

УДК 004.934: 004.42

DOI: 10.18372/2073-4751.84.20901

Рябий, М.О., к.т.н.,
orcid.org/0000-0002-9651-9135,
m.o.ryabyy@gmail.com
Шатохін, М.А., аспірант КАІ
orcid.org/0000-0003-0028-6208
n.shatokhin@gmail.com

ОГЛЯД МЕТОДІВ ВИЗНАЧЕННЯ МІСЦЕЗНАХОДЖЕННЯ ДЖЕРЕЛ ЗВУКУ

Державний університет «Київський авіаційний інститут»

Вступ

Визначення місцезнаходження (локалізація) джерел звуку (Sound Source Localization, SSL) — міждисциплінарна сфера на стику акустики, цифрової обробки сигналів і штучного інтелекту. Вона застосовується в найрізноманітніших галузях — від систем безпеки до робототехніки та доповненої реальності — всюди, де потрібно швидко визначити, звідки походить звук (наприклад, щоб виявити місце пострілу або зорієнтувати робота на голос людини). Використовуючи різні методи обробки сигналів, можна визначити напрямок та відстань до джерела звуку.

Галузь локалізації звуку має понад сторічну історію розвитку. Перші роботи з вивчення біологічних механізмів локалізації звуку з'явилися в середині XIX століття. Одним із першопрохідців є Джон Вільям Стратт, який один з перших порушив питання про механізм локалізації звуку у біологічних системах, та Сільванус Томпсон, який експериментально продемонстрував визначення напрямку на джерело звуку на основі різниці фаз між звуковими хвилями з різних вух [1].

Під час Першої світової війни акустична локалізація набула значного розвитку. Використовувалися акустичні труби, дзеркала та пізніше мікрофонні системи для визначення позицій

ворожого артилерії за допомогою аналізу звукових хвиль від пострілів. Цей метод дозволяв обчислювати координати джерела звуку, що було критично важливим для контрбатарейної боротьби [2].

У міжвоєнний період звукова локалізація продовжувала розвиватися, особливо в контексті протиповітряної оборони. Розроблялися спеціальні акустичні дзеркала та інші пристрої для виявлення наближення літаків за звуком їх двигунів. Однак з появою радарних технологій під час Другої світової війни акустичні методи поступово втратили своє значення у військовій сфері [3].

Після Другої світової війни дослідження в галузі SSL перейшли до цивільних застосувань. З розвитком електроніки та обчислювальної техніки з'явилися точніші та швидші системи локалізації звуку; зокрема, сучасні алгоритми на основі нейронних мереж визначають напрямок на джерело майже миттєво і з мінімальною похибкою, але потребують попереднього тренування на великих об'ємах даних.

Постановка проблеми

З огляду публікацій, що вже існують, вирішення проблеми визначення координат джерела звуку мало велику кількість спроб та успішних досліджень. Хоча за останні десятиліття точність локалізації звуку помітно зросла, деякі фундаментальні проблеми досі лишаються невирішеними. Сильна реверберація

приміщення або високий рівень шуму можуть майже повністю звести нанівець ефективність нинішніх алгоритмів локалізації. Динамічне середовище — рухомі джерела чи переміщення самих сенсорів — також створює додаткові труднощі, які поки що вдається подолати лише частково. Крім того, часто виникає проблема ціни компонентів для побудови системи або обмеження щодо її енергоспоживання, що також спонукає до її вирішення.

Метою даної роботи є аналіз сучасних методів локалізації джерела звуку, порівняння їх характеристик та визначення оптимальних підходів для різних застосувань та порівняння їх з власними підходами.

Класифікація методів

Серед різноманіття алгоритмів локалізації джерел звуку можна виділити кілька основних груп, що відрізняються підходом до обробки сигналу та відновлення просторового розташування джерела. Деякі методи покладаються на часові затримки між мікрофонами, інші працюють із фазовою інформацією або енергетичними характеристиками сигналу, тощо.

З огляду на різноманітність існуючих підходів, сестиматизація методів локалізації джерела звуку є необхідною передумовою для їх ґрунтовного та полідовного аналізу. На рисунку 1 представлена запропонована авторами класифікація методів локалізації джерела звуку.



Рис. 1. Класифікація методів локалізації джерел звуку

1. Методи, що базуються на часових характеристиках сигналу

Ці методи використовують різницю в часі прибуття сигналу до різних сенсорів (мікрофонів або антен), щоб оцінити напрямок приходу хвилі. Основний принцип — вимірювання різниці часу прибуття (Time Difference of Arrival, TDoA), що дозволяє визначити відстань, пройдену сигналом від джерела до сенсора, i , на основі цього, обчислити напрямок. Це класичні методи, які часто використовуються в

задачах, пов'язаних з локалізацією звукових джерел.

Шлях сигналу до кожного мікрофона — відстань:

$$d_i = |q - r_i|, d_j = |q - r_j|, \quad (1)$$

де r_i, r_j — координати мікрофонів та q — координата джерела звуку.

Якщо джерело близько до мікрофонів, тому хвилі розходяться і

потрібно враховувати реальні відстані (ближнє поле):

$$\tau_{ij}(q) = \frac{|q-r_i| - |q-r_j|}{c} \quad (2)$$

Якщо джерело дуже далеко, хвилі приходять майже паралельно. Тоді напрям на джерело описується одиничним вектором $u(\theta)$, і TDoA лінеаризується (далеке поле):

$$\tau_{ij}(\theta) = \frac{(r_i - r_j) \cdot u(\theta)}{c} \quad (3)$$

Оскільки q та $u(\theta)$ невідомі, тому TDoA оцінюється за сигналами за допомогою генералізованої крос-кореляції.

Generalized Cross-Correlation (GCC) є базовим методом знаходження різниці часу прибуття сигналів (TDoA) шляхом обчислення крос-кореляційної функції [4].

Базова формула GCC:

$$R_{ij}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \underbrace{\Psi_{ij}(\omega)}_{\text{вагова функція}} \underbrace{G_{ij}(\omega)}_{\text{функція крос-спектр}} e^{j\omega\tau} d\omega \quad (4)$$

де $G_{ij}(\omega) = X_i(\omega) X_j^*(\omega)$ — крос-спектральна густина (CPSD) між каналами i, j , $X_i(\omega)$ — Фур'є-образ (або STFT-коефіцієнт) сигналу x_i , $\Psi_{ij}(\omega)$ — вагова функція (головне, що відрізняє варіанти GCC), $e^{j\omega\tau}$ — «узгодження» на затримку τ .

На практиці це реалізують через STFT і суму по частотних бінах:

$$R_{ij}[l] \approx \sum_f \Psi_{ij}(f) X_i(f) X_j^*(f) e^{j2\pi f l / F_s} \quad (5)$$

де f — індекс частоти (бін), F_s — частота дискретизації, l — зсув у відліках.

