

КІБЕРНЕТИКА та КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

У статті представлено метод чірплет-перетворення для аналізу звукових сигналів, що поєднує елементи перетворення Гільберта – Хуанга та чірплет-аналізу для підвищення точності сегментації та виділення характеристик аудіосигналів. Метод дозволяє локалізувати часово-частотні особливості сигналу, адаптуючись до складних динамічних змін. Використання чірплетної проєкції на альтернативні нелінійні шкали забезпечує більш точне виявлення мовних сегментів, що сприяє вдосконаленню методів аналізу мови, розпізнавання мовних елементів і зниження рівня шуму в звукових сигналах. Результати експериментів підтверджують ефективність підходу порівняно з класичними методами.

Ключові слова: перетворення чірплет, перетворення Гільберта – Хуанга, емпіричне модова декомпозиція, мел-шкала, альтернативні нелінійні шкали.

© І.А. Безвербний, 2025

УДК 004.93:534.8(07)

DOI:10.34229/2707-451X.25.1.7

І.А. БЕЗВЕРБНИЙ

ЧІРПЛЕТ-АНАЛІЗ МОВНИХ СИГНАЛІВ НА ОСНОВІ ПЕРЕТВОРЕННЯ ГІЛЬБЕРТА – ХУАНГА

Вступ. Аналіз звукових сигналів це одна з ключових задач у сфері цифрової обробки сигналів, що знаходить застосування у різних галузях, зокрема, розпізнаванні мови, обробці музичних даних, аудіофільтрації та інших. Властивості звукових сигналів часто включають нелінійність, нестійкість та швидкі зміни у часі, що робить їх ефективний аналіз складним завданням. Традиційні методи, такі як швидке перетворення Фур'є та вейвлет-перетворення, мають обмеження, пов'язані з фіксованим часово-частотним розподілом. Перетворення Гільберта – Хуанга, завдяки своєму адаптивному підходу до декомпозиції сигналів, відкриває нові можливості для точного аналізу часово-частотних характеристик. Запропонований метод чірплет-перетворення базується на інтеграції перетворення Гільберта – Хуанга з концепціями чірплет-аналізу і дозволяє враховувати нелінійні та нестійкі характеристики звукових сигналів у мел-шкалі або альтернативній нелінійній частотній шкалі (АЧШ) [1]. Такий підхід дозволяє точніше сегментувати сигнал на функціональні компоненти та виділяти його ключові характеристики, що має велике значення для задач розпізнавання мови та інших застосувань.

Сутність проблеми. Сегментація та виділення звукових характеристик мовних сигналів це одна з найбільш важливих задач у цифровій обробці сигналів мовлення. Особливості мовних сигналів, такі як їх нестабільність у часі та широке частотне охоплення, створюють складнощі для ефективного аналізу та обробки. Традиційні методи, такі як дискретне перетворення Фур'є та вейвлет-перетворення, мають певні обмеження. Дискретне перетворення Фур'є добре працює для стаціонарних сигналів, однак має низьку точність у випадку часових змін, оскільки не може забезпечити гнучке розділення часово-частотних характеристик. Вейвлет-перетворення забезпечує часткове розв'язання цієї проблеми, але його ефективність обмежується вибором вейвлет-функції, яка може не відповідати специфічним особливостям аналізованих

сигналів. Крім того, в сучасних системах розпізнавання мови та аудіоаналізу – це важлива можливість сегментації звуку на значущі компоненти, такі як фонемі, голосні, приголосні тощо, а також виділення енергетичних і частотних характеристик у змінному часовому контексті. Для досягнення цього потрібно адаптивне перетворення, яке здатне ефективно локалізувати зміни сигналу в часі та враховувати нелінійні особливості.

Перетворення Гільберта – Хуанга дозволяє адаптивно декомпонувати сигнал, використовуючи емпіричну модову декомпозицію, і визначати його часово-частотну структуру через миттєві частоти. Утім, для аудіоаналізу важливо враховувати нелінійну мел-шкалу або іншу АЧШ, що відображає людське сприйняття і відтворення мовного звуку. У цьому контексті поєднання перетворення Гільберта – Хуанга та чірплет-аналізу дозволяє ефективно вирішувати проблему сегментації звукових сигналів, а конкретно – виділення звукових характеристик людського мовлення, що відкриває нові перспективи для розпізнавання мови та інших застосувань.

