

## ФОРМАЛІЗАЦІЯ ЗАДАЧІ ІНТЕЛЕКТУАЛЬНОГО ГОЛОСОВОГО УПРАВЛІННЯ БПЛА

*Анотація:* У статті проаналізовано еволюцію та поточний стан методів управління безпілотними літальними апаратами (БПЛА), із зосередженням на інтеграції інтелектуальних систем для підвищення їхньої автономності та надійності. Встановлено, що традиційні інтерфейси є недостатньо оперативними в динамічних умовах. Розроблено змістовну та формалізовано математичну моделі задачі інтелектуального голосового керування, яка враховує не тільки пряме виконання команд, але й функціонування в режимі випередження з мінімізацією часу реакції. Обґрунтовано доцільність застосування алгоритмів глибинного навчання (DNN) для надійного розпізнавання мовлення в умовах акустичних перешкод та трансформації нечітких команд у точні керуючі дії. Описано архітектуру програмного забезпечення голосового модуля, що забезпечує контекстно-чутливий семантичний аналіз та стійкість зв'язку з автопілотом (протокол MAVLink). Результати дослідження підтверджують можливість створення високо адаптивної системи управління, що має критичне значення для складних місій, включаючи пошуково-рятувальні та військові операції.

*Ключові слова:* методи управління безпілотними літальними апаратами, інтелектуальне управління, голосове керування, розпізнавання мовлення, математична модель, штучний інтелект.

### Вступ

У контексті стрімкого розвитку безпіотної авіації та інтеграції в неї передових можливостей штучного інтелекту, створення інтелектуальних систем управління БПЛА, особливо за допомогою голосових інтерфейсів, стає не лише інноваційним, а й критично актуальним завданням. Потреба у вдосконаленні професійного управління складними авіаційними комплексами, особливо в умовах обмеженої видимості, високого психологічного навантаження на оператора та потреби у швидкому реагуванні, зростає з кожним днем [1]. Це відкриває широкі горизонти для розробки систем, що можуть сприймати та аналізувати людську мову, виконувати різноманітні завдання від простих маневрів до комплексного керування місією на базі контекстуального розуміння ситуації.

З огляду на це, важливість розробки технології голосового управління БПЛА з вбудованим штучним інтелектом не може бути переоцінена. Така система відіграватиме ключову роль у підвищенні ефективності та безпеки виконання польотних завдань, даючи можливість операторам управляти апаратами за допомогою природної

мови, асистувати в прийнятті рішень на основі аналізу сенсорних даних та забезпечуючи роботу у «режимі випередження» [2].

На сучасному ринку та в академічних дослідженнях існує багато рішень для управління БПЛА, починаючи від систем ручного пілотування (наприклад, ArduPilot, PX4) до високоавтономних платформ, що використовують комп'ютерний зір. Однак більшість з них використовують традиційні інтерфейси (пульт, наземна станція), які не завжди забезпечують необхідну швидкість і гнучкість. Аналіз цих існуючих рішень дозволяє визначити ключові аспекти, які необхідно врахувати при розробці нового продукту, а саме: необхідність надійного розпізнавання мовлення в умовах шуму та мінімальний час затримки між командою та її виконанням, що і є головним об'єктом даного дослідження.

### **Аналіз сучасних методів управління БПЛА**

Сучасні методи управління БПЛА демонструють стрімку еволюцію від простих систем ручного пілотування до високоавтономних платформ, керованих штучним інтелектом. Класифікація цих методів ґрунтується на рівні залучення оператора та ступені автономності системи [3].

Так, за рівнем автономності, можна виділити три основні методи управління:

1. Дистанційно-пілотований (Ручне управління), коли оператор безпосередньо контролює виконавчі механізми апарата, передаючи команди на відхилення рульових площин та керування тягою [4]. Цей метод є простим, але вимагає постійної уваги, має високу затримку реакції та використовується здебільшого для завдань, що не вимагають високої точності чи тривалого польоту (наприклад, візуальна розвідка на короткій дистанції);

2. Навігаційний (Напіваавтономний), коли управління здійснюється шляхом завдання точок маршруту (way-points) або визначення зони патрулювання. Частина обчислень, в цьому випадку, переноситься на борт БПЛА, який самостійно підтримує задану висоту, швидкість та коригує курс за допомогою GPS та інерційних систем [4]. Прикладом такого управління є протокол MAVLink, який забезпечує уніфікований зв'язок між наземною станцією та автопілотом (наприклад, ArduPilot, PX4) [5].