Оцінка TDoA:

$$\hat{\tau}_{ij} = \underset{\tau}{\arg \max} R_{ij}(\tau) \quad (6)$$

Варіації GCC використовують різні вагові функції, які підвищують точність або стійкість до шуму та

реверберації. GCC-PHAT зберігає лише фазову інформацію спектра, що робить його більш стійким до реверберації та ефективним для ширококугових сигналів, зокрема мовлення [4] — на відміну від звичайного GCC, який враховує і амплітуду. GCC-SCOT використовує когерентність сигналів як вагу і може забезпечувати підвищену стійкість у середніх умовах шуму, хоча поступається PHAT у ревербераційних середовищах [5][6]. GCC-ROTH базується на спектральних характеристиках сигналів і дозволяє краще придушувати шум, однак програє PHAT у точності через чутливість до спектральної різниці між каналами [6]. GCC-NLT вводить нелінійне спектральне перетворення перед обчисленням кореляції, що підвищує стабільність у складних акустичних умовах [7]. Нарешті, GCC-PHAT-SM поєднує PHAT з маскуванням мовлення (на основі VAD або нейронної оцінки), що дозволяє виділити мовні частоти та покращити локалізацію мовця в шумі [8] — на відміну від звичайного PHAT, який обробляє повний спектр без фільтрації.

$$\Psi_{ij}^{CC}(\omega) = 1 \quad (7)$$

$$\Psi_{ij}^{PHAT}(\omega) = \frac{1}{|X_i(\omega) X_j^*(\omega)| + \epsilon} \quad (8)$$

$$\Psi_{ij}^{SCOT}(\omega) = \frac{1}{\sqrt{G_{ii}(\omega) G_{jj}(\omega)} + \epsilon} \quad (9)$$

$$\Psi_{ij}^{ROTH}(\omega) = \frac{1}{G_{ii}(\omega) + \epsilon} \quad (10)$$

$$\Psi_{ij}^{WPHAT}(\omega) = \frac{1}{|X_i(\omega) X_j^*(\omega)|^\alpha + \epsilon}, \quad (11)$$

$$0 < \alpha \leq 1$$

$$R_{ij}^{NLT}(\tau) = \sum_f \frac{X_i(f) X_j^*(f)}{|X_i(f) X_j^*(f)|^\alpha + \epsilon} e^{j2\pi f \tau} \quad (12)$$

$$R_{ij}^{PHAT-SM}(\tau) = \sum_f M(f) \frac{X_i(f) X_j^*(f)}{|X_i(f) X_j^*(f)|} \quad (13)$$

де ϵ — числова стабілізація.

Часові методи добре працюють у випадках, коли затримка сигналу може

бути точно виміряна, але вони можуть бути менш ефективними в складних акустичних середовищах.

Альтернативним підходом є підпросторові методи, які аналізують статистичні властивості масиву сигналів, зокрема коваріаційні структури, для виділення сигнального підпростору.

2. Підпросторові методи

Ці методи використовують фазову інформацію, зокрема різницю фаз між сигналами, що надходять на різні сенсори. Саме ця фазова різниця, у поєднанні з просторовим моделюванням, дозволяє точно оцінити напрямки на джерело звуку. За достатньої довжини спостереження та високого співвідношення сигнал/шум підпросторові методи забезпечують роздільність, яка значно перевищує дифракційні межі масиву.

$$R = E \{ x(t) x^H(t) \} \quad (14)$$

$$R = E_s \Lambda_s E_s^H + E_n \Lambda_n E_n^H \quad (15)$$

MUSIC (Multiple Signal Classification) — це підпросторовий метод оцінки напрямку приходу сигналу (DOA), який аналізує власні вектори коваріаційної матриці прийнятих сигналів. Він розділяє простір на сигнальний і шумовий підпростори: сигнали приходять у напрямках, ортогональних до шумового підпростору. Для кожного можливого напрямку MUSIC обчислює псевдоспектр, де піки відповідають оцінкам напрямків джерел. Хоча цей метод і забезпечує дуже високу роздільну здатність (дозволяючи розрізняти кілька джерел навіть за сильного шуму), він має й недоліки: алгоритм потребує значних обчислювальних ресурсів та наперед відомої кількості джерел, до того ж чутливий до похибок оцінки коваріаційної матриці та малих обсягів даних [9].

$$P_{\text{MUSIC}}(\theta) = \frac{1}{a^H(\theta) E_n E_n^H a(\theta)} \quad (16)$$

Приклад псевдоспектру MUSIC наведено на рисунку 2.

Псевдоспектр MUSIC (кут=азимут, радіус=висота; центр=-90°, край=+90°)

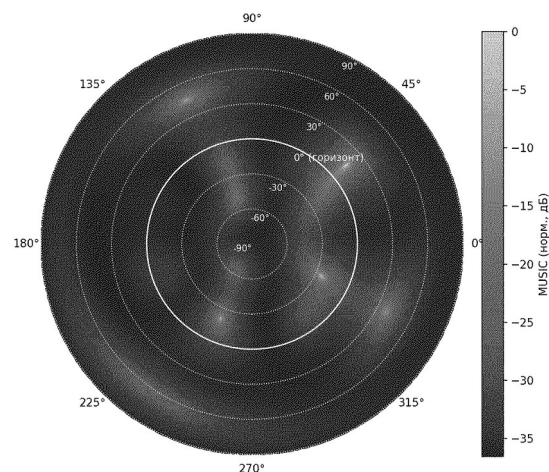


Рис. 2. Псевдоспектр MUSIC (піки на графіку показують найбільш ймовірне розташування джерел звуку)

Існує кілька ефективних варіацій MUSIC. Root-MUSIC уникає повного

перебору напрямків, знаходячи корені полінома, що відповідає пікам

псевдоспектра — це значно знижує обчислювальну складність при збереженні точності, але потребує строгої лінійної геометрії масиву [10] [11]. FAST-MUSIC прискорює алгоритм для майже періодичних сигналів, замінюючи власний розклад на FFT через апроксимацію коваріаційної матриці циркулянтною; це зменшує складність з $O(N^3)$ до $O(N \log N)$, але ефективний лише для спеціальних типів сигналів [12]. GPU-/FPGA-реалізації MUSIC дають значне прискорення за рахунок апаратного паралелізму, дозволяючи використовувати MUSIC у реальному часі, але вимагають спеціалізованої архітектури та ресурсів [13][14].

ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) — ще один підпросторовий метод DOA-оцінки, який використовує властивість обертальної інваріантності між підмасивами сенсорів. На відміну від MUSIC, він не виконує пошук по всьому простору напрямків, а напряму обчислює кути на основі співвідношень між сигнальними підпросторами. Це дає суттєве зниження обчислювальної складності при подібній точності. Основна перевага ESPRIT — швидкість і менша залежність від об'єму даних; головні недоліки — вимога до точної та симетричної геометрії масиву (наприклад, ULA), а також нижча стійкість до геометричних похибок або неповної обсервації [15].

$$E_s^{(2)} = E_s^{(1)} \Phi \quad (17)$$

$$\Psi_1^T \Psi_2 v_i = \lambda_i v_i \quad (18)$$

Серед варіацій ESPRIT: Unitary ESPRIT застосовує унітарне перетворення для роботи лише з дійсними числами, що пришвидшує обчислення та підвищує числову

стійкість [16][17]; Tensor-ESPRIT дозволяє працювати з багатовимірними масивами (наприклад, для просторово-часових або просторово-частотних структур) [18]; Sparse ESPRIT адаптований до рідкісних (неповних) масивів, але може втрачати точність при сильному шумі або неправильному моделюванні структури [19].

Поряд із фазовими оцінками напрямку, широке застосування отримали методи формування променя, які дозволяють активно підсилювати сигнал у бажаному напрямку.