Ступінь розробки. Проблематика аналізу звукових сигналів активно досліджується у сфері цифрової обробки сигналів, де акцентується увага на точному представленні часово-частотної структури. Одним із найпоширеніших методів є дискретне перетворення Фур'є, яке використовується для виділення частотних компонент, однак воно має обмежену здатність відображати короткочасні зміни у сигналі. Згодом широке застосування знайшли вейвлет-перетворення, які дозволяють локалізувати сигнали у часі та частоті, але їх ефективність значною мірою залежить від вибору базової функції. Роботи Дебоші, Коїфмана, Морле та інших науковців [2–4] стали основою для створення низки вейвлетів, які забезпечують адаптивний аналіз сигналів, але мають певні обмеження у разі нелінійних чи нерегулярних сигналів.

Ще один важливий підхід – це сучасні дослідження у сфері розпізнавання мови та аудіоаналізу зосереджені на сегментації звукового сигналу на важливі елементи [5, 6], зокрема, фонемі, голосні та приголосні, а також на виділенні їхніх енергетичних і частотних характеристик у змінному часовому інтервалі. Поряд з використанням перетворення Гільберта – Хуанга у поєднанні з емпіричною модовою декомпозицією це дозволяє адаптивно декомпонувати сигнал на складові моди і визначати миттєві частоти. Перетворення Гільберта – Хуанга зарекомендувало себе як ефективний інструмент для аналізу нестабільних і нелінійних сигналів, таких як біосигнали та геофізичні дані. У сфері обробки мовних сигналів використання перетворення Гільберта – Хуанга дає можливість виділити змінні характеристики сигналу в часовому контексті, але потребує подальшого вдосконалення для роботи з нелінійною АЧШ. Як альтернативні частотні шкали можуть розглядатися шкала Барка (Bark Scale), ERB шкала (Equivalent Rectangular Bandwidth), частотна шкала октав, шкала семитонів (Pitch Scale), тонотопічна шкала (Tonotopic Scale) та інші АЧШ, які дозволяють спеціалізувати визначення особливих точок мовного сигналу.

Впровадження АЧШ-чірплетів у поєднанні з перетворенням Гільберта – Хуанга дозволяє покращити часово-частотне відображення та забезпечує більш ефективну сегментацію звукових сигналів. Подібні підходи вже використовувалися у різних галузях, включаючи розпізнавання мови та аудіоаналіз, але методики, що поєднують перетворення Гільберта – Хуанга і чірплетний аналіз, залишаються предметом активних досліджень [1, 7].

Ідея методу. Чірплет-перетворення на базі перетворення Гільберта – Хуанга для аналізу звукових сигналів базується на синтезі методів емпіричної модової декомпозиції, перетворення Гільберта та чірплетного представлення даних. Основна ідея полягає у тому, щоб забезпечити точну часово-частотну локалізацію компонент сигналу з урахуванням його нелінійних і нестационарних властивостей, а також адаптивно опрацювати звуковий сигнал в індексах мел-шкали або альтернативної їй нелінійної частотної шкали в пошуку його кращої інтерпретованості. Практична реалізація показує, що в рамках цього методу використання мел-шкали для визначення часово-частотних характеристик сигналу не завжди забезпечує розв'язування задач сегментації та класи-

фікації мовних сигналів. Тому важливим моментом є використання саме альтернативних нелінійних шкал частотної індексації.

На першому етапі сигнал розкладається на набір внутрішніх модальних функцій, які є базовими компонентами сигналу. Цей розклад дозволяє виділити адаптивні базові сигнали, кожен із яких відповідає окремій частотній складовій. Для кожної внутрішньої модальної функції виконується перетворення Гільберта, що дає змогу отримати часово-частотне представлення миттєвих частот і амплітуд. Це забезпечує виявлення динамічних змін частотних характеристик у часі. Отримані часово-частотні дані трансформуються у чірплет-простір, де вісь частот представлена в нелінійній мел-шкалі або АЧШ. Це дозволяє сформувати двовимірну матрицю частотних індексів у відповідній нелінійній шкалі. Для візуальної перевірки результати аналізу можуть бути представлені у вигляді 3D-графіка, де вісь X відповідає часовим індексам, вісь Y – частотним індексам у відповідній нелінійній шкалі, а вісь Z – значенням чірплет-амплітуд. На основі отриманих даних виконується сегментація сигналу на фонеми. Аналіз чірплет-амплітуд дозволяє виділити особливі точки, які характеризують сигнал.