3. Автономний (Інтелектуальний), коли БПЛА отримує лише високорівневу місію і самостійно приймає рішення щодо оптимального маршруту, уникнення перешкод та виконання завдань, використовуючи бортові сенсори, комп'ютерний зір та алгоритми машинного навчання (ШН). Серед цих методів найбільш перспективним є метод заснований на навчанні (Learning-based methods), який дозволяє БПЛА адаптуватися до складних динамічних умов, таких як зміна погоди або поява невідомих перешкод.

Що стосується використання систем, що керуються штучним інтелектом, то сучасні дослідження зосереджуються на двох основних підходах [3]:

1. Використання методів, які засновані на оптимізації (Optimization-based): включають класичні алгоритми, такі як A\*, PSO та GWO, і використовуються для швидкого пошуку оптимального шляху в статичних або частково відомих середовищах;

2. Використання методів, які засновані на навчанні (Learning-based): включають застосування глибокого навчання (DL) та навчання з підкріпленням (RL), зокрема алгоритмів DRL та A3C, що забезпечує високу гнучкість і стійкість БПЛА. Ці методи є основою для реалізації системи голосового управління, оскільки вони забезпечують надійне розпізнавання мовлення в умовах шуму та контекстної інтерпретації команд.

Таким чином, інтеграція голосових інтерфейсів є логічним кроком у розвитку автономних систем. Вона мінімізує фізичне втручання оператора і дозволяє йому сконцентруватися на стратегічних аспектах місії.

Реалізація такої системи складається з трьох ключових рівнів обробки: акустичний модуль, лінгвістичний процесор та семантичний інтерпретатор.

### **Постановка та формалізація задачі інтелектуального голосового управління**

Задача інтелектуального голосового керування БПЛА полягає у забезпеченні надійної, швидкої та контекстно-чутливої трансформації голосової команди оператора у послідовність керуючих дій для бортового комп'ютера (автопілота) та включає наступні основні вимоги до системи:

1. Надійність розпізнавання (Assurance): система повинна демонструвати високу точність розпізнавання мовлення (>95%) навіть в умовах високого фонового шуму (вітер, звук двигунів, зовнішні завади) та змінної акустики, що забезпечується використанням вдосконалених глибоких нейронних мереж (DNN), здатних ефективно виділяти акустичні ознаки мови на фоні перешкод [1];

2. Мінімальна затримка (Latency): система управління має функціонувати в режимі реального часу, коли загальний час від моменту вимови команди оператором до початку її фізичного виконання БПЛА ( $T_{total}$ ) не повинен перевищувати критичного порогу (зазвичай менше 500 мс) для запобігання втраті контролю або зіткненню;

3. Контекстна чутливість (Context Awareness): система не повинна просто виконувати команду, а має враховувати поточний стан БПЛА (P), його місце розташування, швидкість, висоту та зовнішні фактори (E). Наприклад, команда «знижуйся» повинна бути інтерпретована як небезпечна, якщо поточна висота менше мінімально дозволена.

Основною особливістю реалізації даної задачі є необхідність забезпечення роботи системи у «режимі випередження». Це означає, що система повинна не тільки реагувати на команду, а й паралельно прогнозувати потенційні ризики (зіткнення, вихід за межі зони польоту, втрата зв'язку) та, у разі виявлення ризику, коригувати або блокувати команду ще до її повного виконання.

Умовна діаграма варіантів використання системи голосового управління БПЛА, яка демонструє її базові варіанти використання, представлена на рис. 1.



Рисунок 1. Умовна діаграма варіантів використання системи голосового управління БПЛА

Для формалізації означеної вище задачі будемо використовувати модель багаторівневого перетворення вводу в керуючу дію. Основна мета – мінімізація загального часу реакції  $T_{total}$  при максимізації ймовірності правильного виконання команди  $P_{correct}$ .

Даний процес складається з наступних етапів:

1. Побудова моделі трансформації команди, яка включає керуючу дію  $C$ , що є результатом інтелектуального алгоритму  $A$ , який опрацьовує акустичний сигнал  $S$  з урахуванням вектора стану середовища  $E$  і вектора внутрішніх параметрів БПЛА  $P$ .

$$C = A(S | E, P),$$

де  $S$  – акустичний сигнал, який являє собою вектор параметрів, отриманий з мікрофона, що включає спектральні характеристики та рівень шуму;  $E = (E_{noise}, E_{wind}, E_{jam})$  – вектор зовнішніх факторів, який описує рівень акустичного шуму, швидкість вітру та рівень GPS-завад;  $P = (P_{coord}, P_{vel}, P_{batt}, P_{health})$  – вектор внутрішніх параметрів, який містить поточні координати, швидкість, заряд батареї та технічний стан, а алгоритм  $A$  є функцією, що включає ASR-модуль (розпізнавання), NLP-модуль (семантичний аналіз) та Decision-Making Unit (прийняття рішень), які зазвичай реалізуються за допомогою рекурентних або трансформерних нейронних мереж.