3. Методи, засновані на формуванні променя (Beamforming)

Методи формування променя реалізують просторову селекцію сигналів, використовуючи масиви мікрофонів для підсилення випромінювань з певного напрямку та пригнічення шуму й інтерференцій. Принцип їх роботи полягає у введенні контрольованих затримок та вагових коефіцієнтів для кожного елемента решітки, що дозволяє формувати вузьку діаграму спрямованості та покращувати співвідношення сигнал/шум навіть у ревербераційних або багатоджерельних середовищах.

Delay-and-Sum (DAS) – класичний метод, у якому сигнали з мікрофонів затримуються відповідно до напрямку на джерело та сумуються. Має низьку обчислювальну складність, але обмежену роздільну здатність через широкий головний промінь і високі бічні пелюстки [20][21].

MVDR (Capon) – адаптивний метод, що мінімізує потужність завади, зберігаючи амплітуду сигналу з цільового напрямку; забезпечує підвищену роздільну здатність у шумових умовах [22].

Steered Response Power (SRP) – метод, що оцінює положення джерела шляхом обчислення просторової енергетичної функції, сформованої з

урахуванням часових затримок між мікрофонами; максимум цієї функції відповідає найбільш ймовірному напрямку [23]. Варіації: SRP-PHAT – підвищує стійкість у ревербераційних умовах за рахунок фазової вагової обробки [24]; SRP-PHAT-HSDA – оптимізує пошук через ієрархічне сканування та відсікання неінформативних напрямків,

зменшуючи обчислювальні витрати [25]; X-SRP – модульна версія SRP з підтримкою різних вагових функцій, стратегій сканування та паралельної обробки [24].

На рисунку 3 схематично показаний принцип роботи алгоритмів формування променя на прикладі метода Delay-and-Sum.

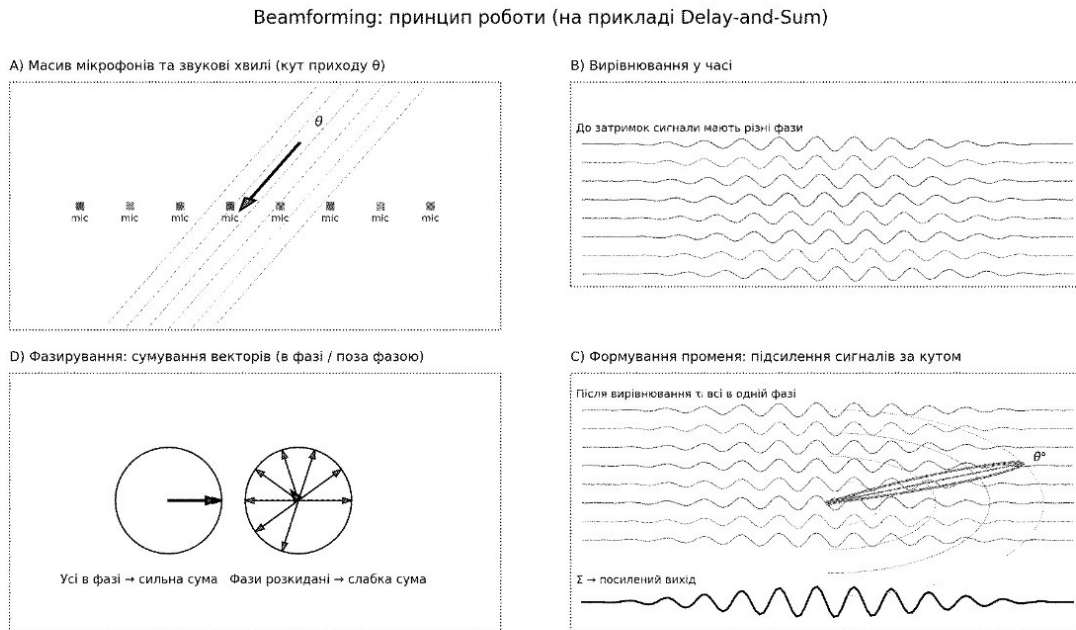


Рис. 3. Принцип роботи алгоритмів формування променя

4. Методи, що базуються на енергетичних характеристиках

На відміну від методів формування променя, які використовують часові затримки та фазову інформацію, енергетичні підходи спираються на аналіз рівнів звукової енергії, зафіксованих різними мікрофонами. Основна ідея полягає у тому, що інтенсивність сигналу зменшується зі збільшенням відстані до джерела, і ця залежність може бути використана для оцінки його місцеположення.

RSSI-Based Acoustic Localization in UWSNs – застосовує індикатор рівня отриманого сигналу (RSSI) з урахуванням зміни швидкості звуку та

синхронізації вузлів, оптимізуючи локалізацію в підводних сенсорних мережах [26].

Energy-Based Sound Source Localization with Low Power Consumption – моделює затухання звукової енергії для локалізації при мінімальному енергоспоживанні, що особливо важливо для сенсорів з обмеженим ресурсом живлення [27].

Acoustic Energy Sensing Flow Exploitation (ESFE) – покращує точність локалізації у складних акустичних сценах шляхом вибору сенсорів, які найкраще відображають напрямки потоку акустичної енергії [28].

Energy-Based Collaborative Source Localization (EBCSL) – оцінює відстань

до джерела за рівнем звукової енергії з кількох сенсорів, формуючи колективну модель просторового розташування [29].

На практиці енергетичні методи використовуються рідко через їх низьку просторову роздільну здатність, сильну залежність від акустичних властивостей середовища та чутливість до перешкод. Вони можуть бути ефективними у випадках, коли важливі простота обчислень, низьке енергоспоживання та робота у середовищах, де інші методи малоефективні (наприклад, підводна акустика або розріджені сенсорні мережі).

5. Методи, що використовують частотні характеристики

Цей клас методів базується на аналізі частотного зсуву, спричиненого рухом джерела відносно приймачів (ефект Доплера). Оцінюючи різницю частот прибуття (FDoA) або поєднання часових і частотних різниць (TDoA/FDoA), можна визначити положення чи напрямок на джерело. Такі методи є особливо корисними, коли фазові або часові вимірювання недоступні чи неточні, зокрема у рухомих або підводних системах.

Frequency Difference of Arrival (FDoA) – локалізація за виключно FDoA-вимірюваннями, з двоетапною обробкою: попередня оцінка через геометрію гіпербол і уточнення методом зважених найменших квадратів із семидефінітною релаксацією [30].

Far-field Linearized DOA Estimation from TDoA/FDoA – спрощена оцінка напрямку у далекому полі шляхом лінеаризації моделі TDoA/FDoA та розв'язання системи лінійних рівнянь, що знижує вплив шуму без складної оптимізації [31].

Algebraic TDoA/FDoA Localization with Unknown Sound Speed – дворівневий підхід, який дозволяє локалізувати джерело без точного знання швидкості звуку, враховуючи похибки у позиціях і швидкостях

сенсорів; початкова лінійна оцінка уточнюється ітераційно для досягнення точності, близької до межі Крамера–Рао [31].

6. Статистичні методи

Статистичні методи локалізації ґрунтуються на побудові моделей, що описують розподіли акустичних ознак та їх варіації. Вони дозволяють враховувати стохастичну природу сигналів, ревербераційні викривлення та багатоджерельність сцени, що робить їх придатними для складних акустичних середовищ.