Метод чірплет-перетворення на основі перетворення Гільберта – Хуанга відрізняється від традиційних методів (таких як швидке перетворення Фур'є або вейвлет-перетворення [8, 9]) здатністю адаптивно підлаштовуватись до структури сигналу. Априорний вибір базисних функцій може бути алгоритмізований, на відміну від вейвлет-аналізу, що є ключовою перевагою для аналізу складних мовних сигналів.

Перетворення Гільберта – Хуанга дозволяє аналізувати мовні сигнали на рівні миттєвих частот, що дає можливість відстежувати зміни у частотному спектрі сигналу в реальному часі. Щоб визначити миттєві частоти мовного сигналу за допомогою перетворення Гільберта – Хуанга, потрібно реалізувати емпіричну модову декомпозицію мовного сигналу на набір внутрішніх модальних функцій.

Вхідний сигнал $x(t)$ розкладається на суму внутрішніх модальних функцій і залишкову компоненту:

$$x(t) = \sum_{i=1}^n IMF_i(t) + r_n(t). \quad (1)$$

Кожна внутрішня модальна функція $IMF_i(t)$ – це коливання на різних часових масштабах.

Після отримання внутрішньої модальної функції $IMF_i(t)$, для кожної з них обчислюється аналітичний сигнал $z_i(t)$ за допомогою перетворення Гільберта:

$$z_i(t) = IMF_i(t) + j \cdot H[IMF_i(t)] \quad (2)$$

– тут $H[IMF_i(t)]$ – це Гільбертове перетворення внутрішньої модальної функції $IMF_i(t)$.

Тоді аналітичний сигнал можна представити у полярній формі:

$$z_i(t) = A_i(t) \cdot e^{j\theta_i(t)}, \quad (3)$$

де $A_i(t)$ – миттєва амплітуда, $\theta_i(t)$ – миттєва фаза.

Далі миттєва частота обчислюється як похідна миттєвої фази:

$$\omega_i(t) = \frac{d\theta_i(t)}{dt}. \quad (4)$$

Отримавши миттєві частоти для кожної внутрішньої модальної функції, можна проаналізувати зміни частотного спектру сигналу. Миттєві частоти $\omega_i(t)$ дозволяють побудувати часо-частотний розподіл сигналу, який показує, як частота сигналу змінюється під час вимови слів, що потрібно для виявлення тональних змін та переходів між різними фонемами та аналізу складних нелінійних коливань у мовних сигналах. На основі отриманих миттєвих частот можна виявляти специфічні патерни мовлення та інші характеристики мовного сигналу шляхом використання АЧШ. Завдяки додатковій фільтрації у розмірній сітці АЧШ і подальшій нормалізації отриманої вигнутої в особливих точках поверхні, такі точки можуть бути виявлені, і відповідно буде проведена фрагментація звукового сигналу якомога ближче до наявних у мовному сигналі ділянок звучання фонем.

Спроба практичного застосування. Для поділу мовного сигналу на фрагменти, які максимально наближені до ділянок звучання фонем, у рамках методу передбачені фільтрація з метою зменшення шуму за допомогою низькочастотного лінійного 64-розрядного фільтру та нормалізація сигналу до однорідного рівня гучності з метою уникнення спотворення від гучних частин.

В процесі реалізації методу був досліджений звуковий файл з частотою дискретизації 16 кГц і глибиною квантування 32 біта, що подається на осцилограмі рис. 1.

На рис. 2 показано 12 внутрішніх модальних функцій, отримані як результат емпіричної модової декомпозиції на основі перетворення Гільберта – Хуанга. Результат встановлення зв'язків між миттєвими частотами і АЧШ – двовимірна матриця, що візуалізується на графіку час-частота на рис. 3.