Керуюча дія, в даному випадку, формується як послідовність команд MAVLink для автопілота (наприклад, MAV\_CMD\_NAV\_WAYPOINT).

2. Обмеження реального часу, коли загальний час реакції  $T_{total}$  визначається як сума часових затримок на кожному етапі:

$$T_{total} = T_{acq} + T_{proc} + T_{comm} + T_{exec},$$

де  $T_{acq}$  – час захоплення та початкової фільтрації акустичного сигналу;  $T_{proc}$  – час інтелектуальної обробки (розпізнавання та семантичний аналіз). Цей показник має бути мінімізований за рахунок оптимізації архітектури DNN [6];  $T_{comm}$  – час передачі команди від наземної станції до бортового комп'ютера через радіоканал (залежить від протоколу MAVLink та швидкості каналу);  $T_{exec}$  – час виконання команди автопілотом.

Основна умова функціонування:

$$\min(T_{total}) \text{ при умові } P_{correct}(C) \geq P_{t_{res}old},$$

де  $P_{t_{res}old}$  – мінімально допустима ймовірність правильного виконання команди (наприклад,  $P_{t_{res}old} = 0,96$ ).

Для верифікації цієї умови проведено емпіричне тестування, результати якого наведені у табл. 1.

Таблиця 1.

Точність розпізнавання команд

№	Команда	Кількість тестових спроб	Кількість правильних розпізнавань	Точність, %
1	«Зліт»	50	48	96 %
2	«Посадка»	50	47	94 %
3	«Повернення додому»	40	37	92,5 %
4	«Розвідка»	30	28	93,3 %
5	«Фото/відео»	30	29	96,7 %
6	«Аварійна посадка»	20	19	95 %
	<b>Середній показник</b>	<b>220</b>	<b>208</b>	<b>94,5 %</b>

### Архітектура та методи розпізнавання мовлення

Для успішної реалізації інтелектуального голосового управління БПЛА критично важливим є вибір оптимальної архітектури програмного забезпечення та методів автоматичного розпізнавання мовлення (ASR), тому що, на відміну від традиційних командних інтерфейсів, голосовий модуль має забезпечити контекстно-чутливий семантичний аналіз у реальному часі.

Таким чином, архітектура системи розпізнавання мовлення для БПЛА має відповідати вимогам модульності, масштабованості та стійкості до зовнішніх і внутрішніх перешкод та будуватися за принципом послідовної інтеграції трьох ключових модулів:

1. Акустичного модуля (Front-end), який здійснює захоплення сирого аудіосигналу з мікрофона та його попередню обробку. Завдяки алгоритмам шумозаглушення відфільтровуються фонові шуми (вітер, технічні завади), а сигнал нормалізується для подальшого аналізу. Далі виконується вилучення акустичних ознак – сигнал

перетворюється у часову послідовність спектральних коефіцієнтів. Найбільш поширеними параметрами тут є мел-частотні кепстральні коефіцієнти (MFCC) [1], які забезпечують стійкість до варіацій тембру голосу, інтонацій та рівня гучності;

2. Лінгвістичного модуля (ASR Core), де отримані спектральні ознаки надходять у ядро системи автоматичного розпізнавання мовлення (Automatic Speech Recognition, ASR). Тут застосовуються моделі глибокого навчання — рекурентні або трансформерні нейронні мережі, які здійснюють перетворення послідовності ознак у текстову транскрипцію. Основним критерієм ефективності цього модуля є мінімізація коефіцієнта помилок у словах (Word Error Rate, WER), що визначає точність розпізнавання мовних команд навіть у складних акустичних умовах;

3. Семантичного модуля (Decision-Making Unit), де розпізнаний текст проходить семантичний аналіз і трансформується у конкретні команди автопілота, наприклад, MAV\_CMD\_NAV\_WAYPOINT або MAV\_CMD\_DO\_SET\_RELAY. Даний модуль інтегрує контекстну інформацію — вектори зовнішніх факторів E (шум, вітер, завади GPS) та внутрішніх параметрів P (координати, швидкість, заряд батареї, технічний стан). Це дозволяє перевірити команду на логічність, безпеку та відповідність ситуаційним умовам, запобігаючи ризиковим або некоректним діям. Такий підхід реалізує так званий *режим випередження* [7], коли система не лише виконує команду, але й оцінює її доцільність у поточному контексті.