Так, метод в роботі [32] ґрунтується на використанні DP-RTF (direct-path relative transfer function) у поєднанні з кластеризацією на основі Gaussian Mixture Models (GMM). Такий підхід дозволяє ефективно групувати ознаки, що відповідають різним джерелам звуку, та оцінювати їх положення у просторі навіть за наявності кількох активних сигналів.

Подальший розвиток цієї ідеї представлений у роботі [33], де GMM використовується разом із спеціальною маскою для придушення дифузної складової. Це підвищує стійкість алгоритму до реверберації та фонового шуму, забезпечуючи більш надійне визначення напрямку на кілька джерел одночасно. Таким чином, статистичні методи демонструють перспективність у багатоджерельних середовищах із підвищеним рівнем перешкод, поєднуючи помірну потребу в навчанні (наприклад, GMM через EM-алгоритм) із гнучкістю щодо геометрії мікрофонних масивів, що робить їх простішими у підготовці, ніж глибокі нейромережеві моделі.

До статистичних підходів варто також віднести роботу [34], де було запропоновано моноуральний імовірнісний метод локалізації. Його головна особливість полягає у здатності працювати лише з одним мікрофоном, що повністю усуває залежність від

геометрії масиву. Водночас такий підхід вимагає значного попереднього навчання на акустичних даних та демонструє нижчу точність і стійкість у складних акустичних середовищах.

7. Нейромережеві методи

Завдяки зростанню обчислювальних потужностей та розвитку штучного інтелекту, нейронні мережі набули широкого застосування у локалізації джерел звуку, особливо в умовах шуму, реверберації та наявності кількох джерел. Серед підходів, що успішно зарекомендували себе, окреме місце займають згорткові нейронні мережі (CNN), здатні ефективно обробляти спектрограми та інші часо-частотні представлення сигналів, виявляючи просторові залежності. У задачах локалізації CNN часто працюють з міжвушними ознаками, такими як різниця фаз (IPD) і різниця рівнів (ILD), що дозволяє оцінювати напрямок і відстань до джерела [35].

Для завдань навчання подібності між сигналами використовуються сіамські мережі, які складаються з двох ідентичних підмереж зі спільними вагами. У локалізації вони навчаються відображати пари аудіосигналів у простір ембедінгів так, щоб відстань між ними відповідала різниці напрямків джерел. Це підвищує узагальнювальну здатність у нових акустичних умовах, зокрема в бінауральних сценаріях [36].

Простіші за структурою прямопоширювальні нейронні мережі (FFNN) також знаходять застосування, особливо при роботі з енергетичними вимірюваннями або часовими різницями приходу сигналів (TDoA), що дозволяє моделювати залежність між акустичними ознаками та положенням джерела [37]. У випадках, коли важливе моделювання часової динаміки, використовують рекурентні нейронні мережі (RNN), здатні обробляти послідовності та відстежувати рухомі джерела у часі [38]. Комбінацією CNN та RNN є згортково-рекурентні мережі

(CRNN), які поєднують сильні сторони просторового аналізу та моделювання часових залежностей, забезпечуючи високу точність навіть у складних середовищах [38].

У глибших архітектурах, таких як ResNet, використання залишкових зв'язків полегшує тренування й дозволяє ефективно працювати з багаторівневими спектральними ознаками. Це дає змогу моделі вловлювати складні залежності, пов'язані з напрямком і відстанню до джерела [39]. Новіші підходи включають архітектури на основі трансформерів, зокрема Binaural Audio Spectrogram Transformer (BAST), який завдяки механізмам багатоголової уваги здатен інтегрувати інформацію з обох каналів і придушувати відлуння, досягаючи середньої кутової похибки близько $1,3^\circ$ навіть у ревербераційних умовах [40].

Іншим напрямом розвитку є автоенкодері, які навчаються стислому представленню даних і використовуються для відновлення просторової інформації навіть за частково відсутніх даних про мікрофони чи сигнал [41]. Біологічно натхнені спайкові нейронні мережі (SNN) застосовують моделі резонансних нейронів Resonate-and-Fire для кодування міжвушної різниці часу (ITD) без перетворення Фур'є, що знижує обчислювальні витрати; додатковий механізм мультислухової уваги (MAA) підвищує точність у шумних середовищах [42]. Нарешті, підхід SSLIDE демонструє, як архітектура типу енкодер–декодер може перевершити класичні алгоритми, такі як MUSIC, у ревербераційних приміщеннях, зберігаючи здатність до узагальнення на нові умови [43].

8. Гібридні методи

Гібридні підходи поєднують класичні методи обробки сигналів із сучасними архітектурами глибокого

навчання, що дозволяє підвищити точність та робастність локалізації джерел звуку. Одним із прикладів є NGCC-PHAT — нейронна версія класичного GCC-PHAT, яка навчається оцінювати TDoA для кількох джерел одночасно, використовуючи навчання, інваріантне до перестановок. Це дає змогу моделі коректно працювати з перекриваючими звуковими подіями [44].

Розширенням цього підходу є wav2pos, де NGCC-PHAT використовується для попередньої обробки сигналів з метою підвищення стійкості до реверберації та шуму. Результати подаються до маскованого автоенкодера, який, використовуючи як аудіодані, так і координати мікрофонів, навчається точно визначати положення джерел, у тому числі рухомих та множинних [41].

BeamLearning пропонує інший шлях — застосування глибоких нейронних мереж безпосередньо до необроблених багатоканальних сигналів для кутової локалізації. Навчання відбувається на сирих даних, що дозволяє уникнути спрощень та втрат інформації, характерних для традиційних передобробок [45].

Частина гібридних методів інтегрує глибинні моделі в класичні субпросторові алгоритми. Так, DAMUSIC поєднує алгоритм MUSIC із рекурентною нейронною мережею, що навчається апроксимувати матрицю коваріації, замінюючи традиційний обчислювальний етап власного розкладу. Це дозволяє ефективно працювати при низькому SNR і обмеженій кількості знімків [46].

Подібну ідею реалізує DeepMUSIC, де згорткові нейронні мережі безпосередньо оцінюють MUSIC-спектр з вхідних даних, обходячи етап обчислення коваріаційної матриці. Це забезпечує високу точність

при значному зменшенні обчислювальної складності [47]. Ще одним сучасним розвитком є TransMUSIC, який використовує трансформерну архітектуру для прямого навчання шумового підпростору з квантованих даних. Такий підхід дозволяє уникнути традиційних розрахунків коваріаційної матриці та її власного розкладу, знижує обчислювальні витрати та підвищує точність навіть при низькій роздільній здатності даних [48].

Порівняльний аналіз методів

У наведеній нижче таблиці 1 використано метод експертної оцінки, що базується на аналізі публікацій та практичних характеристик методів локалізації джерел звуку. Кожен критерій відображає окремий аспект придатності алгоритмів до роботи в реальних умовах.

Кожен критерій було конкретизовано через відповідні показники ефективності. Для джерел оцінювалася здатність алгоритму відслідковувати кілька об'єктів одночасно. Реверберація характеризувала стійкість до акустичних відбиттів: від повної втрати працездатності до стабільної роботи навіть у складних умовах. Чутливість визначається можливістю коректної роботи при низькому відношенні сигнал/шум. Для навчання вищі оцінки отримували методи, що не потребують великих обсягів даних для попередньої адаптації. Точність безпосередньо відображає середню похибку у визначенні напрямку. Критерій шум враховував стабільність роботи за наявності зовнішніх акустичних завад. Обчислювальна складність оцінювалася з огляду на вимоги до ресурсів та можливість використання на обмежених апаратних платформах. Гнучкість описувала здатність алгоритму адаптуватися до зміни умов без необхідності переналаштування.