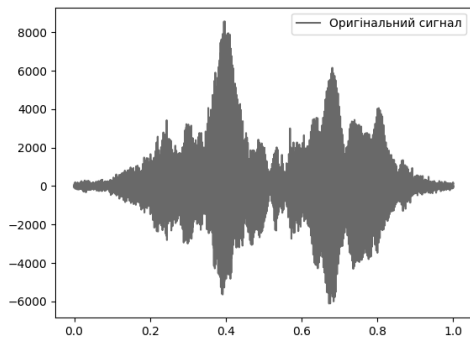


РИС. 1. Оригінальний сигнал

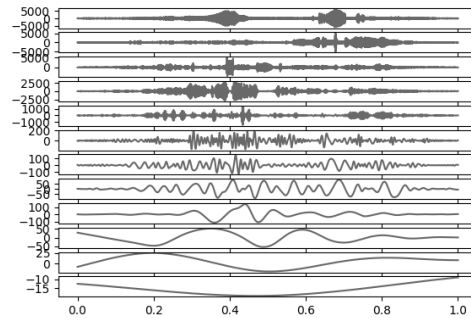


РИС. 2. 12 внутрішніх модальних функцій

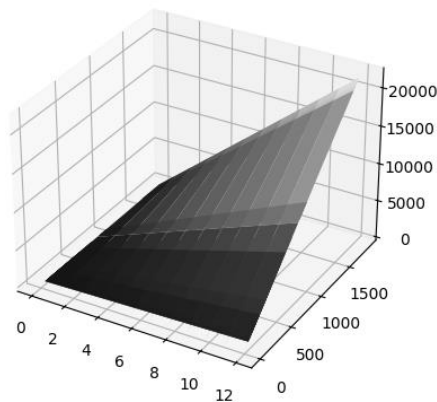


РИС. 3. Візуалізація чірплет-перетворення

Для спрощення процедури проведено на мел-шкалі:

```
def hht_to_mel(imfs, fs):
    mel_frequencies = []
    mel_phase = []
    mel_amplitudes = []
    for imf in imfs:
        analytic_signal = hilbert(imf)
        instantaneous_phase = np.unwrap(np.angle(analytic_signal))
        instantaneous_frequency = np.unwrap(np.diff(instantaneous_phase) /
            (2.0 * np.instantaneous_amplitude = np.array(analytic_signal)
        mel_freq = 2595 * np.log10(1 + abs(instantaneous_frequency) / 700.0)
        mel_frequencies.append(mel_freq)
        mel_phase.append(instantaneous_phase)
        mel_amplitudes.append(instantaneous_amplitude)
    return mel_amplitudes, mel_frequencies, mel_phase
```

Це дозволяє отримати масив особливих точок для подальшого фреймування звукового файлу шляхом фільтрації і нормалізації отриманої викривленої поверхні чірплет-перетворення. Фреймування відбувається за допомогою виявлених особливих точок часово-частотного розподілу сигналу. Оскільки мовний сигнал є неперервним, для збереження зв'язку між сусідніми кадрами використовується перекриття фреймів (20–50%).

```
# Особливі точки, отримані в результаті перетворення Гільберта – Хуанга
specialpoints = [0, ..., len(signal)]
# Розбиття сигналу на фрейми за особливими точками
def segment_signal_by_special_points(signal, specialpoints):
    frames = []
    for i in range(len(specialpoints) - 1):
        start = specialpoints[i]
        end = specialpoints[i + 1]
        frame = signal[start:end]
        frames.append(frame)
    return frames
# Розбиття сигналу
frames = segment_signal_by_special_points(signal, specialpoints)
```

Програмну реалізацію методу здійснено на мові програмування python 3.10. Подальший аналіз отриманих фреймів може бути реалізовано з урахуванням контрольного методу виділення фону мовному сигналі за допомогою ефекту стоячої хвилі [7].

Варто зауважити, що кожен фрейм можна конвертувати у вектор ознак для подальшого розпізнавання. Найпоширеніший метод – це MFCC (Mel-Frequency Cepstral Coefficients) [10], який моделює властивості людського слуху. MFCC використовується для зменшення кількості даних, зберігаючи важливі акустичні характеристики. Також під час роботи з масивом аудіоданих для видалення пауз або неінформативних частин сигналу можна використовувати систему VAD (Voice Activity Detection) інтегровану в мові python для виявлення активних ділянок мовлення.

