокрема його спектральні та часові компоненти. Використання спектрального знешумлювання на основі перетворення Гільберта-Хуанга в поєднанні з F-test показали незадовільні результати для побутових сигналів. Натомість як для аналізу та знешумлювання простих графіків функцій, або, наприклад, отриманих сигналів з електрокардіографії - то метод працює чудово та відносно швидко, адже такі сигнали зазвичай нетривалі. Подальші тести показали, що вилучення особливостей сигналу для класифікації голосу та шуму з адаптивним фільтром IMFs набагато краще справляються з поточною задачею.

З огляду на аналіз результатів можна зазначити, що модифікований метод віконного Empirical Mode Decomposition із класифікацією особливостей сигналу та адаптивним вибором значущих ознак продемонстрував задовільну ефективність у знешумлюванні і збереженні голосових сигналів у різних акустичних середовищах. Додавання попереднього етапу Spectral Gating для кожної отриманої IMF покращило якість видалення шуму, проте призвело до зниження енергії сигналу. Це вказує на перспективність подальшої модифікації методу, зокрема оптимізації використання Spectral Gating, адаптації його параметрів або пошуку альтернативних підходів для мінімізації втрат енергії сигналу та покращення загальної якості обробки. Іншим можливим

направленням є спроба використати такі часо-частотні алгоритми, як емпіричне вейвлет-перетворення (Empirical wavelet transform, EWT)[1] замість EMD, та Teager Energy Operator (TEO)[14], або Teager-Huang Transform (THT)[15] замість перетворення Гільберта-Хуанга.

СПИСОК ЛІТЕРАТУРИ

1. Pachori, Ram Bilas (2023). Time-Frequency Analysis Techniques and their Applications (1st ed.), 238 p.
2. A. B. Gumelar, M. H. Purnomo, E. M. Yuniarno and I. Sugiarto, "Spectral Analysis of Familiar Human Voice Based On Hilbert-Huang Transform,"2018, Surabaya, Indonesia, 2018, pp. 311-316.
3. Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. The Journal of the Acoustical Society of America, 8(3), 185-190.
4. Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357-366.
5. Kumar, Arvind & Solanki, Sandeep & Chandra, Mahesh. (2022). Hilbert spectrum based features for speech/music classification. Serbian Journal of Electrical Engineering. 19. 239-259.
6. Jeon, Hohyub & Jung, Yongchul & Lee, Seongjoo & Jung, Yunho. (2020). Area-Efficient Short-Time Fourier Transform Processor for Time–Frequency Analysis of Non-Stationary Signals. Applied Sciences. URL: https://www.researchgate.net/publication/346243843_Area-Efficient_Short-Time_Fourier_Transform_Processor_for_Time-Frequency_Analysis_of_Non-Stationary_Signals (дата звернення: 06.11.2024).
7. Sejdic, Ervin & Djurovic, Igor & Jiang, Jin. (2009). Time–frequency feature representation using energy concentration: An overview of recent advances. Digital Signal Processing. 2019. URL: https://www.researchgate.net/publication/223900360_Time-frequency_feature_representation_using_energy_concentration_An_overview_of_recent_advances (дата звернення: 08.11.2024).
8. Debnath, Lokenath & Antoine, Jean-Pierre. Wavelet Transforms and Their Applications. Physics Today - PHYS TODAY, 2003. URL: https://www.researchgate.net/publication/238958371_Wavelet_

Transforms_and_Their_Applications (дата звернення: 21.10.2024).

9. Zosso, Dominique & Dragomiretskiy, Konstantin. (2013). Variational Mode Decomposition. IEEE Transactions on Signal Processing. 62.
10. Bian X, Ling M, Chu Y, Liu P, Tan X. Spectral denoising based on Hilbert-Huang transform combined with F-test. Front Chem. 2022 Aug.
11. B. H. Prasetio, D. O. Yusuf, D. Syauqy and S. Chilmi, "Spectral Gating for Noise Reduction in Speech Stress Recognition System,"2024 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), BALI, Indonesia, 2024, pp. 149-155.
12. Richard W. Hamming, Error detecting and error correcting codes, 1950. Bell System Technical Journal 29 (2): pp. 147–160.
13. https://medium.com/@vaibhavtalekar87/a-deep-dive-into-audio-denoising-with-tensorflow-cnn-a996e0c62e16 (дата звернення: 19.11.2024)
14. P. Maragos, J.F. Kaiser and T. Quatieri, ”Energy separation in signal modulation to speech analysis,” IEEE TSP., vol. 41, pp. 3024-3051, 1993.
15. J. -C. Cexus and A. -O. Boudraa, "Nonstationary signals analysis by Teager-Huang Transform (THT),"2006 14th European Signal Processing Conference, Florence, Italy, 2006, pp. 1-5.