Таблиця 1. Порівняння методів локалізації джерел звуку

Категорія	Джерела	Реверб.	Чутл.	Навчання	Точн.	Шум	Обчисл.	Гнуч.	Геом.	Мікроф.	Сума
Часові методи (GCC)	5	6	7	10	6	6	9	7	7	8	71
Підпросторові методи (MUSIC/ESPRIT)	7	4	6	10	8	6	5	6	5	5	62
Beamforming	7	7	7	10	7	7	4	8	8	6	71
Енергетичні	3	3	4	10	4	4	10	6	8	8	60
Частотні (FDoA)	4	5	5	10	6	5	6	5	7	4	57
Статистичні	7	6	7	8	8	8	3	9	9	6	73
Нейромеревеві	8	8	8	2	8	8	5	7	7	6	67
Гібридні	9	9	8	6	9	9	4	9	9	7	79

Геометрія відображала залежність від просторової конфігурації мікрофонів — від жорстко фіксованої лінійної до довільної тривимірної із можливістю перебудови. Нарешті, мікрофони оцінювалися за мінімально необхідною кількістю сенсорів для забезпечення стабільної роботи методу.

Сукупність цих критеріїв дозволяє побудувати збалансоване порівняння алгоритмів і виділити найбільш придатні для використання у практичних сценаріях.

Проведений порівняльний аналіз методів локалізації джерел звуку демонструє, що жоден із розглянутих підходів не забезпечує оптимальних характеристик за всіма критеріями одночасно. Отримані оцінки підтверджують компромісний характер задачі локалізації в реальних акустичних умовах, де вимоги до точності, стійкості до реверберації та шуму, обчислювальної складності й гнучкості часто є взаємовиключними.

Порівняльні оцінки вказують, що часові методи на основі TDOA (зокрема GCC) залишаються одним із найпривабливіших варіантів для практичних систем завдяки відсутності потреби у навчанні, низькій обчислювальній складності та слабкій залежності від конкретної геометрії

масиву. Водночас саме в умовах, які є типовими для реальних застосувань — багатоджерельність, сильна реверберація та нестационарні завади — класичний GCC демонструє деградацію: кореляційна функція містить кілька конкурентних піків, з'являються хибні максимуми через відбиття, а розділення кількох джерел стає нетривіальним [49-51].

Це формує очевидну нішу для гібридизації: зберегти GCC як фізично обґрунтований механізм витягання часової інформації (TDOA-структури), але делегувати нейромереві складну частину — інтерпретацію багатопікових кроскореляцій/кроспектрів у складній акустиці та виділення кількох джерел із оцінкою достовірності.

Запропонований гібридний метод локалізації

У цій роботі запропоновано гібридний підхід до багатоджерельної локалізації, у якому нейронна мережа працює не з "сирими" аудіосигналами або їх спектрами, а з кроспектральними ознаками GCC, що мають прямий фізичний зв'язок із часовими затримками приходу сигналу на мікрофони.

На відміну від типових нейромеревевих моделей DOA-

оцінювання, що прогнозують один напрямок або потребують фіксованого числа виходів із жорсткою прив'язкою до кількості джерел, у роботі застосовано multi-head архітектуру, в якій кожна "голова" мережі відповідає окремому кандидату джерела. Така організація виходу дозволяє моделі одночасно обробляти сцени з кількома джерелами звуку, не зводячи задачу до попереднього розділення сигналів чи кластеризації на проміжному етапі, а також знижує ризик пропуску слабких джерел при конкуренції за домінуючу складову спектра.

Важливим елементом запропонованої концепції є додавання physics-informed регуляризації (фізичних обмежень), що підсилює узгодженість прогнозів мережі з закономірностями поширення хвилі в просторі. Зокрема, вводяться два типи обмежень: геометрично-часові обмеження, які пов'язують оцінені параметри локалізації з очікуваними часовими затримками приходу сигналу на елементи мікрофонної решітки; фазові/спектральні обмеження, що враховують фазові залежності крос-спектральних ознак та дозволяють зменшити кількість фізично некоректних розв'язків, які можуть виникати у чисто data-driven моделях.

З метою демонстрації практичної перспективності підходу виконано

серію попередніх експериментів на синтетичних сценах із варіюванням рівня шуму (SNR), кількості одночасних джерел та умов спостереження, а результати подано у вигляді графіків. Отримані залежності показують, що запропоноване поєднання multi-head нейромережевої інтерпретації та physics-informed обмежень здатне покращувати стабільність оцінювання в багатоджерельних умовах і при зниженні SNR. Детальний опис архітектури, протоколу експериментів, а також розширене кількісне порівняння із базовими методами планується подати в окремій публікації.

Одноджерельний режим

Результати експериментів у режимі одного джерела представлено на рисунку 4. Класичний GCC-PHAT демонструє очікувану поведінку: при високих SNR (≥ 5 dB) він забезпечує найменшу кутову помилку завдяки однозначності фазових співвідношень. Водночас при низьких SNR (-10 dB) GCC деградує до майже випадкової оцінки (MAE $\approx 90^\circ$).

Нейромережеві методи (MLP та PINN) демонструють значно вищу робастність у шумових умовах. Серед них найкращі результати систематично показує PINN-phase, який має найменшу MAE у всьому діапазоні низьких і середніх SNR.

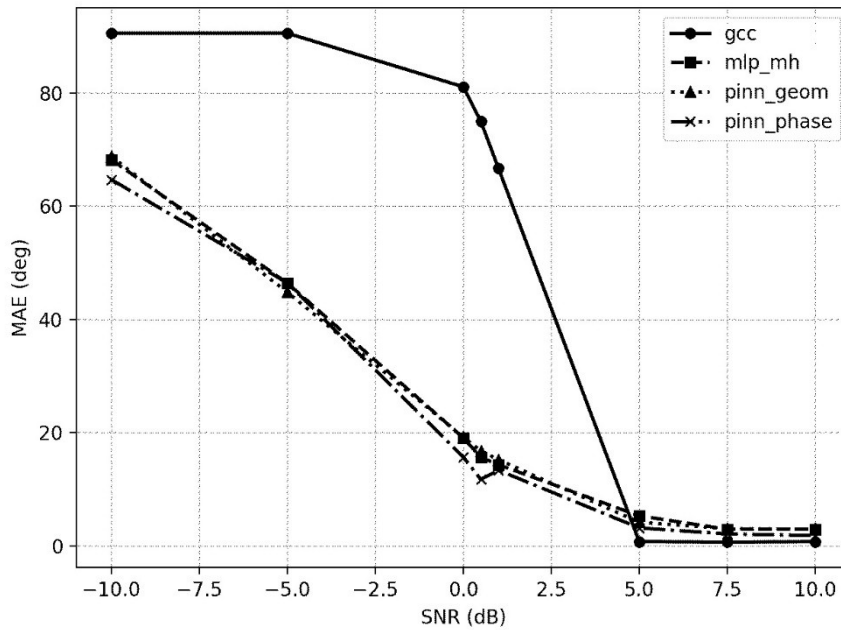


Рис. 4. Середня помилка в сценах з одним джерелом в залежності від SNR

Багатоджерельний режим

У багатоджерельному (до трьох джерел одночасно) режимі класичний GCC-RHAT принципово обмежений: він не має механізму оцінки кількості джерел і не формує стабільної multi-head структури. Хоча за метрикою MAE GCC може іноді показувати відносно низькі значення, ці оцінки не відповідають практичній задачі багатоджерельної локалізації. GCC показує лише помилку кута до найближчого за кутом або найгучнішого джерела, ігноруючи інші. Оскільки джерел в сцені більше, то помилка в

середньому менше, ніж в одноджерельній сцені.