Висновки. У цій статті представлено метод чірплет-перетворення на основі перетворення Гільберта – Хуанга для аналізу звукових сигналів, що дозволяє забезпечити ефективну сегментацію та виділення ключових характеристик сигналу. Запропонований підхід поєднує можливості перетворення Гільберта-Хуанга для адаптивного аналізу нестабільних сигналів із використанням мел-шкали, що підвищує точність представлення звукової інформації у часово-частотному просторі. Завдяки цьому підходу можливо отримати детальніші часово-частотні залежності звукових сигна-

лів, зокрема їх локалізовані зміни, які мають критичне значення для задач розпізнавання мовлення, аудіоаналізу та інших суміжних застосувань.

У подальших дослідженнях передбачено оптимізацію методу з метою зменшення обчислювальної складності та інтеграцію алгоритмів машинного навчання для автоматичної класифікації звукових фрагментів на основі отриманих ознак перетворення Гільберта – Хуанга та чірплет-перетворення. Таким чином, запропонований метод відкриває перспективи для подальшого вдосконалення обробки звукових сигналів з урахуванням їх складної структури та мінливості.

Список літератури

1. Semotiuk M.V., Palagin A.V. Technocratic model of the human auditory system. arXiv preprint arXiv:2310.05639, 2023. <https://doi.org/10.48550/arXiv.2310.05639>
2. Daubechies I. Ten lectures on wavelets. Society for industrial and applied mathematics. 1992. 350 c. <https://jqichina.wordpress.com/wp-content/uploads/2012/02/ten-lectures-of-waveletsefbc88e5b08fe6b3a2e58d81e8aeb2efbc891.pdf>
3. Coifman R.R. Wavelets and their applications past and future. Independent Component Analyses, Wavelets, Neural Networks, Biosystems, and Nanoengineering VII. SPIE, 2009. P. 23–35.
4. Grossmann A., Morlet J. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*. 1984. **15** (4). P. 723–736.
5. Sazhok M., Poltyeva A., Robeiko A., Seliukh R., Fedoryn D. Punctuation Restoration for Ukrainian Broadcast Speech Recognition System based on Bidirectional Recurrent Neural Network and Word Embeddings. *Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021)*. 2021. Vol. I: Main Conference. P. 300–310. <http://ceur-ws.org/Vol-2870/paper25.pdf>
6. Sazhok M.M., Robeiko V.V., Smoliakov Ye.A., Zabolotko T.O., Seliukh R.A., Fedoryn D.Ya., Yukhymenko O.A. Modeling Domain Openness in Speech Information Technologies. *Control Systems and Computers*. 2023. Iss. 4. P. 19–28. <https://doi.org/10.15407/csc.2023.04.019>
7. Безвербний І.А. До питання виділення фонем у мовному сигналі за допомогою ефекту стоячої хвилі. *Комп'ютерні засоби, мережі та системи*. 2019. № 18. С. 32–35. <http://dspace.nbu.gov.ua/handle/123456789/168473>
8. Pesquet-Popescu B., Pesquet J.C. Ondelettes et applications. *Techniques de l'ingénieur*. 2001. 5. P. 215.
9. Lokenath D. Wavelet transforms and time-frequency signal analysis. Springer Science & Business Media, 2012. 350 p.
10. Xie X., Cai H., Li C. A Voice Disease Detection Method Based on MFCCs and Shallow CNN. arXiv preprint arXiv:2304.08708, 2023. <https://arxiv.org/pdf/2304.08708>

Одержано 06.01.2025

Безвербний Ігор Анатолійович,

кандидат технічних наук, старший науковий співробітник
Інституту кібернетики імені В.М. Глушкова НАН України, Київ.
<https://orcid.org/0009-0005-4344-3068>
ihorbezverbnyi@gmail.com

УДК 004.93:534.8(07)

І.А. Безвербний

Чірплет-аналіз мовних сигналів на основі перетворення Гільберта – Хуанга

Інститут кібернетики імені В.М. Глушкова НАН України, Київ
Листування: ihorbezverbnyi@gmail.com

Вступ. У статті пропонується новий підхід до аналізу мовних сигналів на основі чірплет-перетворення, яке об'єднує перетворення Гільберта – Хуанга з чірплет-аналізом. Цей метод забезпечує розширені можливості сегментації та виділення ознак у мовних сигналах, уможливаючи точну ідентифікацію частотно-часових характеристик. Пропонується подолати обмеження традиційних методів, таких як швидке перетворення Фур'є і вейвлет-аналіз, використовуючи більш адаптивне рішення, адаптоване до нелінійної та нестационарної природи мовних сигналів.