Для досягнення більш високої точності розпізнавання мовлення в умовах, які характерні для експлуатації БПЛА, застосовуються глибокі нейронні мережі (Deep Neural Networks, DNN), здатні виявляти складні патерни у послідовних даних [8]. Їх використання забезпечує стійкість до шумів, завад і варіативності мовлення оператора.

Серед найбільш поширених, можна визначити наступні мережі:

1. Рекурентні нейронні мережі (RNN) та LSTM/GRU, які тривалий час були стандартом для обробки послідовностей, зокрема мовних сигналів. Їх головною перевагою є здатність зберігати та передавати інформацію про довгострокові залежності у мовленні. Це критично важливо для розуміння довгих або багатокomпонентних команд, коли значення кінця фрази визначається контекстом її початку. Архітектури LSTM (Long Short-Term Memory) та GRU (Gated Recurrent Unit) розв'язують проблему «згасання градієнта» у класичних RNN, що дозволяє ефективніше навчати моделі на тривалих послідовностях.

2. Згорткові нейронні мережі (CNN), які традиційно застосовуються у задачах комп'ютерного зору. Вони успішно використовуються і в акустичному моделюванні, де їхня роль полягає у виділенні просторово-часових ознак зі спектрограм аудіосигналу. Ці мережі ефективно фільтрують локальні закономірності у спектрі, що дозволяє моделі краще відокремлювати корисний мовний сигнал від шумових компонентів. Таким чином, CNN можуть виступати як фронт-енд для подальших послідовних моделей [1].

3. Трансформерні моделі (Attention Mechanism), які є сучасним стандартом у розпізнаванні мовлення та їхні численні варіації, зокрема BERT, Whisper та інші. В їх основі лежить механізм уваги (Attention Mechanism), який дозволяє оцінювати значущість кожного слова чи акустичного фрагмента відносно всієї послідовності. Це радикально знижує залежність від порядку слів і значно підвищує точність розпізнавання. На практиці трансформери забезпечують кращу узагальнювальну здатність і швидше навчання порівняно з RNN/LSTM, що робить їх основою найсучасніших ASR-систем [9].

Отже, еволюція архітектур – від RNN та LSTM/GRU через CNN до трансформерів – відображає рух від класичних послідовних моделей до більш універсальних і контекстно-орієнтованих рішень, що є оптимальними для складних умов експлуатації БПЛА.

Зважаючи на те, що експлуатація безпілотних літальних апаратів неминує буде супроводжуватись високим рівнем акустичних перешкод, серед яких будуть домінувати: шум двигунів, вплив вітру та фактори зовнішнього середовища, то для забезпечення стабільної роботи системи розпізнавання мовлення, у таких умовах, слід застосовувати спеціалізовані методи обробки сигналу та адаптації моделей. На етапі попередньої обробки сигналу слід використовувати шумозаглушення, що реалізується за допомогою цифрових фільтрів та автоенкодерів, навчених відокремлювати голос оператора від фонового шуму, завдяки чому на вхід глибоким нейронним мережам надходить більш чистий сигнал. Додатково слід застосовувати адаптивні моделі: наприклад, нейронні мережі, які забезпечують донавчання на спеціалізованих вибірках (містять характерні шуми польоту), що дозволяє значно підвищити їхню продуктивність у реальних умовах експлуатації [8].

Загалом, ефективність системи безпосередньо визначається досягненням високої точності розпізнавання, тому емпіричні дані, що наведені в табл. 1, підтверджують результативність описаних методів. Зокрема, дана таблиця демонструє, що навіть за умов сильного акустичного шуму система здатна підтримувати заданий рівень надійності  $P_{threshold}$ .

### Висновки

Аналіз сучасних методів управління БПЛА показав їх критичну залежність від глибокої та ефективної інтеграції технологій штучного інтелекту, зокрема у сфері голосового керування. Встановлено, що традиційні методи командного керування недостатньо ефективні, оскільки вони не здатні забезпечити необхідну гнучність, швидкість реакції та адаптивність в умовах динамічного середовища та високих акустичних перешкод.