MLP multi-head суттєво перевершує GCC, забезпечуючи структурований вихід у вигляді кількох слотів з напрямками та рівнями впевненості. Додавання фізичної регуляризації покращує точність слотів, причому найбільший вигравш забезпечує PINN-phase, який демонструє найменшу MAE для всіх SNR.

Результати експериментів в режимі декількох джерел на сцені наведено на рисунку 5.

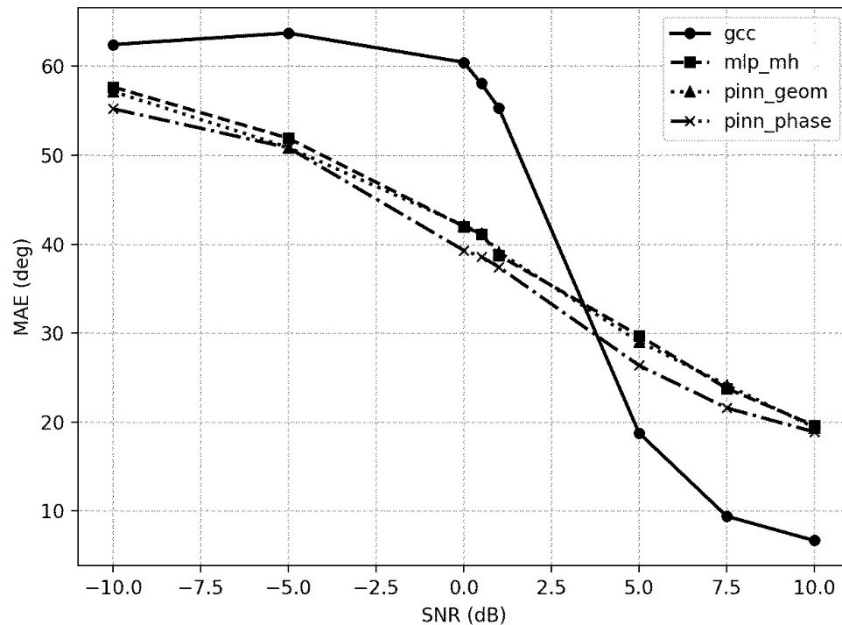


Рис. 5. Середня помилка в сценах з декількома джерелами в залежності від SNR

Порівняння фізичних гіпотез

Результати експериментів чітко показують, що фазово-узгоджена фізична гіпотеза (PINN-phase) систематично перевершує як суто data-driven MLP, так і геометричний PINN.

Геометричний PINN забезпечує лише помірний виграв над MLP. Це пояснюється тим, що геометрична інформація вже неявно присутня у вхідних ознаках нейронної мережі: крос-спектральні фазові співвідношення між мікрофонними парами безпосередньо визначаються геометрією масиву. Внаслідок цього MLP здатна частково відновлювати геометричні обмеження самостійно.

Натомість PINN-phase вводить принципово сильніше фізичне обмеження, змушуючи модель відтворювати всю комплексну багатоджерельну крос-спектральну суміш, а не лише узгоджені часові затримки. Саме це забезпечує зменшення кутової помилки, особливо у шумових і багатоджерельних сценах.

Висновок

У цій роботі виконано огляд і систематичне порівняння сучасних методів акустичної локалізації джерел звуку, включаючи часові методи на

основі TDoA (GCC), підпросторові методи (MUSIC/ESPRIT), методи формування променя, статистичні підходи, а також нейромереві та гібридні рішення. Проведений аналіз підтверджує, що жоден із підходів не є універсальним, а вибір методу визначається умовами застосування та компромісом між точністю, робастністю, обчислювальними витратами і вимогами до конфігурації мікрофонного масиву.

Показано, що класичні часові методи (зокрема GCC-PHAT) залишаються привабливими завдяки простоті реалізації, низькій обчислювальній складності та фізичній обґрунтованості. Водночас у сценах із кількома джерелами, реверберацією та низьким SNR вони схильні до деградації через появу кількох конкурентних піків і хибних максимумів у кореляційних оцінках, що ускладнює стабільне виділення та супровід кількох джерел. Підпросторові методи демонструють високу точність за виконання модельних припущень, однак їх практична ефективність суттєво залежить від геометрії масиву та рівня когерентних відбиттів. Статистичні методи є ефективними при

багатоджерельних сценах, оскільки дозволяють інтегрувати різномірні спостереження й враховувати невизначеність, проте часто потребують значних обчислювальних ресурсів. Нейромережеві методи забезпечують високу адаптивність і робастність у складній акустиці, але потребують навчання та можуть втрачати якість при зміні умов або конфігурації сенсорів.

На основі узагальнених оцінок встановлено, що найбільш практичним напрямом розвитку є гібридні багаторівневі системи, де фізично інтерпретовані ознаки використовуються для формування інформативного представлення сигналів, а статистичні та/або нейромережеві моделі — для інтерпретації складних багатопікових структур, стабілізації оцінок та обробки багатоджерельних сцен. У цьому контексті коротко описано запропоновану гібридну концепцію: використання крос-спектральних ознак GCC як вхідних даних multi-head нейромережевої моделі для оцінювання кількості активних джерел і їх напрямків разом із мірою достовірності прогнозу. Наведені у роботі результати демонструють перспективність такого підходу як основи для подальшого розвитку практичних систем локалізації, трекінгу та мультистанційної триангуляції.

Перелік літератур

1. *Yost W. A.* History of sound source localization: 1850–1950 // *Proceedings of Meetings on Acoustics*. 2017. Т. 30, № 1. Art. 050002. DOI: <https://doi.org/10.1121/2.0000529>
2. *Таланов А. В.* Звуковая разведка артиллерии [Акустична розвідка артилерії]. Москва: Военное издательство Министерства Вооруженных сил Союза ССР, 1948.
3. *Zimmerman D.* Tucker's acoustical mirrors: Aircraft detection before radar // *War & Society*. 1997. Т. 15, № 1. С. 73–99. DOI: <https://doi.org/10.1179/072924797791201003>
4. *Knapp C. H., Carter G. C.* The generalized correlation method for estimation of time delay // *IEEE Transactions on Acoustics, Speech, and Signal Processing*. 1976. Т. 24, № 4. С. 320–327. DOI: <https://doi.org/10.1109/TASSP.1976.1162830>
5. *Carter G., Nuttall A., Cable P.* The smoothed coherence transform // *Proceedings of the IEEE*. 1973. Т. 61, № 10. С. 1497–1498. DOI: <https://doi.org/10.1109/PROC.1973.9300>
6. *Chen L., Liu Y., Kong F., He N.* Acoustic source localization based on generalized cross-correlation time-delay estimation // *Procedia Engineering*. 2011. Т. 15. С. 4912–4919. DOI: <https://doi.org/10.1016/j.proeng.2011.08.915>
7. *Pena D. S., Lima A. D. L., de Sousa Jr. V. A., Silveira L. F., Martins A. M.* Robust time delay estimation based on non-Gaussian impulsive acoustic channel // *Journal of Communication and Information Systems*. 2020. Т. 35, № 1. С. 86–93. [Електронний ресурс]. URL: <https://jcis.sbrt.org.br/jcis/article/view/687/482>
8. *Wang J., Qian X., Pan Z., Zhang M.* GCC-PHAT with speech-oriented attention for robotic sound source localization // 2021 IEEE International Conference on Robotics and Automation (ICRA). 2021. С. 13752–13758. DOI: <https://doi.org/10.1109/ICRA48506.2021.956188>
9. *Schmidt R. O.* Multiple emitter location and signal parameter