Мета роботи. Створення чисельно-аналітичного методу фонетичного аналізу мовного сигналу. Центральним у методології є поєднання емпіричної модової декомпозиції від перетворення Гільберта – Хуанга з чірплет-проекціями на альтернативні нелінійні масштаби, такі як мел-шкала. Цей підхід забезпечує чудову локалізацію динамічних змін у частотно-часовій області, одночасно узгоджуючи їх із перцептивними характеристиками людського слуху. Використовуючи чірплет-перетворення, запропонований метод покращує виявлення лінгвістичних елементів, включаючи фонему та інші сегменти мови, навіть за наявності компонентів, що накладаються.

Результати. Практична реалізація цього методу продемонстрована шляхом експериментального аналізу мовних сигналів. Результати вказують на покращення точності сегментації порівняно зі звичайними підходами. Часово-частотні візуалізації ілюструють адаптивність методу для обробки складних мовних сигналів із різними динамічними властивостями.

Висновки. Це дослідження сприяє прогресу в аналізі мовлення, розпізнаванні та обробці аудіосигналів, пропонуючи потенційні застосування в таких сферах, як системи з голосовим керуванням, лінгвістичні дослідження та технології розпізнавання мовлення. Запропонований підхід можна додатково вдосконалити та інтегрувати з алгоритмами машинного навчання для автоматизації класифікації та аналізу мовних сегментів. Стаття забезпечує основу для майбутніх досліджень перетину чірплет-перетворень і нелінійної обробки сигналів, підкреслюючи їхню роль у вирішенні реальних проблем у мовних і аудіотехнологіях.

Ключові слова: перетворення чірплет, перетворення Гільберта – Хуанга, емпіричне модова декомпозиція, мел-шкала, альтернативні нелінійні шкали.

UDC 004.93:534.8(07)

Ihor Bezverbnyi

Chirplet Analysis of Speech Signals Based on the Hilbert–Huang Transform

V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine, Kyiv

Correspondence: ihorbezverbnyi@gmail.com

Introduction. This article proposes a novel approach to speech signal analysis based on the chirplet transform, which integrates the Hilbert – Huang transform with chirplet analysis. This method provides enhanced segmentation and feature extraction capabilities, enabling accurate identification of time-frequency characteristics in speech signals. It is proposed to overcome the limitations of traditional methods such as Short-Time Fourier transform and wavelet analysis, by offering a more adaptive solution tailored to the non-linear and non-stationary nature of speech signals.

The purpose of the work is to develop a numerical-analytic method for phonetic analysis of speech signals. The central feature of the methodology is the combination of empirical mode decomposition from Hilbert – Huang transform with chirplet projections onto alternative nonlinear scales, such as the mel-scale. This approach ensures superior localization of dynamic changes in the frequency-time domain, while ensures superior with the perceptual characteristics of human hearing. By leveraging chirplet transforms, the proposed method enhances the detection of linguistic elements, including phonemes and other speech segments, even in the presence of overlapping components.

Results. The practical implementation of this method is demonstrated through experimental analysis of speech signals. The results indicate an improvement in the accuracy of segmentation and noise suppression compared to conventional approaches. Time-frequency visualizations illustrate the adaptability of the method in handling complex speech signals with varying dynamic properties.

Conclusions. This research contributes to advancements in speech analysis, recognition, and audio signal processing, offering potential applications in areas such as voice-controlled systems, linguistic studies, and speech recognition technologies. The proposed approach can be further refined and integrated with machine learning algorithms to automate the classification and analysis of speech segments. The article provides a foundation for future studies on the intersection of chirplet transforms and nonlinear signal processing, emphasizing their role in addressing real-world challenges in speech and audio technologies.

Keywords: chirplet transform, Hilbert – Huang transform, empirical mode decomposition, mel-scale, alternative nonlinear scales.