Проведена формалізація задачі інтелектуального голосового управління, яка була визначена як задача забезпечення мінімальної загальної затримки  $T_{total}$  та максимальної надійності розпізнавання  $A_{accuracy}$ , чітко окреслює багатоступеневу структуру перетворення: від акустичного сигналу (що вимагає попередньої обробки для усунення домінуючого шуму пропелерів) до керуючої команди MAVLink (яка повинна бути контекстно-залежною).

Обґрунтовано, що вирішення проблеми затримки  $T_{total}$  вимагає реалізації системи у «режимі випередження» (Context-Aware Control), де семантичний аналізатор не лише інтерпретує поточну команду, а й прогнозує наступні дії пілота на основі параметрів польоту  $P$  та стану оточення  $E$ . Для практичної реалізації цього режиму критично важливою є оптимізація архітектури глибоких нейронних мереж (DNN) з використанням підходів Edge Computing, що має забезпечувати мінімізацію часу обробки  $T_{proc}$  на ресурсно-обмежених бортових пристроях.

Застосування гібридних моделей глибоких нейронних мереж (CNN, Transformer) для акустичного моделювання та семантичного аналізу підтверджує можливість створення надійних, високопродуктивних, контекстно-чутливих і стійких до акустичних перешкод систем керування БПЛА нового покоління. Отримані результати та сформульовані вимоги до архітектури системи створюють необхідну теоретичну базу для її детальної розробки та моделювання програмного забезпечення голосового модуля.

### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ashwini P., Dr. Bharathi S, Ananya K.Nayaka. A Review on Different Feature Recognition Techniques for Speech Process in Automatic Speech Recognition. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*. 2019. Vol. 8, no. 09. P. 1953–1957. URL: <https://www.ijstr.org/final-print/sep2019/A-Review-On-Different-Feature-Recognition-Techniques-For-Speech-Process-In-Automatic-Speech-Recognition.pdf>.
2. MAVLink Basics – Dev documentation. *ArduPilot - Versatile, Trusted, Open*. URL: <https://ardupilot.org/dev/docs/mavlink-basics.html>.
3. ВСЕБІЧНИЙ ОГЛЯД МЕТОДІВ НАВІГАЦІЇ БПЛА НА БАЗІ ШТУЧНОГО ІНТЕЛЕКТУ | ТЕЛЕКОМУНІКАЦІЙНІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ. *ТЕЛЕКОМУНІКАЦІЙНІ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ*. URL: <https://tit.dut.edu.ua/index.php/telecommunication/article/view/2621>.
4. Іваненко Ю. В., Ляшенко О. С., Філімончук Т. В. ОГЛЯД МЕТОДІВ КЕРУВАННЯ БЕЗПІЛОТНИМИ ЛІТАЛЬНИМИ АПАРАТАМИ. *Control, Navigation and Communication Systems*. 2023. № 1. С. 26–30. URL: <https://journals.nupp.edu.ua/sunz/article/download/2822/2230/3959>.
5. MAVLink Interface – Dev documentation. *ArduPilot - Versatile, Trusted, Open*. URL: <https://ardupilot.org/dev/docs/mavlink-commands.html>.
6. JIASI CHEN, XUKAN RAN. Deep Learning With Edge Computing: A Review. *PROCEEDINGS OF THE IEEE*. 2019. Vol. 107, no. 8. P. 1655–1674. URL: [https://www.cs.ucr.edu/~jiasi/pub/deep\\_edge\\_review.pdf](https://www.cs.ucr.edu/~jiasi/pub/deep_edge_review.pdf).

7. Context-Aware Deep Reinforcement Learning for Autonomous Robotic Navigation in Unknown Area / Jingsong Liang et al. URL: <https://proceedings.mlr.press/v229/liang23a/liang23a.pdf>.

8. Yanjue Song, Stijn Kindt, Nilesh Madhu. Drone Ego-Noise Cancellation for Improved Speech Capture using Deep Convolutional Autoencoder Assisted Multistage Beamforming. 2022. URL: [https://www.researchgate.net/publication/362960350\\_Drone\\_Ego-Noise\\_Cancellation\\_for\\_Improved\\_Speech\\_Capture\\_using\\_Deep\\_Convolutional\\_Autoencoder\\_Assisted\\_Multistage\\_Beamforming](https://www.researchgate.net/publication/362960350_Drone_Ego-Noise_Cancellation_for_Improved_Speech_Capture_using_Deep_Convolutional_Autoencoder_Assisted_Multistage_Beamforming).

9. Attention Is All You Need. *arXiv.org*. URL: <https://arxiv.org/abs/1706.03762>.