- estimation // IEEE Transactions on Antennas and Propagation. 1986. T. 34, № 3. С. 276–280. DOI: <https://doi.org/10.1109/TAP.1986.1143830>
10. *Hwang H. K., Aliyazicioglu Z., Grice M., Yakovlev A.* Direction of arrival estimation using a Root-MUSIC algorithm // International MultiConference of Engineers and Computer Scientists 2008 (IMECS 2008). 2008. Vol. II.
 11. *Liu X., Liu C., Liao G.* Polynomial coefficient finding for Root-MUSIC // Journal of Electronics (China). 2009. T. 26, № 5. С. 543–548. DOI: <https://doi.org/10.1007/s11767-009-0142-7>
 12. *Das O., Abel J. S., Smith J. O.* FAST MUSIC—An efficient implementation of the MUSIC algorithm for frequency estimation of approximately periodic signals // 21st International Conference on Digital Audio Effects (DAFx-18). 2018.
 13. *Huang Q., Lu N.* Optimized real-time MUSIC algorithm with CPU–GPU architecture // IEEE Access. 2021. DOI: <https://doi.org/10.1109/ACCESS.2021.3070980>
 14. *Aaltonen T.* FPGA implementation of MUSIC direction of arrival algorithm using high-level synthesis : магістер. дис. Tampere University, 2023. [Електронний ресурс]. URL: <https://urn.fi/URN:NBN:fi:tuni-202401091317>
 15. *Roy R., Kailath T.* ESPRIT—Estimation of signal parameters via rotational invariance techniques // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1989. T. 37, № 7. С. 984–995. DOI: <https://doi.org/10.1109/29.32276>
 16. *Haardt M., Zoltowski M. D., Mathews C. P., Nossek J. A.* 2D unitary ESPRIT for efficient 2D parameter estimation // 1995 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-95). 1995. T. 3. С. 2096–2099. DOI: <https://doi.org/10.1109/ICASSP.1995.478488>
 17. *Haardt M., Nossek J. A.* Unitary ESPRIT: How to obtain increased estimation accuracy with a reduced computational burden // IEEE Transactions on Signal Processing. 1995. T. 43, № 5. С. 1232–1242. DOI: <https://doi.org/10.1109/78.382406>
 18. *Römer F., Haardt M., Del Galdo G.* Analytical performance assessment of multi-dimensional matrix- and tensor-based ESPRIT-type algorithms // IEEE Transactions on Signal Processing. 2014. T. 62, № 10. С. 2611–2625. DOI: <https://doi.org/10.1109/TSP.2014.2313530>
 19. *Zeng W., He J., Li H., Zhu X.* A SVT-ESPRIT estimation algorithm in sparse array // International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2016). 2016. С. 12–17. DOI: <https://doi.org/10.2991/iccia-16.2016.3>
 20. *Ramos A. L. L., Holm S., Gudvangen S., Otterlei R.* Delay-and-sum beamforming for direction of arrival estimation applied to gunshot acoustics // Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense X (Proc. SPIE). 2011. T. 8019. Art. 80190U. DOI: <https://doi.org/10.1117/12.886833>

21. *Perrot V., Polichetti M., Varray F., Garcia D.* So you think you can DAS? A viewpoint on delay-and-sum beamforming // *Ultrasonics*. 2021. T. 111. Art. 106309. DOI: <https://doi.org/10.1016/j.ultras.2020.106309>
22. *Capon J.* High-resolution frequency-wavenumber spectrum analysis // *Proceedings of the IEEE*. 1969. T. 57, № 8. C. 1408–1418. DOI: <https://doi.org/10.1109/PROC.1969.7278>
23. *Brandstein M. S., Silverman H. F.* A practical methodology for speech source localization with microphone arrays // *Computer Speech & Language*. 1997. T. 11, № 2. C. 91–126.
24. *Grinstein E., Tengan E., Çakmak B., Dietzen T., Nunes L., van Waterschoot T., Brookes M., Naylor P. A.* Steered response power for sound source localization: A tutorial review [Электронный ресурс]. DOI: <https://doi.org/10.48550/arXiv.2405.02991>
25. *Grondin F., Michaud F.* Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations // *Robotics and Autonomous Systems*. 2019. T. 113. C. 63–80. DOI: <https://doi.org/10.1016/j.robot.2019.01.002>
26. *Sathish K., Chinthaginjala R., Kim W., Rajesh A., Corchado J. M., Abbas M.* Underwater wireless sensor networks with RSSI-based advanced efficiency-driven localization and unprecedented accuracy // *Sensors*. 2023. T. 23, № 15. Art. 6973. DOI: <https://doi.org/10.3390/s23156973>
27. *Deng F., Guan S., Yue X., Gu X., Chen J., Lv J.* Energy-based sound source localization with low power consumption in wireless sensor networks // *IEEE Transactions on Industrial Electronics*. 2017. T. 64, № 6. C. 4894–4902. DOI: <https://doi.org/10.1109/TIE.2017.2652394>
28. *Alves M., Coelho R., Dranka E.* Effective acoustic energy sensing exploitation for target sources localization in urban acoustic scenes. [Электронный ресурс]. DOI: <https://doi.org/10.48550/arXiv.1910.02709>
29. *Hu Y. H., Li D.* Energy-based collaborative source localization using acoustic micro-sensor array // *IEEE Workshop on Multimedia Signal Processing (MMSP 2002)*. 2003. C. 509–512. DOI: <https://doi.org/10.1109/MMSP.2002.1203323>
30. *Khalaf-Allah M.* Emitter location using frequency difference of arrival measurements only // *Sensors*. 2022. T. 22, № 24. Art. 9642. DOI: <https://doi.org/10.3390/s22249642>
31. *Zhang B., Hu Y., Wang H., Zhuang Z.* Underwater source localization using TDoA and FDOA measurements with unknown propagation speed and sensor parameter errors // *IEEE Access*. 2018. T. 6. C. 36645–36661. DOI: <https://doi.org/10.1109/ACCESS.2018.2852636>
32. *Li X., Girin L., Horaud R., Alameda-Pineda X.* Multiple sound source localization with DP-RTF features and GMM-based clustering. [Электронный ресурс]. URL: <https://arxiv.org/abs/1611.01172>
33. *Park M., Sim K., Yang H.* Multiple sound source localization using GMM-based mask with diffuse component suppression // *INTER-NOISE 2020*. 2020. T. 261, № 3.

- [Електронний ресурс]. URL: <https://www.ingentaconnect.com/contentone/ince/incesp/2020/00000261/00000003/art00084>
34. *Fuchs J.* Monaural sound localization: A probabilistic approach. Graz University of Technology, 2008. [Електронний ресурс]. URL: <https://www.spsc.tugraz.at/system/files/MonauralSoundLocalization.pdf>
35. *Tan T.-H., Lin Y.-T., Chang Y.-L., Alkhaleefah M.* Sound source localization using a convolutional neural network and regression model // *Sensors*. 2021. Т. 21, № 23. Art. 8031. DOI: <https://doi.org/10.3390/s21238031>
36. *Tang D., Taseska M., van Waterschoot T.* Toward learning robust contrastive embeddings for binaural sound source localization // *Frontiers in Neuroinformatics*. 2022. Т. 16. Art. 942978. DOI: <https://doi.org/10.3389/fninf.2022.942978>
37. *Correia S. D., Tomic S., Beko M.* A feed-forward neural network approach for energy-based acoustic source localization // *Journal of Sensor and Actuator Networks*. 2021. Т. 10, № 2. Art. 29. DOI: <https://doi.org/10.3390/jsan10020029>
38. *Adavanne S., Politis A., Nikunen J., Virtanen T.* Sound event localization and detection of overlapping sources using convolutional recurrent neural networks // *IEEE Journal of Selected Topics in Signal Processing*. 2018. Т. 13, № 1. С. 34–48. DOI: <https://doi.org/10.1109/JSTSP.2018.2885636>
39. *Hu F., Song X., He R., Yu Y.* Sound source localization based on residual network and channel attention module // *Scientific Reports*. 2023. Т. 13. Art. 32657. DOI: <https://doi.org/10.1038/s41598-023-32657-7>
40. *Kuang S., Shi J., van der Heijden K., Mehrkanoon S.* BAST: Binaural audio spectrogram transformer for binaural sound localization [Електронний ресурс]. URL: <https://arxiv.org/abs/2207.03927v2>
41. *Berg A., Gulin J., O'Connor M., Zhou C., Åström K., Oskarsson M.* wav2pos: Sound source localization using masked autoencoders // *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. 2024. DOI: <https://doi.org/10.1109/IPIN62893.2024.10786105>
42. *Zhang D., Wang S., Belatreche A., Wei W., Xiao Y., Zheng H., Zhou Z., Zhang M., Yang Y.* Spike-based neuromorphic model for sound source localization // *NeurIPS* 2024. 2024. [Електронний ресурс]. URL: <https://openreview.net/forum?id=CyCDqnrymT>
43. *Wu Y., Ayyalasomayajula R., Bianco M. J., Bharadia D., Gerstoft P.* SSLIDE: Sound source localization for indoors based on deep learning [Електронний ресурс]. DOI: <https://doi.org/10.48550/arXiv.2010.14420>
44. *Berg A., Engman J., Gulin J., Åström K., Oskarsson M.* Learning multi-target TDoA features for sound event localization and detection [Електронний ресурс]. URL: <https://arxiv.org/abs/2408.17166>
45. *Pujol H., Bavu É., Garcia A.* BeamLearning: An end-to-end deep learning approach for the angular localization of sound sources using raw multichannel

- acoustic pressure data // The Journal of the Acoustical Society of America. 2021. Т. 149, № 6. С. 4069–4081. DOI: <https://doi.org/10.1121/10.0005046>
46. Merkofer J. P., Revach G., Shlezinger N., van Sloun R. J. G. Deep augmented MUSIC algorithm for data-driven DoA estimation // ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. 2022. С. 4613–4617. DOI: <https://doi.org/10.1109/ICASSP43922.2022.9746637>
47. Elbir A. M. DeepMUSIC: Multiple signal classification via deep learning // IEEE Sensors Letters. 2020. Т. 4, № 6. С. 1–4. DOI: <https://doi.org/10.1109/LSSENS.2020.2980384>
48. Ji J., Mao W., Xi F., Chen S. TransMUSIC: A transformer-aided subspace method for DOA estimation with low-resolution ADCs [Електронний ресурс]. DOI: <https://doi.org/10.48550/arXiv.2309.08174>
49. Chen J., Hudson R. E., Yao K. A comparative study on time delay estimation in reverberant and noisy environments // Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005). 2005. DOI/URL: <https://www.jingdongchen.com/conferencespapers/%282005%29A%20comparative%20study%20on%20time%20delay%20estimation%20in%20reverberant%20and%20noisy%20environments.pdf>
50. DiBiase J. H. A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays : PhD thesis. Providence, Rhode Island : Brown University, 2000. [Електронний ресурс]. URL: <https://www.glat.info/ma/av16.3/2000-DiBiaseThesis.pdf>
51. Brandstein M. S., Ward D. B. Microphone arrays: signal processing techniques and applications. Berlin : Springer, 2001. [Електронний ресурс]. URL: <https://link.springer.com/book/10.1007/978-3-662-04619-7>

Рябий, М.О., Шатохін, М.А

ОГЛЯД МЕТОДІВ ВИЗНАЧЕННЯ МІСЦЕЗНАХОДЖЕННЯ ДЖЕРЕЛ ЗВУКУ

У роботі виконано огляд та систематичне порівняння сучасних методів локалізації джерел звуку, зокрема часових (TDoA, GCC), підпросторових (MUSIC, ESPRIT), методів формування променя, статистичних підходів трекінгу та нейромережових і гібридних рішень. Для кожного класу методів проаналізовано ключові переваги та обмеження, а також придатність до роботи в реальних акустичних умовах, зокрема за наявності шуму, реверберації та кількох одночасних джерел. На основі порівняльного аналізу зроблено висновок, що найбільш перспективними для практичних систем є гібридні архітектури, які поєднують фізично інтерпретовані ознаки з адаптивними моделями машинного навчання. Запропоновано гібридну концепцію багатоджерельної локалізації, що використовує крос-спектральні ознаки GCC та multi-head нейромережову модель для оцінювання кількості активних джерел, їх напрямків і впевненості прогнозу.

Ключові слова: акустична локалізація джерел, мікрофонний масив, часова затримка приходу сигналу, глибоке навчання, фізично-інформоване навчання, багатоджерельна акустична сцена

Ryaby, M., Shatokhin, M

OVERVIEW OF METHODS FOR DETERMINING THE LOCATION OF SOUND SOURCES

The paper provides an overview and systematic comparison of modern methods for sound source localization, including time-based (TDoA, GCC), subspace-based (MUSIC, ESPRIT), beamforming methods, statistical tracking approaches, and neural network and hybrid solutions. For each class of methods, the key advantages and limitations are analyzed, as well as their suitability for use in real acoustic conditions, in particular in the presence of noise, reverberation, and multiple simultaneous sources. Based on a comparative analysis, it is concluded that hybrid architectures that combine physically interpreted features with adaptive machine learning models are the most promising for practical systems. A hybrid concept of multi-source localization is proposed, which uses GCC cross-spectral features and a multi-head neural network model to estimate the number of active sources, their directions, and the confidence of the prediction.

Keywords: *acoustic source localization, microphone array, time difference of arrival, deep learning, physics-informed learning, multi-source acoustic scene*

Стаття подана до редакції: 18/11/2025

Стаття прийнята до опублікування: 2/12/2025

Стаття опублікована: 30/12/2025

Стаття поширюється на умовах ліцензії CC BY 4.0